HAL
open science

Etude de représentations parcimonieuses des statistiques d'erreur d'observation pour différentes métriques. Application à l'assimilation de données images

Vincent Chabot

To cite this version:

Vincent Chabot. Etude de représentations parcimonieuses des statistiques d'erreur d'observation pour différentes métriques. Application à l'assimilation de données images. Traitement du signal et de l'image [eess.SP]. Université de Grenoble, 2014. Français. NNT : 2014GRENM094 . tel-01073582v2

HAL Id: tel-01073582

https://theses.hal.science/tel-01073582v2

Submitted on 22 May 2017

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 GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Mathématiques Appliquées**

Arrêté ministériel : 7 aout 2006

Présentée par

Vincent Chabot

Thèse dirigée par **Arthur Vidard**
et codirigée par **Maëlle Nodet**

préparée au sein du **Laboratoire Jean Kuntzmann**
et de **École Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique**

Étude de représentations parcimonieuses des statistiques d'erreur d'observation pour différentes métriques. Application à l'assimilation de données images.

Thèse soutenue publiquement le **11 juillet 2014**,
devant le jury composé de :

Mme. Valérie Perrier
Professeur à Grenoble INP, Présidente
M. Thibaut Montmerle
Ingénieur des Ponts, des Eaux et des Forêts, Météo–France, Rapporteur
M. Thomas Corpetti
Directeur de recherche CNRS (Rennes), Rapporteur
M. Didier Auroux
Professeur à l'Université de Nice Sophia Antipolis, Examinateur
M. Arthur Vidard
Chargé de recherche INRIA, Directeur de thèse
Mme. Maëlle Nodet
Maître de conférences à l'Université Joseph Fourier, Co-Directrice de thèse
Mme. Claire Lauvernet
Chargée de recherche IRSTEA, (Lyon), Invitée
M. Nicolas Papadakis
Chargé de recherche CNRS (Bordeaux), Invité

Sommaire

Remerciements

Introduction

Chapitre 1 Introduction à l'assimilation de données **7**

1.1 Notations 8

1.2 Formulation variationnelle du problème d'assimilation 10

1.3 Préconditonnement du problème 14

1.4 Influence de la matrice B 15

1.5 Construction de la matrice B 16

Chapitre 2 Transformée en ondelettes discrète **19**

2.1 Introduction 19

2.2 Transformée en ondelettes discrète 21

2.2.1 Analyse multi-résolution 21

2.2.2 Le rôle des filtres 24

2.2.3 Propriétés des bases d'ondelettes 29

2.2.4 Opérations de seuillage 30

2.3 Ondelettes séparables en dimension 2 32

2.3.1 Construction 33

2.3.2 Obtention des coefficients 33

2.3.3 Illustration sur un exemple du stockage de l'information dans un espace d'ondelettes 35

2.3.4 Autre représentation en dimension 2 38

Chapitre 3 Utilisation d'images en assimilation de données **41**

3.1 Introduction 41

3.1.1 Atmospheric motion vectors 42

3.1.2 Estimation du champ de vitesse par mesure du flot optique 43

3.1.3 Assimilation directe d'image 45

3.2 Méthode d'assimilation utilisée dans ce manuscrit 48

3.2.1 Principe 48

3.2.2 Mesure de la distance entre deux images 49

3.3 Expériences jumelles avec des données parfaites . . . 56
3.3.1 Contexte expérimental . . . 56
3.3.2 Diagnostic utilisé . . . 59
3.3.3 Validation des opérateurs . . . 61
3.3.4 Influence de la longueur de la fenêtre d'assimilation . . . 64
3.3.5 Impact de la régularisation par l'ébauche . . . 67
3.4 Conclusion . . . 70

Chapitre 4 Impact d'un bruit d'observation non corrélé en espace 73
4.1 Introduction . . . 73
4.1.1 Rappel . . . 73
4.1.2 Notations . . . 74
4.1.3 Matrice de covariance ou matrice de corrélation ? . . . 74
4.2 Matrice de covariance dans différents espaces . . . 75
4.2.1 Distance entre gradients . . . 76
4.2.2 Dans une base d'ondelettes orthonormées . . . 79
4.2.3 Distance dans une trame de curvelettes . . . 79
4.2.4 Distance entre les gradients normalisés . . . 82
4.3 Expériences effectuées . . . 85
4.3.1 Impact du niveau de bruit . . . 86
4.3.2 Intérêt de la perte d'information en présence de biais . . . 88
4.3.3 Impact de la fréquence d'observation . . . 90

Chapitre 5 Impact d'un bruit d'observation corrélé en espace 93
5.1 Introduction . . . 93
5.2 Pixels, gradients et gradient normalisés . . . 97
5.2.1 Approximation dans une base de pixels . . . 97
5.2.2 Approximation dans l'espace des gradients . . . 97
5.2.3 Approximation pour les gradients normalisés . . . 99
5.3 Approximation scalaire dans une base d'ondelettes . . . 99
5.3.1 Construction . . . 99
5.3.2 Pourquoi est-ce une mauvaise approximation ? . . . 100
5.4 Construction formelle de la matrice de covariance dans une base d'ondelettes . . . 103
5.4.1 Obtention de la matrice de covariance en mode boîte noire . . . 103
5.4.2 Utilisation de la connaissance de la transformée en ondelette . . . 103
5.4.3 Mais au final, à quoi ça ressemble ? . . . 110
5.4.4 Propriété facilitant la construction lorsque le bruit est identiquement distribué . . . 111
5.5 Approximation diagonale dans une base d'ondelettes . . . 113
5.5.1 Impact de la diagonalisation . . . 113
5.5.2 Renormalisation . . . 116

5.5.3 Propriété intéressante pour la renormalisation (et autres) 117
5.6 Estimation de la matrice dans une trame de curvelettes 119
5.6.1 Utilisation de matrice de poids symétrique définie positive 119
5.6.2 Restriction à des matrices diagonales 122
5.7 Estimation de la matrice de covariance dans un espace de Fourier 128
5.8 Expériences 129
5.8.1 Influence du niveau de bruit 131
5.8.2 Impact de R sur le choix de B 139
5.8.3 Quelle quantité d'information est contenue dans les statistiques d'erreurs d'observation ? 142
5.8.4 Attention danger : statistiques à manipuler avec précaution ! 143
5.8.5 Anisotropie et choix de la représentation 147

Chapitre 6 Gestion des occultations dans des bases d'ondelettes 151
6.1 Données parfaites et nuages 151
6.1.1 Fonction coût dans l'espace des pixels 152
6.1.2 Expériences jumelles 153
6.1.3 Bilan partiel 155
6.2 Impact des occultations sur le modèle de corrélations 156
6.3 Comparaison d'images occultées dans un espace d'ondelettes 158
6.3.1 Approximation de la matrice de covariance sans prise en compte des occultations 160
6.3.2 Que se passe-t-il lorsques certains pixels viennent à manquer ? Illustration dans l'espace des gradients. 162
6.3.3 Pourquoi les coefficients de petite échelle sont peu affectés par un bruit corrélé ? 164
6.4 Modélisation de la variance en présence d'occultation 166
6.4.1 Pourcentage d'information disponible 166
6.4.2 Calcul du terme d'inflation 169
6.4.3 Approximation de la variance de coefficients partiellement occultés 171
6.5 Conclusion 176

Chapitre 7 Conclusion et perspectives 177
7.1 Conclusion 177
7.2 Perspectives 178
7.2.1 Vers l'utilisation d'autres espaces multi–échelles 178
7.2.2 Prise en compte des corrélations temporelles 180
7.2.3 Gestion des occultations 181
7.2.4 Approximation par blocs diagonaux 182
7.2.5 Vers des problèmes réalistes 183
7.2.6 Autres définitions de la distance entre deux images 184

Annexe A Pourquoi la hauteur d'eau n'est-elle pas corrigée ? 187

Bibliographie 191

Remerciements

Je souhaite tout d'abord remercier mes directeurs de thèse Arthur Vidard et Maëlle Nodet pour avoir eu l'occasion d'effectuer cette thèse sous leur direction. Je les remercie en effet pour leur disponibilité, leur patience et le temps qu'ils ont pu consacrer à l'encadrement de ces travaux. Les nombreuses discussions que j'ai pu avoir avec eux ont toujours été à la fois agréables et enrichissantes. Leur adaptabilité et leur rapidité de compréhension d'un discours souvent "trop direct et trop technique" m'a été d'un grand secours durant ces années. Je tiens aussi à les remercier d'avoir su introduire une méthodologie dans mon travail. Enfin, je tiens à les remercier pour les bons moments passés, notamment autour de bières ou de jeux, durant ces quelques années à Grenoble.

Cette thèse n'aurait sans doute pas été la même sans l'implication de Nicolas Papadakis. Je tiens à le remercier pour le temps qu'il a pu consacrer à ce projet, son engouement et son foisonnement d'idées. Je tiens aussi à le remercier pour avoir su plusieurs fois me remotiver.

Les années passées à Grenoble n'auraient pas été aussi bonnes sans un environnement de travail agréable. Je tiens à remercier l'ensemble de l'équipe MOISE pour son accueil. Parmi ses membres, je tiens tout particulièrement à remercier Anne Pierson qui a toujours été disponible pour m'aider à surmonter les tâches administratives afférentes à la préparation d'un doctorat.

Un grand merci à Alexandre Hoffman que j'ai eu le plaisir d'encadrer lors de son stage de licence. Merci pour son implication et ses différentes idées sur un sujet pas toujours aisé au premier abord.

Je tiens à remercier Thibaut Montmerle et Thomas Corpetti pour le temps qu'ils ont pu consacrer à effectuer un travail de rapport de très grande qualité. Leurs nombreuses critiques, très constructives, m'ont permis d'améliorer ces travaux de thèse. Je dois la découverte du monde des ondelettes (très présent dans ces travaux) à Valérie Perrier que je remercie d'avoir accepté de présider mon jury de thèse. Enfin, je tiens à remercier Didier Auroux et Claire Lauvernet pour le temps consacré à une lecture minutieuse et pour l'intérêt porté à ces travaux.

Je n'aurai certainement pas réalisé cette thèse si Gérald Desroziers et Loïk Berre ne m'avaient pas accueilli lors de mon stage de licence au sein du GMAP. Je tiens à les remercier de m'avoir fait découvrir très tôt le monde de l'assimilation de données et de m'avoir fait comprendre l'importance des matrices de corrélations d'erreur au sein de cette discipline.

J'ai apprécié au cours de ma dernière année de thèse d'avoir pu collaborer avec Jean–François Coeurjolly. Je le remercie pour son encadrement de ma pratique de l'enseignement des statistiques au sein de l'UPMF. Je le remercie au passage pour le temps passé à répondre à mes différentes questions concernant cette discipline aussi bien pour le côté enseignement que pour le côté recherche scientifique.

Si j'ai été aussi heureux de passer quatre années au sein du LJK, cela tient en partie aux personnes que j'ai pu y côtoyer. Je pense tout d'abord à mes co–bureaux : Gaëlle et ses réactions

spontanées, aux histoires et autres râleries de Bertrand, aux déambulations inopinées de Romain. Je tiens aussi à remercier les abeilles du 36 et leur essaimage spontané grâce auquel j'ai co–burauté avec Mathias, Nelson et les mots croisés. Je tiens aussi à remercier Chloé Mimeau pour les nombreuses discussions que nous avons pu avoir (et aussi pour avoir accepté de relire et corriger une partie de ce travail de thèse). Enfin, j'ai pu apprécier de nombreuses autres personnes d'apprécier : Emilie, Roland, Innocent, Madison, Thomas, PJ, PO, Morgane (merci pour tes encouragements dans des moments pas facile), Jean-Mathieu, Lukas, Eurielle, Bénédicte, Meriem.

Cette thèse n'aurait sans doute pas été la même sans l'ensemble de mes colocataires avec qui j'ai pu passer de très bon moments, et ce aussi bien à La Tronche qu'à Venon.

Dernier point, et non des moindres, je tiens à remercier la machine à café du 7ème d'avoir produit (périodiquement) un café qui m'a permis de tenir physiquement (ainsi que de faire de nombreuses pauses) au cours de la réalisation de ces travaux.

Introduction

Pourquoi les météorologistes ont-ils tant de peine à prédire le temps avec quelque certitude ? Pourquoi les chutes de pluie, les tempêtes elles-mêmes nous semblent-elles arriver au hasard, de sorte que bien des gens trouvent tout naturel de prier pour avoir la pluie ou le beau temps, alors qu'ils jugeraient ridicule de demander une éclipse par une prière ? Nous voyons que les grandes perturbations se produisent généralement dans les régions où l'atmosphère est en équilibre instable. Les météorologistes voient bien que cet équilibre est instable, qu'un cyclone va naître quelque part ; mais où, ils sont hors d'état de le dire ; un dixième de degré en plus ou en moins en un point quelconque, le cyclone éclate ici et non pas là, et il étend ses ravages sur des contrées qu'il aurait épargnées. Si on avait connu ce dixième de degré, on aurait pu le savoir d'avance, mais les observations n'étaient ni assez serrées, ni assez précises, et c'est pour cela que tout semble dû à l'intervention du hasard. Ici encore nous retrouvons le même contraste entre une cause minime, inappréciable pour l'observateur, et des effets considérables, qui sont quelquefois d'épouvantables désastres.

Henri Poincaré (1854-1912)

Le caractère chaotique de la météorologie fut pressenti très tôt par Henri Poincaré. Comme on peut le voir dans la citation précédente, le caractère non prévisible de l'état de l'atmosphère n'est ainsi pas dû au hasard mais découle de la *sensibilité du système aux conditions initiales* mal connues et de la propriété de récurrence du système [1].

Comme le pressentait Henri Poincarré, afin d'effectuer de bonnes prévisions météorologiques, il est essentiel d'avoir accès à la meilleure condition initiale possible. Il est tout aussi fondamental, à partir de ces conditions initiales, de savoir comment les propager dans le temps afin d'obtenir une trajectoire de l'atmosphère précise à courte échéance. Au vu de la complexité des lois physiques régissant le système, arriver à prévoir la trajectoire du système à moyen terme est un travail complexe que nous sommes peu capable de faire seul. Jay Forrester (fondateur de la dynamique des systèmes) résume en peu de mot pourquoi nous ne sommes pas capables d'appréhender le comportement des systèmes complexes :

La difficulté, c'est que ces systèmes sont contre–intuitifs. Ils ne se comportent pas comme on s'y attendrait. Et il y a des raisons à cela. Le cerveau humain a formé son intuition et son expérience à partir de systèmes simples. Conduire une voiture est sans doute le système le plus complexe, le plus élaboré, que nous maîtrisons pleinement. Face à des systèmes plus complexes nous appliquons l'expérience acquise sur des systèmes simples mais les systèmes complexes ont une dynamique très différente. C'est la structure du système complet qui détermine son comportement.

Jay Forrester (1918–)

1. Cette dernière propriété signifie que pour presque toute condition initiale du système, le système repasse au cours du temps une infinité de fois arbitrairement près de cette configuration initiale.

Dans les années 1920, Richardson tenta le premier de résoudre numériquement les équations différentielles simplifiées afin de produire une prévision météorologique de six heures. Par manque de moyen de calcul, le résultat fut décevant mais son expérience fut profitable aux générations suivantes. Ce n'est cependant que dans les années cinquante, après que furent inventés les premiers ordinateurs, que [Charney et al., 1950] firent la première prévision numérique couronnée de succès.

De manière générale, l'arrivée des ordinateurs a permis de mieux appréhender le comportement de ces système complexes grâce à la résolution numérique d'équation (décrivant de manière idéalisée les processus au centre des systèmes complexes), et ce, aussi bien en météorologie que dans bien d'autres domaines.
Depuis le début des prévisions numériques du temps, la qualité des modèles numériques (permettant de propager les conditions initiales dans le temps) s'est accrue, et ce en partie grâce à la croissance de la puissance de calcul des ordinateurs. Cependant, sans condition initiale fiable, l'utilité de l'accroissement de la qualité de ces modèles (dans le but d'effectuer des prévisions) est discutable. Pour obtenir ces conditions initiales, il est nécessaire d'accéder à des informations sur l'état actuel de l'atmosphère. Pour ce faire, la quantité d'information récoltée par le réseau d'observation mondiale (Global Observing System) s'est accrue au fil des années.

Remarque 0.1

Il est bon de noter que la météorologie est sans doute un des domaines où la coopération internationale est la plus aboutie. En effet, via l'Organisation Météorologique Mondiale les données récoltées par chaque pays sont accessibles aux autres membres en temps réel. Il est ainsi possible pour chaque centre météorologique de disposer d'observations un peu partout sur Terre sans avoir à les effectuer eux mêmes.

Ce réseau comprend, entre autre, de nombreux satellites. L'information en provenance de ces derniers a crû de manière rapide depuis maintenant une vingtaine d'année. Aujourd'hui la majeure part des informations utilisées en météorologie provient des satellites malgré l'important réseau d'observation conventionnelle actuellement en place. En météorologie l'impact des observations satellites est particulièrement important au–dessus des régions peu observées par les moyens conventionnels. Ainsi, les satellites sont, de loin, la principale source d'information au niveau des pôles, au dessus de l'océan et, dans une moindre mesure, dans les pays du Sud [Desroziers et al., 2009]. Ces informations sont cruciales car elles contribuent à réduire l'écart de qualité entre les prévisions effectuées dans les pays du Nord et du Sud. En océanographie, science pauvre en observations conventionnelles, leur impact est encore plus grand : les satellites permettent d'accéder aux processus se déroulant à la surface de ce dernier de manière de plus en plus précise.

Dans ces deux domaines, afin d'effectuer les meilleures prévisions possibles, une bonne prise en compte de ces observations est primordiale dans le but d'estimer de manière optimale les conditions initiales du système. Ces images peuvent fournir de nombreuses informations sur des variables directement observées, comme la hauteur de surface de l'océan en océanographie, ou une information indirecte sur la dynamique du système. En observant des séquences d'images (comme par exemple celle présentées lors des prévisions météorologiques à la télévision), l'information sur la dynamique du système (ici le champ de vitesse) est aisément perçue par l'oeil humain. Cependant, cette information est actuellement sous–utilisée dans les systèmes opérationnels de prévision numérique :

— en météorologie, l'information est intégrée sous la forme d'observation indirecte sévèrement sous-échantillonnées en temps et en espace,
— en océanographie, elle n'est tout simplement pas (encore) utilisée.

Depuis maintenant quelques années, de nouvelles méthodes ont été proposées pour utiliser ce type d'information [Papadakis and Mémin, 2008, Korotaev et al., 2008, Corpetti et al., 2009,

Titaud et al., 2010, Michel, 2011]. Dans ce mémoire nous utilisons une méthode dans la continuité des travaux de [Souopgui, 2010, Titaud et al., 2010] réalisés au sein de l'équipe projet INRIA MOISE [2] afin d'extraire des séquences d'images les informations sur la dynamique du système.

Les séquences d'images satellites, bien que fournissant des données denses en espace ne sont pas sans poser certains problèmes.

Le premier problème consiste à savoir comment extraire à partir d'une séquence d'images des informations sur la dynamique du système.

Le second problème provient du moyen utilisé pour créer ces séquences d'images. En effet, les satellites mesurent des rayonnements électromagnétiques émis ou réfléchis par les objets étudiés (nuages, surface de l'eau, ...) dans un certain domaine de fréquences. Ceci est rendu possible par le fait que les objets étudiés émettent ou réfléchissent du rayonnement à différentes longueurs d'onde et intensités selon leur état. Pour avoir accès aux quantités physiques nous intéressant, il est nécessaire d'effectuer un pré–traitement de ces données. Ce pré–traitement tend à créer des corrélations au sein des champs d'erreurs d'observations, corrélations qui doivent être prise en compte lors de la phase d'assimilation de données.

En océanographie, les séquences d'images sont souvent en partie occultées par la présence de nuages. L'espace observé changeant d'un instant à l'autre il est dans ce cas bien plus difficile de prendre en compte de manière peu coûteuse les corrélations au sein des erreurs d'observations : les approximations effectuées sur la matrice de covariance lorsque l'ensemble de l'image est observé ne sont plus valables et doivent être adaptées.

Organisation du document

Ce travail de thèse est présenté en 6 chapitres. Le premier chapitre porte sur la méthode d'assimilation de données utilisée dans la suite du manuscrit. Le second chapitre présente la transformée en ondelettes orthogonales. Le troisième chapitre introduit le problème de l'assimilation d'images ainsi que la méthode employée dans ce manuscrit. Le quatrième porte sur la gestion d'un bruit d'observation décorrélé en espace et en temps. Ce chapitre est relativement court au vu du suivant : il est en effet peu envisageable d'avoir à décrire des erreurs décorrélées en temps et en espace pour des observations dont la densité spatiale et temporelle s'accroit continuellement.

Au cours du cinquième chapitre sont décrites différentes manières de prendre en compte un bruit corrélé en espace via l'utilisation des propriétés des bases d'ondelettes et des trames de curvelettes. Dans ce chapitre, l'intérêt d'appauvrir l'information (en travaillant par exemple dans l'espace des gradients) est aussi présenté.

Le dernier chapitre aborde le problème de la gestion des occultations lorsque les corrélations sont modélisées par une matrice diagonale dans une base d'ondelettes. Enfin, une dernière partie conclut et propose quelques perspectives à ces travaux.

2. Equipe où cette thèse a été préparée.

1

Introduction à l'assimilation de données

Sommaire

1.1 Notations 8
1.2 Formulation variationnelle du problème d'assimilation 10
1.3 Préconditonnement du problème 14
1.4 Influence de la matrice B 15
1.5 Construction de la matrice B 16

En novembre 1854, au cours de la Guerre de Crimée, une violente tempête provoqua le naufrage de 41 navires français en Mer Noire. Cette tempête avait, avant d'arriver en Crimée, traversée l'ensemble de l'Europe. Cependant, personne ne fut en mesure de la signaler. Pour prévenir ce genre d'évènement désagréable dans le futur, à la demande de Napoléon III, Urbain le Verrier mit en place un réseau d'observatoires météorologiques en France (24 observatoires) qui fut étendu par la suite à l'ensemble de l'Europe (59 observatoires). A partir des données récoltées par ces observatoires, les premières prévisions météorologiques en France eurent lieu dès 1863 via l'étude de l'évolution des phénomènes observés sur des cartes.
Ces premières prévisions étaient à l'époque destinées aux marins pour les prévenir des tempêtes arrivant. Elles étaient effectuées via l'utilisation d'**observations** et par *"déduction"* de la trajectoire de l'atmosphère.

Aujourd'hui et ce depuis l'apparition des premiers calculateurs, de nombreuses choses ont changé en ce qui concerne la prévision du temps. Actuellement, on parle de prévision *numérique* du temps, le terme numérique signifiant que la majorité du travail de prévision est effectuée par des ordinateurs compilant des informations en provenance de sources hétérogènes. Ces informations comprennent toujours les **observations** du système (qui sont un rappel vers la réalité qui est en train de se dérouler au dessus de notre tête) et un a priori sur l'état de l'atmosphère (nommée **ébauche**). Ce qui a profondément changé depuis l'arrivée des premiers calculateurs est l'utilisation de **modèles numériques** d'évolution intégrant des équations mathématiques décrivant la physique des processus étudiés.

Les calculateurs exécutant simplement les calculs demandés par l'utilisateur, un domaine de recherche important consiste à savoir comment leur demander d'extraire et de combiner (de manière calculatoire) l'information en provenance de différentes sources que nous ne sommes pas capable d'analyser sans utiliser leurs capacités de calcul.
Ces méthodes, en météorologie et océanographie, sont nommées méthodes *d'assimilation de données*. Elles permettent de combiner d'une manière optimale (dans un sens à définir) les différentes sources d'informations à notre disposition.

Dans quel but peut–on utiliser l'assimilation de données ?

Les méthodes d'assimilations de données peuvent être utilisées dans différents buts.

1. **La prévision :** Les informations passées et présentes sont utilisées au mieux afin de fournir la meilleure prévision possible au travers de la définition d'une condition initiale optimale propagée par un modèle.
2. **La réanalyse :** Les observations passées sont utilisées afin de construire la meilleure trajectoire du système étudié au cours du temps. Elles permettent d'étudier à posteriori un phénomène particulier ou de mieux comprendre l'évolution d'un système sur de longues périodes de temps. Ces réanalyses sont l'occasion d'utiliser des nouveaux systèmes de prévisions sur des cas anciens afin d'évaluer leur impact (et notamment savoir si à l'heure actuelle certains événements sont mieux représentés et prévus par les modèles).

   Ces deux premiers usages de l'assimilation de données sont les usages "traditionnels" de l'assimilation de données. Un autre usage, traditionnellement moins répandu, est actuellement plus fréquemment utilisé.
3. **La calibration des modèles :** Les informations passées et présentes peuvent en effet être utilisées pour mieux comprendre et régler les paramètres utilisés dans les modèles numériques. L'utilisation des techniques d'assimilations peut en effet permettre d'optimiser la calibration de certains *"paramètres"* du modèle en les confrontant à des observations.

Dans ce chapitre, nous introduisons le concept de l'assimilation variationelle de données. Bien que d'autres méthodes pour effectuer de l'assimilation de données existent et soient employées quotidiennement, nous ne les présentons pas ici. Le lecteur intéressé par une review des différentes méthodes de filtrage peut se tourner vers [Bonan, 2013].

Dans cette thèse, nous ne développons pas de nouvelle méthode d'assimilation de données. La littérature portant sur la description des différentes méthodes d'assimilation de données étant conséquente et bien détaillée, il a été choisit de décrire uniquement la méthode utilisée dans ces travaux. C'est pourquoi seul le cadre d'un **4D–Var** classique est introduit. Cette introduction pourrait paraître brève et peu détaillée pour le lecteur intéressé découvrant le sujet. Si tel est le cas, ce dernier est invité à consulter différentes notes de cours [Blayo and Nodet, 2012, Bocquet, 2011] ou travaux présentant bien plus en détails les différentes techniques d'assimilation de données [Vidard, 2001, Souopgui, 2010, Mirouze, 2010, Neveu, 2011, Bonan, 2013, Nodet, 2013].

1.1 Notations

Avant d'entrer dans les détails algorithmiques, nous allons commencer par introduire le vocabulaire et les notations employées dans ce chapitre.

Le vecteur d'état $\boldsymbol{X}$

Le vecteur d'état représente l'ensemble des variables pronostiques du modèle, c'est–à–dire les variables dont la dérivée par rapport au temps est prise en compte explicitement dans les équations du modèle. Il est noté $\boldsymbol{X}(t)$ et contient l'état des différentes variables du système étudié à l'instant t. Dans un système numérique, le vecteur d'état est une *représentation discrète* des variables continues du système. Sa dimension est notée n_x. L'état du système à l'instant t_k, i-e $\boldsymbol{X}(t_k)$, est noté $\boldsymbol{X}_k$.

L'état vrai $\boldsymbol{X}^t$

L'état vrai $\boldsymbol{X}^t$ représente l'état à estimer au moyen d'un processus d'assimilation. Cet état est par nature discret et ne peut donc représenter le véritable état continu $\boldsymbol{X}^c$. Il est cependant possible de relier $\boldsymbol{X}^c$ à $\boldsymbol{X}^t$ au moyen d'un opérateur $\boldsymbol{\Pi}$ permettant la projection de la variable continue sur la discrétisation employée. Nous avons ainsi :

$$\boldsymbol{X}_i^t = \boldsymbol{\Pi}\boldsymbol{X}_i^c. \tag{1.1}$$

Le modèle

Le modèle représente l'implémentation numérique des équations d'évolution du système. A partir d'un état initial $\boldsymbol{X}_0$, le modèle $\mathcal{M}$ permet de fournir un estimé de l'état du système $\boldsymbol{X}_k$.

$$\begin{cases} \dfrac{\partial \boldsymbol{X}(t)}{\partial t} = \mathcal{M}(\boldsymbol{X}(t)) \\ \boldsymbol{X}(t_0) = \boldsymbol{X}_0 \end{cases} \tag{1.2}$$

Cependant, le modèle n'est qu'une représentation partielle de la réalité des processus physiques. Ainsi, même si le véritable vecteur d'état, noté $\boldsymbol{X}^t$, est connu à l'instant t_k son intégration par le modèle jusqu'au temps t_{k+1} comporte des erreurs :

$$\boldsymbol{X}_{k+1}^t = \mathcal{M}_{t_k \to t_{k+1}}\left(\boldsymbol{X}_k^t\right) + \eta_k. \tag{1.3}$$

Le vecteur η_k est nommé erreur modèle. Cette erreur modèle dépend, entres autres, de l'état continu $\boldsymbol{X}_k^c$. Sans informations supplémentaires sur la nature de l'erreur modèle, cette dernière peut être modélisée comme une variable aléatoire gaussienne de moyenne :

$$\overline{\eta_k} = \mathrm{E}\left(\eta_k\right) \tag{1.4}$$

et de matrice de covariance :

$$\mathbf{Q}_k = \mathrm{E}\left((\eta_k - \overline{\eta_k})(\eta_k - \overline{\eta_k})^T\right). \tag{1.5}$$

Ces perturbations représentent l'ensemble des processus non modélisés. Ils englobent les processus non résolus par le modèle car de trop petite échelle.

L'ébauche $\boldsymbol{X}^b$

L'ébauche est une première estimation du vecteur d'état. Typiquement, il s'agit d'une sortie du modèle en provenance de la phase de prévision de la période précédente d'assimilation. L'ébauche est une connaissance a priori de l'état du système.

De la même manière que nous avons défini l'erreur modèle, nous pouvons définir une erreur d'ébauche :

$$\boldsymbol{X}^b = \boldsymbol{X}_0^t + \epsilon^b \tag{1.6}$$

Dans la suite, on suppose que ces erreurs suivent une loi gaussienne et ne sont pas biaisées :

$$\epsilon^b \sim \mathcal{N}(0, \mathbf{B}). \tag{1.7}$$

Les observations $\mathbf{y^o}$

Afin de pouvoir contrôler le système, il est nécessaire d'avoir des informations sur les processus qui sont en train d'avoir lieu. C'est le rôle des observations de fournir ce type d'information.

Ces observations, notées $\mathbf{y^o}$, proviennent de sources diverses. Elles ne sont pas nécessairement des observations des variables des modèles ou situées au point de grille de notre discrétisation.

Afin de relier les observations à un instant t_k, notée $\mathbf{y^o}_k$, au vecteur d'état $\boldsymbol{X}_k$ il est nécessaire de définir un opérateur d'observation discret $\mathcal{H}_k$ tel que :

$$\mathbf{y_k^o} = \mathcal{H}_k(\boldsymbol{X}_k^t) + \epsilon_k^o \tag{1.8}$$

où ϵ_k^o représente l'erreur d'observation au temps t_k. Les erreurs d'observations peuvent provenir de différentes sources dont les erreurs de mesure, les erreurs de pré-processing et les erreurs de représentativité. Ces dernières peuvent notamment être liées au fait que certains phénomènes observés ne sont pas résolus par le modèle.

Les erreurs d'observations sont supposées **décorrélées des erreurs d'ébauche** et non biaisées. Leur matrice de covariance est donnée par :

$$\mathbf{R}_i = \mathrm{E}\left(\epsilon_k^o \epsilon_k^{oT}\right) \tag{1.9}$$

1.2 Formulation variationnelle du problème d'assimilation

L'assimilation de données consiste à résoudre un problème inverse posé sur un système physique dont les paramètres d'entrée sont mal connus et dont les sorties ne sont que partiellement observées. Le principe des méthodes variationnelles est de résoudre ce problème par minimisation d'une fonctionnelle prenant en compte l'ensemble de l'information disponible.

Dans notre cas, le but est d'estimer la condition initiale $\boldsymbol{X}_0$ permettant d'expliquer la trajectoire observée de l'atmosphère, en prenant en compte nos connaissances sur la dynamique et un a priori sur la condition initiale. Cette condition initiale constitue notre **vecteur de contrôle**, c'est à dire l'ensemble des paramètres que nous allons chercher à estimer.

Remarque 1.1

Dans la suite de ce chapitre, pour décrire le concept de l'assimilation de données variationnelle, nous utilisons uniquement un modèle linéaire (noté $\boldsymbol{M}$) et un opérateur d'observation linéaire (noté $\mathbf{H}$) par soucis de simplicité. Toujours par soucis de simplicité, les statistiques d'erreurs d'observation et d'ébauche sont supposées gaussiennes tandis que les erreurs modèles sont négligées.

La fonctionnelle à minimiser s'écrit sous la forme :

$$\begin{aligned} J(\boldsymbol{X}_0) &= \frac{1}{2}(\boldsymbol{X}_0 - \boldsymbol{X}^b)^T \mathbf{B}^{-1}(\boldsymbol{X}_0 - \boldsymbol{X}^b) + \frac{1}{2}\sum_{i=0}^{N}(\mathbf{y_i^o} - \mathbf{H}_i(\boldsymbol{X}_i))^T \mathbf{R}_k^{-1}(\mathbf{y_i^o} - \mathbf{H}_i(\boldsymbol{X}_i)) \\ &= \underbrace{\|\boldsymbol{X}_0 - \boldsymbol{X}^b\|_{\mathbf{B}}^2}_{J^b} + \underbrace{\sum_{i=0}^{N} \|\mathbf{y_i^o} - \mathbf{H}_i(\boldsymbol{X}_i)\|_{\mathbf{R}_k}^2}_{J^o} \end{aligned} \tag{1.10}$$

Le minimum de cette fonctionnelle est nommé **état analysé** $\boldsymbol{X}^a$:

$$\boldsymbol{X}^a = \underset{\boldsymbol{X}_0}{\operatorname{argmin}}\, J(\boldsymbol{X}_0) \tag{1.11}$$

Cet état analysé peut être décomposé sous la forme de l'ébauche sur la condition initiale à laquelle un incrément a été ajouté :

$$\boldsymbol{X}^a = \boldsymbol{X}^b + \delta_{\boldsymbol{X}^a}. \tag{1.12}$$

Cet incrément $\delta_{\boldsymbol{X}^a}$ est nommé incrément d'analyse.

Le terme J^o est nommé terme d'écart aux observations. Il contient l'écart entre le vecteur d'observations $\mathbf{y^o}$ et l'état du modèle dans l'espace des observations. L'écart entre les observations et l'état du modèle est "pondéré" par l'inverse de la matrice de covariance d'erreur d'observation.

Le terme J^b est le terme d'écart à l'ébauche. Ce premier terme joue deux rôles :

— Il contient l'information a priori dont on dispose sur l'état du système ;
— Il permet de régulariser le problème mathématique à résoudre. Il y a en général trop peu d'observations pour que le problème sans ce terme d'écart à l'ébauche puisse être bien posé.

Remarque 1.2

Les observations à l'instant t_i sont liées au vecteur de contrôle $\boldsymbol{X}_0$ de manière très "indirecte". En effet, le lien entre l'état du système à l'instant t_i et l'état initial est fourni par le modèle d'évolution en temps :

$$\begin{aligned} \boldsymbol{X}_i &= \mathbf{M}_{(t_0 \to t_i)}(\boldsymbol{X}_0) \\ &= \mathbf{M}_i \mathbf{M}_{i-1} \cdots \mathbf{M}_1(\boldsymbol{X}_0) \end{aligned}$$

où on a noté $\mathbf{M}_i = \mathbf{M}_{(t_{i-1} \to t_i)}$.

Résolution par méthode adjointe

Au vu de la taille du vecteur de contrôle, minimiser directement la fonctionnelle (1.10) via des méthodes de descente de gradient est très coûteux lorsque le gradient est calculé par différences finies. L'utilisation des méthodes adjointes rend cependant ce calcul possible à un coût raisonnable [LeDimet, 1982, LeDimet and Talagrand, 1986] au prix du développement des modèles et opérateurs adjoints [3].

Le gradient de la fonction coût (1.10) introduite précédemment est donné par :

$$\begin{aligned} \nabla J(\boldsymbol{X}_0) &= \mathbf{B}^{-1}(\boldsymbol{X}_0 - \boldsymbol{X}^b) - \sum_{i=0}^{N} \mathbf{M}_1^T \cdots \mathbf{M}_i^T \mathbf{H}_i^T \mathbf{R}_i^{-1} \left[\mathbf{y_i^o} - \mathbf{H}_i \mathbf{M}_i \cdots \mathbf{M}_1 \boldsymbol{X}_0\right] \\ &= \nabla J^b(\boldsymbol{X}_0) + \nabla J^o(\boldsymbol{X}_0) \end{aligned} \tag{1.13}$$

La partie du gradient correspondant au rappel à l'ébauche s'obtient aisément (si on est capable d'inverser la matrice de covariance d'erreur d'ébauche). La seconde partie est cependant plus ardue à calculer. Elle fait intervenir le modèle et l'opérateur d'observation ainsi que leurs adjoints respectifs (notés $\mathbf{M}^T$ et $\mathbf{H}^T$). Cette seconde partie est, en assimilation variationelle, obtenue par méthode adjointe.

Afin de visualiser comment ces méthodes fonctionnent, nous pouvons réécrire $\nabla J^o(\boldsymbol{X}_0)$ en utilisant le vecteur d'innovation $\mathbf{d}_i$ défini par :

$$\mathbf{d}_i = \mathbf{y_i^o} - \mathbf{H}_i \mathbf{M}_i \cdots \mathbf{M}_1 \boldsymbol{X}_0. \tag{1.14}$$

qui correspond donc à la différence entre les observations et l'état du modèle à l'instant t_i.

3. A ce sujet, des dérivateurs automatiques tels TAPENADE développé dans l'équipe projet INRIA TROPICS [Hascoet and Pascual, 2013] existent pour faire une partie du travail de dérivation à notre place.

Avec cette nouvelle notation, la part du gradient imputable aux observations s'écrit :

$$\begin{aligned}
-\nabla J^o(\boldsymbol{X}_0) &= \sum_{i=0}^{N} \mathbf{M}_1^T \cdots \mathbf{M}_i^T \mathbf{H}_i^T \mathbf{R}_i^{-1} \left[\mathbf{y_i^o} - \mathbf{H}_i \mathbf{M}_i \cdots \mathbf{M}_1 \boldsymbol{X}_0\right] \\
&= \sum_{i=0}^{N} \mathbf{M}_1^T \cdots \mathbf{M}_i^T \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{d}_i \\
&= \mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{d}_0 + \mathbf{M}_1^T \mathbf{H}_1^T \mathbf{R}_1^{-1} \mathbf{d}_1 + \mathbf{M}_1^T \mathbf{M}_2^T \mathbf{H}_2^T \mathbf{R}_2^{-1} \mathbf{d}_2 + \cdots + \\
&\quad \mathbf{M}_1^T \cdots \mathbf{M}_N^T \mathbf{H}_N^T \mathbf{R}_N^{-1} \mathbf{d}_N \\
&= \mathbf{H}_0^T \mathbf{R}_0^{-1} \mathbf{d}_0 + \mathbf{M}_1^T \Big[\mathbf{H}_1^T \mathbf{R}_1^{-1} \mathbf{d}_1 + \mathbf{M}_2^T \Big[\mathbf{H}_2^T \mathbf{R}_2^{-1} \mathbf{d}_2 + \\
&\quad \mathbf{M}_3^T \left[\cdots + \mathbf{M}_N^T \mathbf{H}_N^T \mathbf{R}_N^{-1} \mathbf{d}_N\right]\Big]\Big]
\end{aligned}$$

En écrivant de cette manière le gradient du terme d'écart aux observations, on s'aperçoit qu'il est possible de l'obtenir par une intégration du modèle adjoint (en supposant qu'on ait déjà effectué une intégration directe pour obtenir les vecteurs d'innovations).

L'intégration du modèle adjoint fonctionne de la manière suivante :

$$\begin{cases} \hat{\boldsymbol{X}}_N = \mathbf{H}_N^T \mathbf{R}_N^{-1} \mathbf{d}_N \\ \hat{\boldsymbol{X}}_i = \mathbf{M}_{i+1}^T \hat{\boldsymbol{X}}_{i+1} + \mathbf{H}_i^T \mathbf{R}_i^{-1} \mathbf{d}_i \quad \text{pour } i = N-1, 0 \end{cases} \tag{1.15}$$

Par identification, en fin d'algorithme, on a donc $\nabla J^o(\boldsymbol{X}_0) = -\hat{\boldsymbol{X}}_0$.

Remarque 1.3

Remarquons que l'ensemble des états adjoints se calculent dans le sens contraire au sens du calcul des états directs du modèle.

Ce moyen d'obtenir le gradient de la fonction coût est à la base de l'application du $\mathbf{4D-Var}$ qui, à partir d'une connaissance à priori du vecteur d'état $\boldsymbol{X}^b$ peut calculer la fonctionnelle J, son gradient.

L'algorithme du $\mathbf{4D-Var}$ peut se résumer de la manière suivante :

Algorithme 4D − Var

1. Initialisation de l'algorithme : $\boldsymbol{X}_0 = \boldsymbol{X}^b$,
2. Intégration du modèle direct $\mathbf{M}$ en partant de $\boldsymbol{X}_0$. Cette intégration permet d'obtenir les vecteurs d'innovations $\mathbf{d}_i$.
3. Intégration du modèle adjoint. Cette intégration permet d'obtenir le gradient de la fonctionnelle J.
4. Appel à une méthode de descente permettant d'obtenir une nouvelle condition initiale $\boldsymbol{X}_0 = \boldsymbol{X}^b + \delta_{\boldsymbol{X}_0}$
5. Si le critère de minimisation est atteint, on s'arrête. Sinon, on reprend à la seconde étape.

La figure 1.1 résume le principe de l'assimilation variationnelle utilisant l'information sur la dynamique du modèle ($4D-Var$). Tout d'abord, l'ébauche est propagée par le modèle. L'assimilation variationnelle consiste alors à comparer l'ensemble des observations disponibles sur la fenêtre au vecteur d'état (à l'instant d'observation) pour corriger la condition initiale. Cela permet d'obtenir un nouvel état initial qui peut à son tour être propagé par le modèle puis corrigé

en tenant compte de la distance aux observations. Lorsqu'après un certain nombre d'itérations on arrive à convergence (ou on estime que la correction réalisée est suffisante), l'état obtenu est nommé état analysé (et noté $\boldsymbol{X}^a$). Cet état analysé nous permet d'effectuer au moyen d'un modèle (pas nécessairement le même) une prévision.

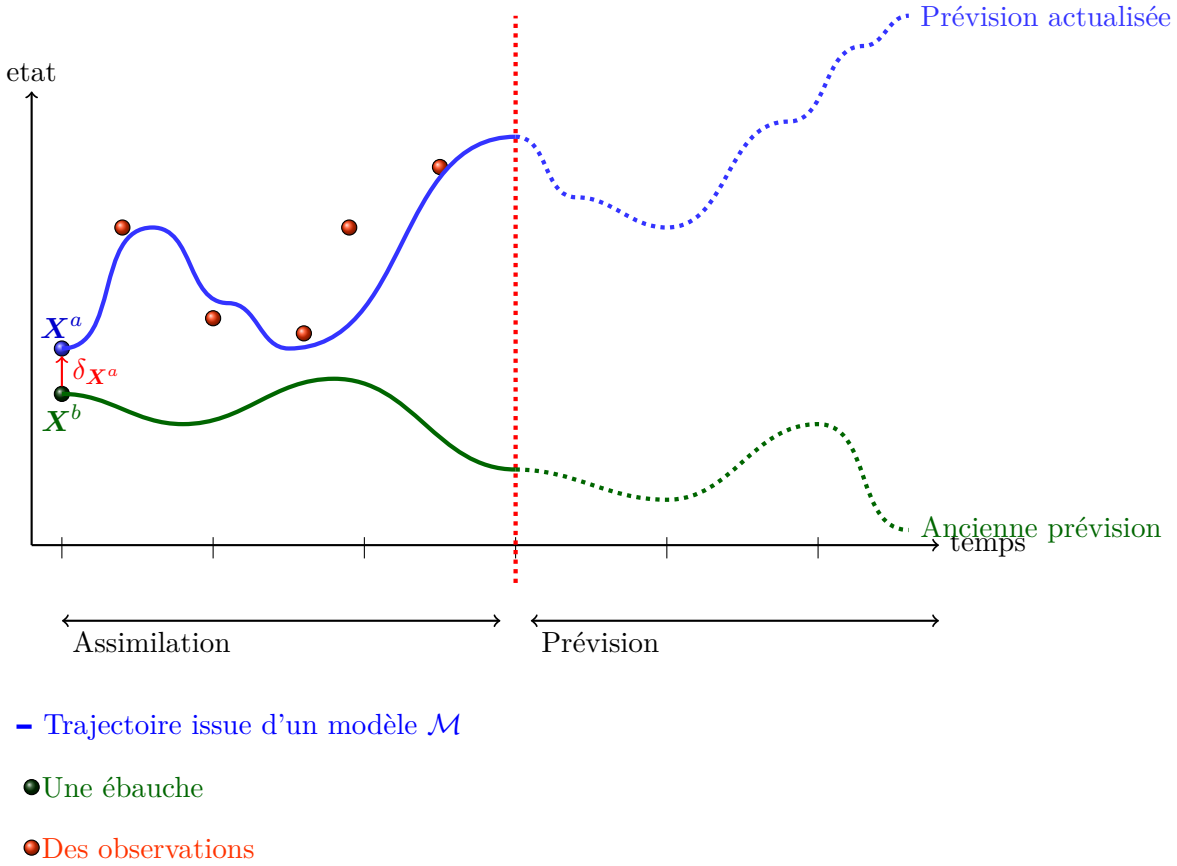


FIGURE 1.1 – Principe du $4D-Var$

Reformulation du problème

Afin de simplifier les notations employées dans la suite de cette section, celles–ci sont adaptées. Elles font apparaître un opérateur d'observation généralisé $\mathbf{G}$ contenant l'intégration par le modèle de l'état initial et l'opérateur d'observation :

$$\mathbf{y^o} = \begin{pmatrix} \mathbf{y_0^o} \\ \mathbf{y_1^o} \\ \vdots \\ \mathbf{y_N^o} \end{pmatrix} ; \quad \mathbf{M} = \begin{pmatrix} \mathbf{M}_{t_0 \to t_0} \\ \mathbf{M}_{t_0 \to t_1} \\ \vdots \\ \mathbf{M}_{t_0 \to t_N} \end{pmatrix} ; \quad \mathbf{R} = \begin{pmatrix} \mathbf{R}_0 & 0 & \cdots & 0 \\ 0 & \mathbf{R}_1 & 0 & 0 \\ \cdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{R}_N \end{pmatrix} ; \quad \mathbf{H} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1 \\ \vdots \\ \mathbf{H}_N \end{pmatrix}$$

$$\mathbf{G} = \mathbf{HM}$$

En utilisant les notations précédentes, la fonction côut (1.10) est donnée par :

$$J(\boldsymbol{X}_0) = \frac{1}{2}\left(\boldsymbol{X}_0 - \boldsymbol{X}^b\right)^T \mathbf{B}^{-1}\left(\boldsymbol{X}_0 - \boldsymbol{X}^b\right) + \frac{1}{2}\left(\mathbf{y^o} - \mathbf{G}\boldsymbol{X}_0\right)^T \mathbf{R}^{-1}\left(\mathbf{y^o} - \mathbf{G}\boldsymbol{X}_0\right) \tag{1.16}$$

En utilisant l'incrément par rapport à l'ébauche

$$\delta_{\boldsymbol{X}_0} = \boldsymbol{X}_0 - \boldsymbol{X}^b$$

et le fait que le modèle et l'opérateur d'observations sont linéaires :

$$\mathbf{y^o} - \mathbf{G}\boldsymbol{X}_0 = \mathbf{y^o} - \mathbf{G}\boldsymbol{X}^b - \mathbf{G}\delta_{\boldsymbol{X}_0} = \mathbf{d} - \mathbf{G}\delta_{\boldsymbol{X}_0}$$

cette fonction coût peut se réécrire sous la forme :

$$J(\delta_{\boldsymbol{X}_0}) = \frac{1}{2}(\delta_{\boldsymbol{X}_0})^T \mathbf{B}^{-1}(\delta_{\boldsymbol{X}_0}) + \frac{1}{2}(\mathbf{d} - \mathbf{G}\delta_{\boldsymbol{X}_0})^T \mathbf{R}^{-1}(\mathbf{d} - \mathbf{G}\delta_{\boldsymbol{X}_0}). \tag{1.17}$$

1.3 Préconditonnement du problème

Une itération de minimisation fait intervenir une intégration du modèle direct et une intégration du modèle adjoint. Chacune de ces intégrations a un coût. En pratique, réduire au maximum le nombre d'itérations de minimisation afin de résoudre le problème est particulièrement important. Dans les faits, l'algorithme est bien souvent arrêté avant de satisfaire un critère de convergence (norme du gradient, valeur de la fonction coût). Le critère d'arrêt étant alors le nombre d'itérations de minimisation que l'on peut se permettre d'effectuer. Il est donc important d'avoir accès à un algorithme convergent rapidement.

Hessienne

Une des sources affectant l'efficacité de la minimisation est le conditionnement de la Hessienne. Ce conditionnnement est défini par le ratio entre la plus grande et la plus petite valeur propre. Plus ce conditionnement est grand, plus le problème est mal conditionné.

La matrice hessienne du poblème (1.17) est donnée par :

$$\mathbf{Hess} = \mathbf{B}^{-1} + \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} \tag{1.18}$$

Afin de réduire le conditionnement d'un système, il est possible de minimiser un problème équivalent, mais mieux conditionné, en effectuant un changement de variable.

Changement de variable

Le changement de variable idéal pour bien préconditionner le système est le suivant :

$$v = (\mathbf{Hess})^{1/2} \delta_{\boldsymbol{X}_0} \tag{1.19}$$

Cependant, n'ayant généralement pas accès à la Hessienne, ce changement de variable optimal ne peut être envisagé.

Un préconditionnement classique est celui proposé par [Courtier, 1997] consistant à utiliser la matrice de covariance d'erreur d'ébauche. Le changement de variable effectué est :

$$v = \mathbf{B}^{-1/2} \delta_{\boldsymbol{X}_0} \Leftrightarrow \delta_{\boldsymbol{X}_0} = \mathbf{B}^{1/2} v \tag{1.20}$$

Le système résolu est bien entendu le même que précédemment en écrivant la fonction coût (1.17) sous la forme :

$$J_v(v) = \frac{1}{2}[\mathbf{B}^{1/2} v]^T \mathbf{B}^{-1} [\mathbf{B}^{1/2} v] + \frac{1}{2}(\mathbf{d} - \mathbf{G}\mathbf{B}^{1/2} v)^T \mathbf{R}^{-1} (\mathbf{d} - \mathbf{G}\mathbf{B}^{1/2} v) \tag{1.21}$$

Le gradient de la fonction coût par rapport au vecteur de contrôle v s'écrit alors sous la forme :

$$\nabla_v J(v) = \mathbf{B}^{T/2} \mathbf{B}^{-1} \mathbf{B}^{1/2} v - \mathbf{B}^{T/2} \mathbf{G}^T \mathbf{R}^{-1} (\mathbf{d} - \mathbf{G}\mathbf{B}^{1/2} v) \tag{1.22}$$

$$= \mathbf{B}^{T/2} \nabla J(\delta_{\boldsymbol{X}_0}) \tag{1.23}$$

tandis que sa Hessienne est donnée par :

$$\begin{aligned} \mathbf{Hess}_v &= \mathbf{B}^{T/2} \mathbf{B}^{-1} \mathbf{B}^{1/2} + \mathbf{B}^{T/2} \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} \mathbf{B}^{1/2} \\ &= \mathbf{I} + \mathbf{B}^{T/2} \mathbf{G}^T \mathbf{R}^{-1} \mathbf{G} \mathbf{B}^{1/2} \end{aligned}$$

où la matrice $\mathbf{B}^{T/2}\mathbf{G}^T\mathbf{R}^{-1}\mathbf{G}\mathbf{B}^{1/2}$ est une matrice symétrique semi-définie positive.

En effectuant ce préconditionnement, on est donc assuré que la plus petite valeur propre du système soit supérieur à 1. Lorsque l'espace d'observation est de petite taille (par rapport à l'espace d'état), le système possède de nombreuses valeurs propres valant 1 car une grande partie des valeurs propres de $\mathbf{B}^{T/2}\mathbf{G}^T\mathbf{R}^{-1}\mathbf{G}\mathbf{B}^{1/2}$ sont nulles.

Algorithme du 4D – Var préconditionné par $\mathbf{B}^{1/2}$

- Initialisation : $v = 0$
  1. Passage dans l'espace physique : $\boldsymbol{X}_0 = \boldsymbol{X}^b + \mathbf{B}^{1/2}v$
  2. Intégration du modèle directe
  3. Intégration du modèle adjoint. On obtient ainsi $\nabla J(\delta_{\boldsymbol{X}_0})$
  4. Retour dans l'espace de contrôle : $\nabla J_v(v) = \mathbf{B}^{T/2}\nabla J(\delta_{\boldsymbol{X}_0})$, $J_v(v) = J(\delta_{\boldsymbol{X}_0})$
  5. Mise à jour du vecteur de contrôle v par la méthode de minimisation choisie,
  6. Retour à 1 si le critère d'arrêt n'est pas respecté.

Remarque 1.4

On observe que lorsque le préconditionnement par $\mathbf{B}^{-1/2}$ est effectué, nous avons "seulement" à manipuler $\mathbf{B}^{1/2}$ et $\mathbf{B}^{T/2}$. La matrice $\mathbf{B}$ ne doit plus être inversée. Cependant, il est nécessaire d'avoir accès à la racine carrée de la matrice de covariance d'erreur d'ébauche ce qui peut s'avérer complexe.

Remarque 1.5

D'autres changement de variables, comme $v = \mathbf{B}^{-1}\delta_{\boldsymbol{X}_0}$, peuvent être effectuées pour préconditionner le problème. Le lecteur intéressé par le sujet est invité à aller consulter la thèse de [Mirouze, 2010] pour plus de détails sur ces changements de variables.

1.4 Influence de la matrice B

Quand la fonction coût est convexe, l'optimum est obtenu lorsque $\nabla J(\delta_{\boldsymbol{X}^a}) = 0$ où le gradient calculé par la formule (1.13) et $\delta_{\boldsymbol{X}^a}$ est l'incrément d'analyse.
La valeur de l'incrément d'analyse est donc :

$$\begin{aligned} & \mathbf{B}^{-1}\delta_{\boldsymbol{X}^a} + \mathbf{G}^T\mathbf{R}^{-1}(\mathbf{G}\delta_{\boldsymbol{X}^a} - \mathbf{d}) = 0 \\ \Leftrightarrow \quad & \delta_{\boldsymbol{X}^a} = \left[\mathbf{B}^{-1} + \mathbf{G}^T\mathbf{R}^{-1}\mathbf{G}\right]^{-1}\mathbf{G}^T\mathbf{R}^{-1}\mathbf{d} \end{aligned}$$

En utilisant la formule de Sherman–Morrison–Woddbury :

$$\left[\mathbf{A}^{-1} + \mathbf{K}^T\mathbf{P}^{-1}\mathbf{K}\right]^{-1}\mathbf{K}^T\mathbf{P}^{-1} = \mathbf{A}\mathbf{K}^T\left[\mathbf{K}\mathbf{A}\mathbf{K}^T + \mathbf{P}\right]^{-1} \tag{1.24}$$

on obtient

$$\delta_{\boldsymbol{X}^a} = \mathbf{B}\mathbf{G}^T\left[\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R}\right]^{-1}\mathbf{d}. \tag{1.25}$$

Le terme $\left[\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R}\right]^{-1}\mathbf{d}$ correspond à une correction du vecteur de contrôle dans l'espace des observations. Si seul ce terme était présent, l'incrément d'analyse ne serait corrigé qu'aux endroits observés. Le terme $\mathbf{G}^T$, comprenant à la fois l'opérateur d'observation adjoint et le modèle rétrograde, permet de diffuser l'information via le modèle rétrograde aux endroits non observés. La matrice $\mathbf{B}$ va de nouveau diffuser cette information en espace aux différentes variables de l'état initial. Par le biais de l'intégration rétrograde et de l'application de la matrice de covariance d'erreur d'ébauche, l'information fournie par les observations est donc transmise vers

des variables non modifiées jusqu'à présent. Dans les zones densément observées, les corrélations présentes au sein de **B** permettent aussi de lisser l'état final. Définir de manière adéquate cette matrice est donc essentiel au bon déroulement de l'algorithme.

1.5 Construction de la matrice B

Les dimensions de la matrice de covariances d'erreur d'ébauche étant très grandes, cette matrice n'est jamais stockée ou définie sous forme matricielle. Comme en pratique nous avons seulement besoin de la multiplication matrice vecteur, ces matrices sont souvent stockées sous forme d'opérateurs permettant d'effectuer la multiplication entre une matrice et un vecteur pour un faible coût de calcul. Il existe différentes manières de construire la matrice de corrélation d'erreur d'ébauche. Cette dernière peut être spécifiée par l'utilisateur en se basant sur une connaissance "a priori" ou provenir d'une méthode d'ensemble.
Dans ces travaux, la matrice de covariance d'erreur d'ébauche est construite en utilisant l'opérateur de diffusion généralisé [Weaver and Courtier, 2001].

Décomposition de la matrice de covariance d'erreur d'ébauche
Pour décrire la construction de la matrice **B**, nous allons nous placer dans le cas du modèle Shallow–Water. Le vecteur d'état est ici composé de 3 variables $\begin{pmatrix} u, & v, & h \end{pmatrix}$. La matrice de covariance d'erreur d'ébauche peut s'écrire sous la forme :

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{uu} & \mathbf{B}_{uv} & \mathbf{B}_{uh} \\ \mathbf{B}_{vu} & \mathbf{B}_{vv} & \mathbf{B}_{vh} \\ \mathbf{B}_{hu} & \mathbf{B}_{hv} & \mathbf{B}_{hh} \end{pmatrix} \tag{1.26}$$

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

Classiquement la matrice de covariance d'erreur d'ébauche **B** peut être décomposée sous la forme :

$$\mathbf{B} = \mathbf{K}_b \Sigma \mathbf{C} \Sigma \mathbf{K}_b^T \tag{1.27}$$

où

— Σ est une matrice diagonale contenant les écarts types,
— **C** une matrice de corrélation d'erreur d'ébauche diagonale par bloc :

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{uu} & 0 & 0 \\ 0 & \mathbf{C}_{vv} & 0 \\ 0 & 0 & \mathbf{C}_{hh} \end{pmatrix} \tag{1.28}$$

— $\mathbf{K}_b$ est un opérateur d'équilibre qui permet d'imposer à l'incrément de satisfaire certaines propriétés physiques connues.

La matrice Σ, au vu de sa taille, peut être manipulée explicitement. Ce n'est pas le cas de la matrice de corrélation **C** et l'opérateur d'équilibre $\mathbf{K}_b$ qui sont généralement de trop grande dimension pour être manipulables. Ce sont ces matrice qui souvent codées sous forme d'opérateurs.

Modélisation des corrélations monovariées
L'approche classique consiste à définir les fonctions de corrélations spatiales de chaque variable comme des gaussiennes dépendant de la distance entre deux points :

$$\mathbf{C}_{i,j} = \exp\left(-\frac{d_{i,j}}{c_1}\right) \tag{1.29}$$

où $d_{i,j}$ est la distance entre les points de grilles i et j et c_1 un facteur de normalisation. En définissant les corrélations de cette manière, l'influence du point i sur j décroit avec la distance entre ces points.

Dans ce cas, les corrélations d'erreur d'ébauche sont modélisées par la convolution avec une gaussienne. La figure 1.2 présente l'effet de la convolution d'un dirac avec une gaussienne.

L'exemple précédent fournit l'incrément d'analyse obtenu si une seule observation (située au temps initial) est utilisée. L'incrément est donné par :

$$\delta_{\boldsymbol{X}^a} = \mathbf{B}\mathbf{G}^T \left[\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R}\right]^{-1} \mathbf{d}. \tag{1.30}$$

où $\mathbf{B} = \mathbf{C}$, $\mathbf{G} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}$ et $\mathbf{R} = \sigma_o^2$.
Il se réécrit donc :

$$\delta_{\boldsymbol{X}^a} = \mathbf{C} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \dfrac{\mathbf{d}}{\sigma_o^2 + 1} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{1.31}$$

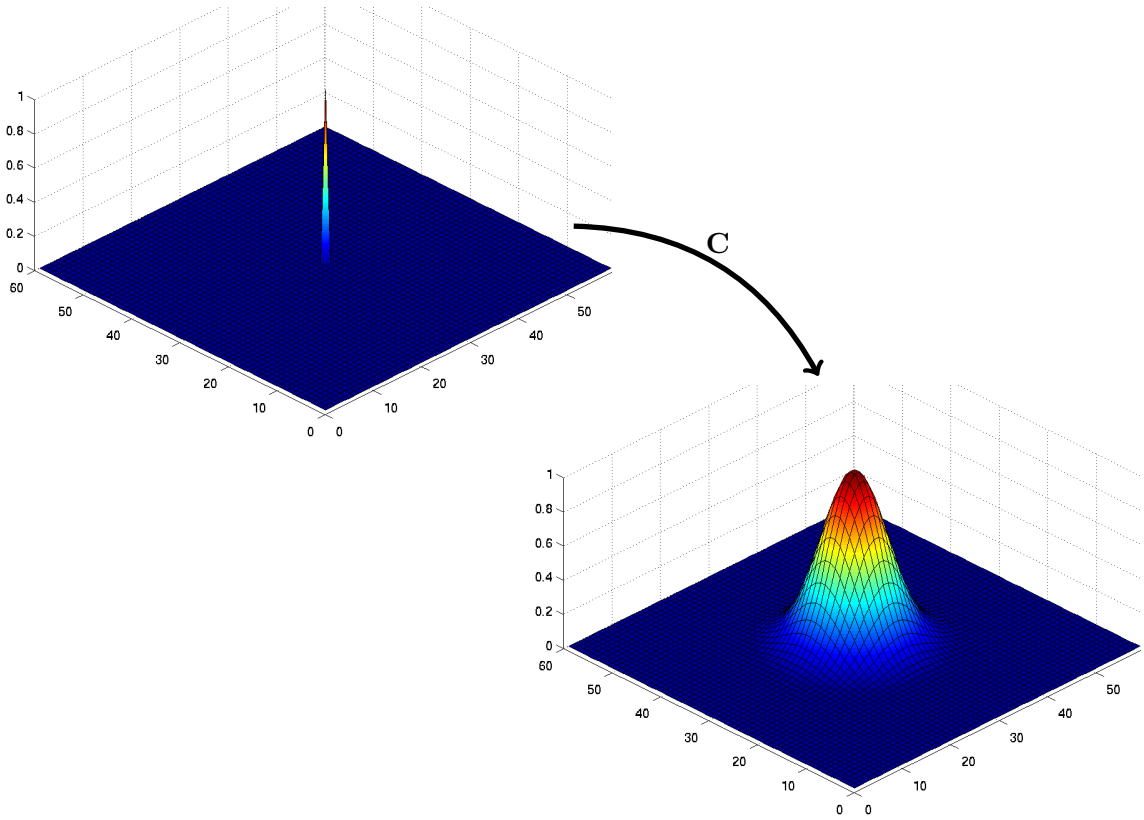


FIGURE 1.2 – Effet de l'application de la convolution d'un dirac avec une gaussienne.

La méthode de l'opérateur de diffusion généralisée [Weaver and Courtier, 2001] améliore l'idée précédente en déformant la structure gaussienne qui peut suivre la côte ou le flot. Cette méthode est utilisable car nous avons généralement uniquement besoin d'un produit matrice–vecteur. Le caclul de ce produit peut–être effectué au moyen de l'intégration d'une équation de

diffusion (dont la solution est, à un facteur près, la convolution d'un vecteur initial par un noyau gaussien).
En effet, si on prend l'équation de diffusion $1D$:

$$\frac{\partial \eta}{\partial t} - \kappa \frac{\partial^2 \eta}{\partial z^2} = 0 \tag{1.32}$$

où $\eta = \eta(z,t)$ est un champ scalaire et κ un coefficient de diffusion, la solution à la fin d'un intervalle de pseudo-temps $0 \leq t \leq T$ est donné par :

$$\eta(z,T) = \frac{1}{\sqrt{4\pi\kappa T}} \int_{z'} e^{-(z-z')^2/4\kappa T} \eta(z',0)\,\mathrm{d}z' = \frac{1}{\lambda}(g \star \eta(z,0)). \tag{1.33}$$

$\eta(z,T)$ est ainsi le résultat de la convolution de $\eta(z,0)$ par une fonction de corrélation gaussienne à un facteur de renormalisation λ près. L'échelle de la corrélation gaussienne est donné par $L = \sqrt{2\kappa T}$.

Remarque 1.6

Lorsque le coeffiicent de diffusion κ dépend des coordonées spatiales, il est possible de faire varier géographiquement les échelles de corrélations.

C'est cette approche qui est utilisée dans ces travaux pour modéliser les corrélations monovariées.

Modélisation des corrélations mutlivariées

Jusqu'à présent, seules les corrélations entre variables de même type sont modélisées. Lorsqu'une seule observation (sur une variable) est utilisée cela permet de corriger uniquement ce type de variable. Il est cependant possible d'intégrer de corrélations entre les différentes variables. Ces corrélations peuvent par exemple forcer le fait que l'incrément d'analyse satisfasse certaines propriétés physiques connues (afin d'obtenir un incrément équilibré). Cependant, dans ces travaux nous n'utilisons pas d'opérateur d'équilibre.
Nous reviendrons sur ce fait lors de la définition du contexte expérimental en section 3.3.3 (via une explication fourni en annexe A).

2

Transformée en ondelettes discrète

Sommaire

2.1 Introduction **19**
2.2 Transformée en ondelettes discrète **21**
2.2.1 Analyse multi-résolution 21
2.2.2 Le rôle des filtres 24
2.2.3 Propriétés des bases d'ondelettes 29
2.2.4 Opérations de seuillage 30
2.3 Ondelettes séparables en dimension 2 . . . . . **32**
2.3.1 Construction 33
2.3.2 Obtention des coefficients 33
2.3.3 Illustration sur un exemple du stockage de l'information dans un espace d'ondelettes 35
2.3.4 Autre représentation en dimension 2 38

2.1 Introduction

Les analyses multi-échelles sont très populaires dans la communauté des mathématiques appliquées et particulièrement dans le domaine du traitement du signal et des images. Parmis ces dernières, les transformées en ondelettes ont été introduites et popularisées dans les années 80 par les travaux (entre autres) de [Morlet and Grossman, 1984], [Daubechies, 1988] et [Mallat, 1989] et ce bien que la première ondelette soit apparue [4] dans les travaux d'Alfred Haar en 1910.

Ces transformées ont aujourd'hui de nombreuses applications notamment en traitement du signal et en statistiques. Ces applications comprennent entre autre la compression d'images (par exemple la norme jpeg 2000 utilise les ondelettes), le débruitage de signal ou d'images (aussi bien pour un bruit additif [Donoho and Johnstone, 1994] que multiplicatif [Durand et al., 2010]), la segmentation d'images (comme par exemple la création d'un détecteur de Canny multi-échelles [Le Cadet, 2004]), la reconstruction d'un signal partiellement observé [Elad et al., 2005] et plus généralement dans toutes les disciplines liées au traitement du signal.

Qu'est ce qu'une transformée en ondelettes ?

D'un point de vue "terre à terre", la transformation d'un signal dans une base d'ondelettes (orthonormale ou non) constitue uniquement un changement de base. Ainsi, n'importe quel signal $f \in L^2(\mathbb{R})$ peut s'écrire sous la forme :

$$f(x) = \sum_{j,\mathbf{k},l} c_{j,\mathbf{k},l}\varphi_{j,\mathbf{k},l}(x) \tag{2.1}$$

4. Mais baptisée ondelette bien plus tard

lorsque la famille $\{\varphi_{j,\mathbf{k},l}\}_{j,\mathbf{k},l}$ forme une base de $L^2(\mathbb{R})$.

La particularité des représentations multi-échelles réside dans la définition des éléments de ces bases. Ceux-ci *capturent* les propriétés du signal à des *échelles* j différentes.
Chaque fonction $\varphi_{j,\mathbf{k},l}$ posséde, en plusieurs dimensions, une *orientation* de capture l privilégiée et est *localisée* en espace (en fonction de $\mathbf{k}$, un vecteur de position).

Pourquoi effectuer une transformation du signal ?

Ces transformations ont pour intêret de faciliter l'analyse du signal : l'information est représentée dans un espace structuré et non plus au niveau du pixel. De nombreuses utilisations des transformées multi-échelles, en traitement du signal, sont basées (au moins implicitement) sur l'hypothèse que certains types de signaux peuvent s'écrire de manière parcimonieuse dans des espaces de $X - let$ [5].

Cette hypothèse constitue un *a priori* lors de la recherche d'un signal original perturbé par un bruit ou dont une partie a été "perdue" (ou non observée). Lorsque cette hypothèse est valide, pour débruiter une image il est possible de se restreindre à travailler uniquement avec un petit nombre de coefficients portant les structures principales du signal. De ce fait, on considère que les autres coefficients sont trop perturbés par le bruit pour être d'un quelconque intérêt.

Quel intérêt en assimilation de données ?

Ce type de transformée permet de pouvoir écrire certains signaux de manière parcimonieuse dans une base d'ondelettes. Cette parcimonie peut être employée pour deux choses différentes : la compression de l'information (comme par exemple effectué par la norme jpeg2000) et la représentation des erreurs (voir par exemple [Deckmyn and Berre, 2005] [Pannekoucke et al., 2007] [Pannekoucke, 2009] [Pannekoucke et al., 2013]).

La compression d'information

La première raison réside dans le fait que la quantité de données disponibles croît de manière rapide notamment concernant les données acquises par des satellites observant la surface et l'atmosphère terrestre. A l'heure actuelle, pour la prévision météorologique, le nombre d'observations utilisées au cours d'un cycle d'assimilation est de l'ordre 10^7. Pour comparaison, la taille d'une seule image en provenance du satellite METEOSAT est de l'ordre de 2.5 10^6 pixels. Ce satellite fournit différentes images toutes les 15 minutes. Par conséquent, sur une fenêtre d'assimilation dont la durée en météorologie est de l'ordre de quelques heures, utiliser l'ensemble des données acquises par cet unique satellite reviendrait à multiplier la taille du vecteur d'observation par quelques dizaines. Il est donc nécessaire de compresser l'information présente dans ce type d'observations ou d'en extraire uniquement une partie.
Une première idée est donc d'utiliser l'écriture parcimonieuse des images dans des bases d'ondelettes pour compresser l'information utile sur un jeu de coefficients de taille réduite. Ce point sera principalement développé dans les chapitres 3 et 4.

La représentation des erreurs

Un autre intêret de travailler dans un espace structuré réside dans la possibilité de décrire les erreurs d'une manière différente. Ces dernières ne sont plus décrites au niveau du pixel mais au niveau de la fonction de base (ici d'ondelettes). Par exemple, un bruit corrélé en espace n'affecte pas de manière identique les coefficients d'ondelettes aux différentes échelles [Johnstone and Silverman, 1997]. *Dès lors en raisonnant en terme de perturbation sur les structures et non plus au niveau du pixel, on peut extraire (ou utiliser) plus aisément une partie de*

5. Terme popularisé par [Do, 2001] désignant une transformation multi-échelles quelconque

l'information portant sur les statistiques du bruit. Nous reviendrons sur ce point au chapitre 5 et au chapitre 6.

Organisation du chapitre

Dans une première partie, la transformée en ondelette orthonormale discrète en $1D$ est introduite. Par la suite, cette transfomée est étendue en plusieurs dimensions. Le rôle de ce chapitre est de fournir une présentation sommaire des transformées multi-échelles en présentant brièvement la théorie sous-jacente illustrée par des exemples afin d'entrevoir l'intérêt de ces représentations. Le lecteur souhaitant plus de détails sur ces transformées peut se tourner vers les ouvrages de [Daubechies et al., 1992] [Truchetet, 1998] et [Mallat, 2008].

2.2 Transformée en ondelettes discrète

2.2.1 Analyse multi-résolution

Les ondelettes orthogonales sont fortement liées à la recherche d'une approximation multi-résolution d'un signal.
Une analyse à la résolution j d'un signal f (dans $L^2(\mathbb{R})$) est obtenue par l'application d'un opérateur linéaire A_j sur f de telle sorte que :

$$A_j f \in V_j \tag{2.2}$$

où V_j est un sous-espace de $L^2(\mathbb{R})$, et A_j est un projecteur sur ce sous-espace.

Une analyse multi-résolution consiste à construire une succession d'analyses dans des espaces $\{V_j : j \in \mathbb{Z}\}$ emboîtés qui *approchent* $L^2(\mathbb{R})$. Le passage d'une résolution j à une autre peut alors être compris comme un changement d'échelle de l'analyse.

Remarque 2.1 (*Convention sur les indices d'échelles*)

On considère qu'on incorpore plus de détails dans l'analyse lorsque j augmente (la précision du signal augmente) ou qu'on prend du recul pour mieux visualiser les structures principales lorsque j diminue.

La succession d'emboîtements choisie est la suivante :

$$\cdots \subset V_0 \subset V_1 \subset \cdots \subset V_j \subset V_{j+1} \subset \cdots \tag{2.3}$$

La définition d'analyse multi-résolution de [Mallat, 1989] est utilisée :

Définition 2.1

Une analyse multi-résolution de $L^2(\mathbb{R})$ met en jeu une succession d'espaces emboîtés $\{V_j\}_{j\in\mathbb{Z}}$ tels que :

$$\begin{array}{llll}
\forall j \in \mathbb{Z}, & V_{j-1} \subset V_j & \subset & \ldots \subset L^2(\mathbb{R}) \\
& \bigcap_{j\in\mathbb{Z}} V_j & = & \{0\} \\
& \overline{\bigcup_{j\in\mathbb{Z}} V_j} & = & L^2(\mathbb{R}) \\
& f(x) \in V_j & \Leftrightarrow & f(2x) \in V_{j+1} \qquad (2.4) \\
& f(x) \in V_0 & \Leftrightarrow & \forall n \in \mathbb{Z}, f(x-n) \in V_0 \qquad (2.5) \\
& \exists \phi \in V_0 & \textit{telle que} & \{\phi(x-n) : n \in \mathbb{Z}\} \textit{ soit une base orthonormée de } V_0.
\end{array}$$

où ϕ, fonction génératrice de V_0, est appelée ***fonction d'échelle*** *de l'analyse multi-résolution.*

Des exemples de fonctions d'échelle sont présentés en figure 2.1. La multi-résolution est introduite par l'équation (2.4). Ainsi, l'espace V_j définit une représentation deux fois plus grossière que la représentation définie par V_{j+1}. Chacune des fonction d'échelles présentée engendre une analyse multi-résolution différente.

Remarque 2.2

Des transformées en ondelettes ont été proposées pour changer le facteur 2 entre la résolution de l'espace V_j et l'espace V_{j+1}. Elles sont nommées transformées en ondelettes M–bandes où M est le facteur de changement de résolution. Voir par exemple la thèse de [Chaux, 2006] pour une étude poussée de ces transformées et de leurs avantages sur les transformées en ondelettes classiques.

Lorsque la famille $\{\phi(t-n)\}_{n\in\mathbb{Z}}$ définit une base orthogonale de V_0, les propriétés (2.4) et (2.5) permettent de prouver que la famille

$$\{\phi_{j,n}\}_{n\in\mathbb{Z}} = \left\{2^{j/2}\phi(2^j t - n)\right\}_{n\in\mathbb{Z}} \tag{2.6}$$

définit une base orthonormale de V_j pour tout $j \in \mathbb{Z}$.

Lors du passage de la résolution $j+1$ à une résolution j plus *grossière* du signal, certains *détails* spécifiques à la résolution $j+1$ sont perdus. Le principe de la transformée en ondelettes est de *stocker* ces détails dans un espace noté W_j.

Définition 2.2

L'espace W_j est le complémentaire orthogonal de V_j dans V_{j+1} :

$$V_{j+1} = W_j \oplus V_j \quad \text{avec } V_j \perp W_j \tag{2.7}$$

L'espace V_j est l'*espace d'approximation* (à l'échelle j) tandis que l'espace W_j est l'*espace de détail*.

Les propriétés suivantes sur les espaces de détails W_j peuvent être déduites des définitions précédentes :

Définition 2.3

$$\begin{array}{rcl} \forall j, j' \in \mathbb{Z} j \neq j' & \Rightarrow & W_j \perp W_{j'}, \\ \bigoplus_{j\in\mathbb{Z}} W_j & = & L^2(\mathbb{R}), \\ f(x) \in W_j & \Leftrightarrow & f(2x) \in W_{j+1} \\ f(x) \in W_0 & \Leftrightarrow & \forall n \in \mathbb{Z}, f(x-n) \in W_0, \\ \exists \psi \in W_0 & \textit{telle que} & \{\psi(x-n) : n \in \mathbb{Z}\} \textit{ soit une base orthonormée de } W_0. \end{array} \tag{2.8}$$

ψ, fonction génératrice de l'espace W_0, est appelée ondelette mère de l'analyse multirésolution.

A partir de la fonction ψ génératrice de l'espace W_0, pour les mêmes raisons que précédemment, il est possible de définir une base de W_j (par dilatation et translation ψ) :

$$\{\psi_{j,n}\}_{n\in\mathbb{Z}} = \left\{2^{j/2}\psi(2^j t - n)\right\}_{n\in\mathbb{Z}} \text{ est une base orthonormale de } W_j. \tag{2.9}$$

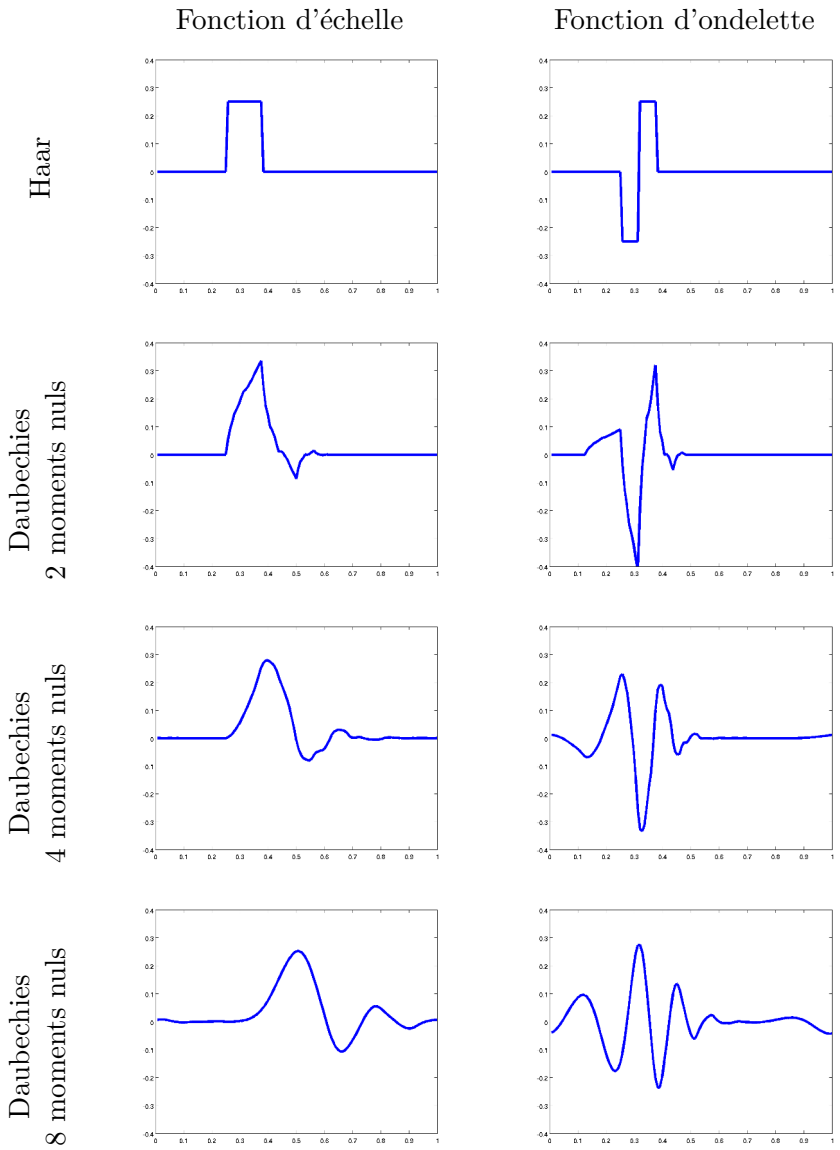


FIGURE 2.1 – Exemple de fonctions d'échelles ϕ (à gauche) et d'ondelettes ψ (à droite) pour une échelle j donnée. Pour générer l'espace d'approximation à cette échelle, il est nécessaire d'utiliser 8 fonctions de bases (qui sont juste les translatés des fonctions présentées ici). On observe que le support de ces fonctions dépend de la base choisie.

Principe de l'analyse multi-résolution

L'idée derrière une analyse multi-résolution consiste, lorsqu'on étudie un signal f vivant dans un espace V_j, à décomposer l'espace d'approximation V_j sous la forme :

$$V_j = V_{j_0} \bigoplus_{p=j_0}^{j-1} W_p \tag{2.10}$$

où l'échelle j_0 constitue le niveau d'approximation grossière désiré (j_0 pouvant être choisie de manière arbitraire).
Le signal $f(x)$ s'écrit dans ce nouvel espace sous la forme :

$$f(x) = \underbrace{\sum_{k\in\mathbb{Z}} c^{j_0}[k]\phi_{j_0,k}(x)}_{\text{Projection dans } V_{j_0}} + \sum_{p=j_0}^{j-1} \underbrace{\sum_{k\in\mathbb{Z}} d^p[k]\psi_{p,k}(x)}_{\text{Projection dans } W_p} \tag{2.11}$$

Ainsi, un signal peut être représenté par une approximation grossière auquel est ajouté l'ensemble des détails perdus lors de l'approximation.

Un exemple est présenté en figure 2.2.

Remarque 2.3

Les coefficients d'approximation $c^j[k]$ et de détail $d^j[k]$ sont donnés par :

$$c^j[k] = \langle f, \phi_{j,k}\rangle \tag{2.12}$$
$$d^j[k] = \langle f, \psi_{j,k}\rangle\,. \tag{2.13}$$

2.2.2 Le rôle des filtres

On peut montrer qu'une analyse multi-résolution est entièrement caractérisée par la fonction d'échelle ϕ utilisée pour générer les espaces V_j. En pratique, cette fonction d'échelle est spécifiée au travers d'un filtre discret h.

Les filtres

Ces filtres sont liés à la définition de l'analyse multi-résolution. On peut les "exhiber" en caractérisant la relation liant les différentes échelles de l'approximation.

On sait en effet que $V_{-1} \subset V_0$, ce qui signifie que toute fonction de V_{-1} peut s'écrire comme une combinaison linéaire des fonctions de base de V_0 (dont la base est $\{\phi(x-n)\}_{n\in\mathbb{Z}}$).
C'est en particulier le cas de la fonction $\frac{1}{\sqrt{2}}\phi\left(\frac{x}{2}\right)$ utilisée pour générer par translation l'espace V_{-1} (voir équation (2.6)). Il existe donc un jeu de coefficients $h = \{h[k]\}_{k\in\mathbb{Z}}$ permettant de décomposer $\frac{1}{\sqrt{2}}\phi\left(\frac{x}{2}\right)$ dans V_0 sous la forme :

$$\frac{1}{\sqrt{2}}\phi\left(\frac{x}{2}\right) = \sum_{-\infty}^{+\infty} h[n]\phi(x-n) \tag{2.14}$$

avec

$$h[n] = \left\langle \frac{1}{\sqrt{2}}\phi\left(\frac{x}{2}\right), \phi(x-n)\right\rangle \tag{2.15}$$

Pour les mêmes raisons (voir équation (2.9)), il existe une suite de coefficients $g = \{g[k]\}_{k\in\mathbb{Z}}$ telle que la fonction d'ondelettes $\frac{1}{\sqrt{2}}\psi\left(\frac{x}{2}\right) \in W_{-1}$ s'écrive de la manière suivante dans V_0 :

$$\frac{1}{\sqrt{2}}\psi\left(\frac{x}{2}\right) = \sum_{-\infty}^{+\infty} g[n]\phi(x-n) \tag{2.16}$$

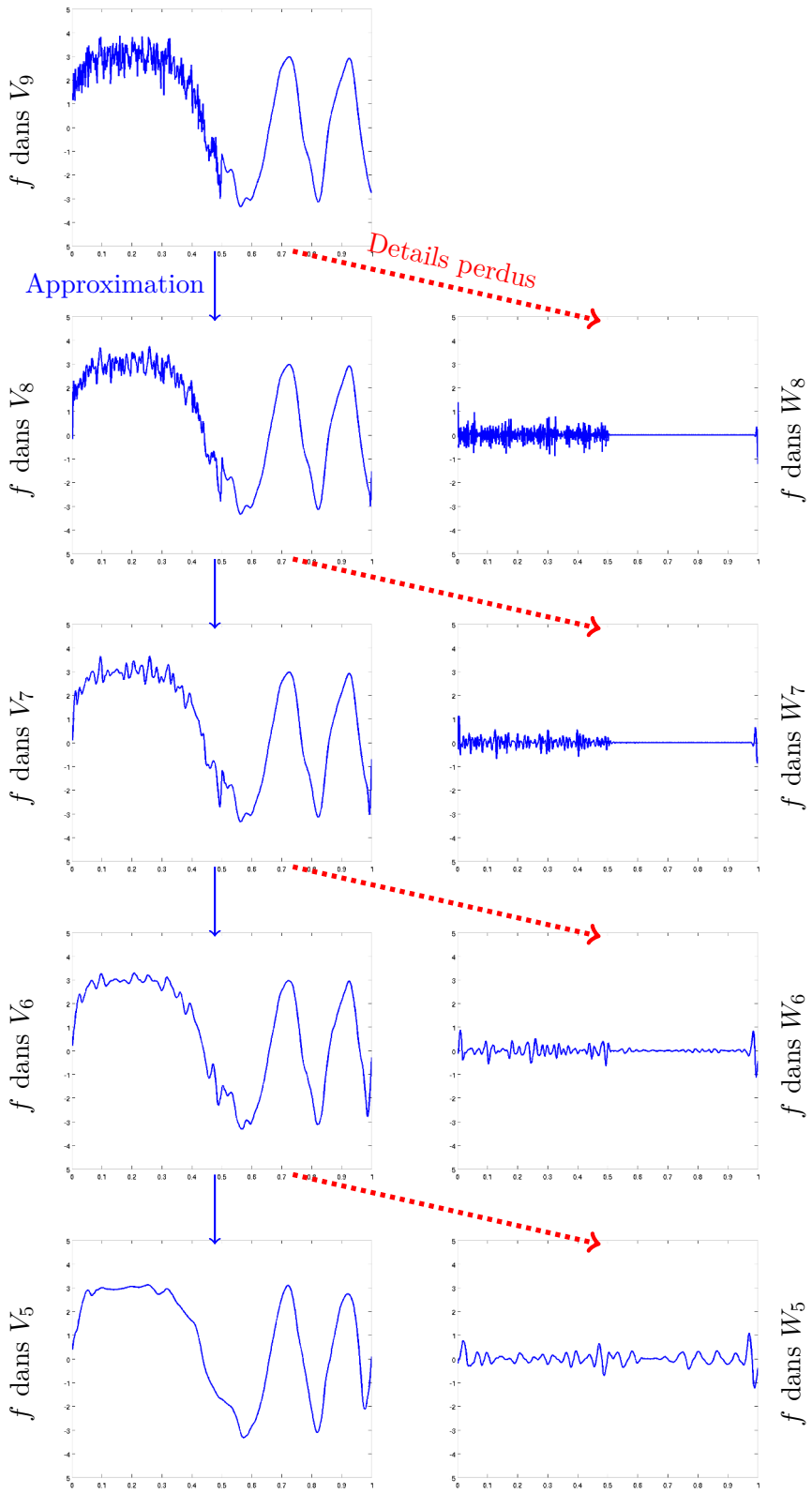


FIGURE 2.2 – Décomposition d'un signal dans une base d'ondelettes. Les approximations d'un signal f vivant dans l'espace V_9 sont présentés dans différents espace V_i de taille de plus en plus petite (à gauche). Les détails perdus entre deux approximations sont présentés à droite.

avec

$$g[n] = \left\langle \frac{1}{\sqrt{2}}\psi\left(\frac{x}{2}\right), \phi(x-n)\right\rangle \tag{2.17}$$

Les relations (2.14) et (2.16) s'appellent les **relations à deux échelles**. Elles caractérisent les liens existants entre l'espace V_j et les espaces V_{j-1} et W_{j-1}. Les filtres h et g sont le "reflet" de cette relation entre échelles. Dans les faits, le filtre h est un filtre passe-bas tandis que le filtre g est un filtre passe-haut.

Remarque 2.4

En pratique, les fonctions d'échelles ϕ et les fonctions d'ondelettes ψ ne sont jamais utilisées de manière formelle. Des conditions sur la transformée de Fourier de ces filtres donnent des conditions suffisantes pour assurer que les fonctions ψ et ϕ correspondantes soient les fonctions génératrices des espaces V_0 et W_0 utilisés en définitions 2.1 et 2.3 (voir par exemple [Mallat, 2008] théorème 7.2 et Lemme 7.3).

Les coefficients $c^j[k]$ et $d^j[k]$ ne sont pas déterminés de manière effective par les équations (2.12) et (2.13). En pratique, la transformée en ondelettes discrète d'un signal vivant dans l'espace V_j est implémentée via l'utilisation des filtres h et g. Ces filtres jouent le rôle principal dans l'implémentation (dont le coût est en $O(N)$ avec N le nombre de pixels du signal) de la transformée en ondelettes discrète.

Remarque 2.5

Au vu des équations (2.15) et (2.17), il est légitime de penser qu'un lien existe entre le filtre h et le filtre g. Ce lien n'est pas défini de manière unique (entre les coefficients des filtres) mais il existe bel et bien. Par exemple dans [Mallat, 2008] la relation est donnée par

$$g^1[n] = (-1)^{1-n}h[1-n]. \tag{2.18}$$

tandis que pour d'autres auteurs tels [Truchetet, 1998, Vidakovic and Mueller, 1994] la relation est donnée par :

$$g^2[n] = (-1)^{n}h[1-n]. \tag{2.19}$$

Dans la suite, nous utiliserons la seconde définition. Les filtres obtenus par ces deux définitions définissent des ondelettes symétriques par rapport à l'axe des abscisses ($g^1[n] = -g^2[n]$). Le changement de définition influe donc uniquement sur le signe du coefficient de détails.

Algorithme de décomposition

En pratique, on suppose que le signal unidimensionel f est décrit sur $N = 2^J$ points. Ce signal est représenté par un vecteur $c^J = \left\{c^J[k]\right\}_{k=0,\cdots,2^J-1}$ de coefficients correspondants aux différentes fonctions de l'espace V_J de telle sorte que :

$$f(x) = \sum_{k=0}^{2^J-1} c^J[k]\phi_{j,k}(x) \tag{2.20}$$

Effectuer une décomposition en ondelettes de ce signal consiste à le réécrire dans $V_{j_0} \oplus W_{j_0} \oplus \cdots \oplus W_{J-1}$ avec $j_0 \geq 0$.
Le fait que $V_J = W_{J-1} \oplus V_{J-1}$ nous permet d'obtenir les coefficients d'approximations et de détails à l'échelle $J-1$. Pour cela, la décomposition des fonctions d'échelles $\phi_{J-1,k}$ et d'approximations $\psi_{J-1,k}$ dans V_J est utilisée. Au final (voir par exemple [Mallat, 2008] chap. 7) les

relations de récurrence suivantes sont obtenues :

$$c^{j-1}[k] = \sum_n h[n-2k]c^j[n] \quad \forall k = 0 \cdots 2^{j-1} - 1 \tag{2.21}$$

$$d^{j-1}[k] = \sum_n g[n-2k]d^j[n] \quad \forall k = 0 \cdots 2^{j-1} - 1 \tag{2.22}$$

Ces égalités peuvent s'exprimer en termes de convolutions circulaires :

$$c^{j-1}[k] = c^j \star \overline{h}[2k] \quad \forall k = 0 \cdots 2^{j-1} - 1 \tag{2.23}$$

$$d^{j-1}[k] = d^j \star \overline{g}[2k] \quad \forall k = 0 \cdots 2^{j-1} - 1 \tag{2.24}$$

avec $\overline{h}[k] = h[-k]$ et où les vecteurs c^j et d^j sont donnés par $c^j = \left\{c^j[k]\right\}_{k=0,\cdots,2^j-1}$ ainsi que $d^j = \left\{d^j[k]\right\}_{k=0,\cdots,2^j-1}$. Ce sont ces dernières formules qui sont utilisées pour implémenter les transformées en ondelettes orthogonales.

Remarque 2.6

On remarque au travers de ces équations que les vecteurs c^{j-1} et d^{j-1} comportent deux fois moins d'éléments que le vecteur c^j.

La figure 2.3 résume les opérations appliquées au vecteur c^j pour obtenir les vecteurs c^{j-1} et d^{j-1}.

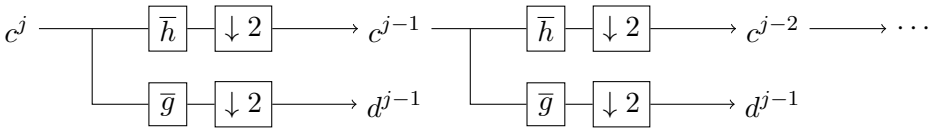


FIGURE 2.3 – Illustration de la transformée rapide en ondelettes. Cette dernière est calculée par une convolution avec les filtres $\overline{h}$ et $\overline{g}$ suivis d'une opération de décimation d'un facteur 2.

Exemple 2.1 (*Decomposition d'un signal*)

Afin d'illustrer les étapes de décomposition d'un signal, nous reprenons ici l'exemple utilisé dans [Vidakovic and Mueller, 1994] (qui constitue une très bonne source si on souhaite appréhender de manière intuitive ce qu'est une transformée en ondelettes).
Dans cet exemple, on utilise la transformée en ondelettes de Haar.
Le filtre $h = \begin{pmatrix} h[0] & h[1] \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$: il correspond ici simplement à la moyenne (renormalisée) du signal sur un intervalle.
Le filtre g est le filtre "miroir" du filtre h donné par l'équation (2.19).
Il s'écrit $g = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix}$ et correspond donc à un calcul de "gradient" (renormalisé) en un point donné.

Soit un signal de départ

$$c^3 = \begin{pmatrix} 1 & 0 & -3 & 2 & 1 & 0 & 1 & 2 \end{pmatrix}. \tag{2.25}$$

Ce signal peut être décomposé en deux parties $\begin{pmatrix} c^2 & d^2 \end{pmatrix}$ où les coefficients de c^2 correspondent à des "moyennes" tandis que ceux de d^2 correspondent à des "gradients" (ou détails perdus en effectuant la moyenne). En utilisant les filtres de Haar définis précédemment, on obtient l'approximation suivante :

$$c^2 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{3}{\sqrt{2}} \end{pmatrix} \tag{2.26}$$

tandis que les coefficients de détail (perdus lors de l'approximation du signal) sont :

$$d^2 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{-5}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \tag{2.27}$$

Le vecteur c^2 peut alors être décomposé en $\begin{pmatrix} c^1 & d^1 \end{pmatrix}$ et ainsi de suite avec tous les vecteurs d'approximations obtenus (jusqu'à ce que cela n'ait plus de sens de le décomposer).
Au final, l'équivalent du vecteur c^3 dans la base de Haar est donné par le vecteur :

$$\begin{pmatrix} c^0 & d^0 & d^1 & d^2 \end{pmatrix} = \begin{pmatrix} \sqrt{2} & -\sqrt{2} & 1 & -1 & \frac{1}{\sqrt{2}} & \frac{-5}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \tag{2.28}$$

Remarque 2.7 (*Ondelettes sur l'intervalle*)

Comme dans l'exemple précédent, les signaux manipulés sont en pratique bornés sur un intervalle. Il est donc nécessaire de définir des bases d'ondelettes sur un intervalle borné. Plusieurs manières d'effectuer cette opération existent (voir par exemple [Mallat, 2008]). On choisit dans notre cas de considérer que le signal est périodique sur l'intervalle (ainsi que les différentes fonctions d'échelles et d'approximation) ce qui explique la convolution circulaire utilisée en équation (2.23) et (2.24).

Remarque 2.8 (*Sur les notations*)

En pratique, comme par exemple en section 2.3 et dans le chapitre 5, lorsqu'il n'est pas nécessaire d'avoir le détail des formules de décomposition nous n'utilisons pas les formules (2.21) et (2.22). A la place, nous utilisons les opérateurs H et G "modélisant" l'action des filtres h et g sur le vecteur c_{j+1} et de la décimation :

$$c^j = Hc^{j+1} \tag{2.29}$$
$$d^j = Gc^{j+1} \tag{2.30}$$

Il est ainsi plus aisé de décrire les différentes étapes liées à la décomposition (ou recomposition) à l'aide de ces notations. Ainsi, dans l'exemple précédent, la transformée en ondelette de c_3 peut être résumée en une ligne par :

$$\begin{pmatrix} c^0 & d^0 & d^1 & d^2 \end{pmatrix} = \begin{pmatrix} H^3c^3 & GH^2c^3 & GHc^3 & Gc^3 \end{pmatrix}. \tag{2.31}$$

Algorithme de recomposition

De manière analogue, les formules de recomposition s'obtiennent en exprimant les fonctions d'échelles $\phi_j \in V_j$ dans $V_{j-1} \oplus W_{j-1}$. Les coefficients d'approximation à l'échelle j s'écrivent :

$$c^j[k] = \sum_n h[k-2n]c^{j-1}[n] + \sum_n g[k-2n]d^{j-1}[n] \quad \forall k = 0, \cdots, 2^j - 1 \tag{2.32}$$

On remarque dans cette formule qu'un coefficient sur deux du filtre n'est pas utilisé lors de la recomposition du signal (pour une position k donnée).
En posant :

$$\breve{x}[n] = \begin{cases} x[p] & \text{si } n = 2p \\ 0 & \text{sinon} \end{cases} \tag{2.33}$$

on peut réécrire la formule précédente sous forme de convolution circulaire :

$$c^j[k] = \breve{c}^{j-1} \star h[k] + \breve{d}^{j-1} \star g[k] \tag{2.34}$$

La figure 2.4 illustre la recomposition d'un signal à partir de sa représentation dans une base d'ondelettes.

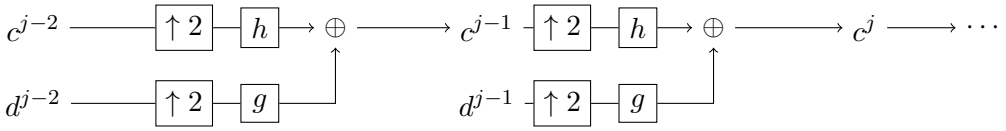


FIGURE 2.4 – Illustration de la recomposition des coefficients d'approximation à l'échelle j. Ces derniers sont calculés, après un sur-échantillonage des coefficients de détail et d'approximation à l'échelle précédente $j-1$, par convolution circulaire avec les filtres h et g.

Remarque 2.9

Le vecteur c^3, vivant dans l'espace des pixels, s'écrit à partir des opérateurs H et G sous la forme :

$$\begin{aligned} c^3 &= H^T c^2 + G^T d^2 \\ &= H^T(H^T(H^T c^0 + G^T d^0) + G^T d^1) + G^T d^2. \end{aligned}$$

2.2.3 Propriétés des bases d'ondelettes

Les points expliqués précédemment, bien que nous présentant une manière de construire une analyse du signal dépendant de la résolution (en accord avec la définition 2.1), ne nous expliquent pas pourquoi utiliser une telle décomposition permet d'écrire certaines classes de signaux de manière parcimonieuse. Dans les faits, c'est le nombre de moments nuls imposés aux fonctions d'ondelettes ψ qui rend possible cette parcimonie.

Nombre de moments nuls d'une ondelette

Définition 2.4

On dit d'une fonction ψ qu'elle possède p moments nuls si

$$\int_{-\infty}^{+\infty} x^k \psi(x)\, dx = 0 \quad \forall 0 \leq k \leq p-1 \tag{2.35}$$

Une fonction ayant p moments nul est donc orthogonale à tout polynôme de degré $p-1$. Dès lors, lorsque sur le support de la fonction $\psi_{j,k}$, le signal f est bien approché par son développement en série de Taylor d'ordre $p-1$, les coefficients de détails $d^j[k]$ sont proches de 0. C'est cette caractéristique qui induit la parcimonie de la représentation dans une base d'ondelettes.

Remarque 2.10

Par contre, si sur le support de $\psi_{j,k}$ le signal présente une singularité, l'amplitude du coefficient $d^j[k]$ est élevée (bien que l'ensemble de la singularité ne soit pas décrite par les coefficients de détails). Les coefficients $d^j[k]$ laissent donc passer la partie lisse et "capturent" uniquement les discontinuités du signal.

Ce phénomène est illustré en figure 2.5.

Taille du support d'une ondelette

La propriété précédente est à mettre en lien avec la taille du support de l'ondelette. Pour que le nombre de moment d'une ondelette ait un intérêt quelconque (dans le but d'écrire un signal de manière parcimonieuse) le signal doit être suffisament *lisse* sur le *support* de l'ondelette.

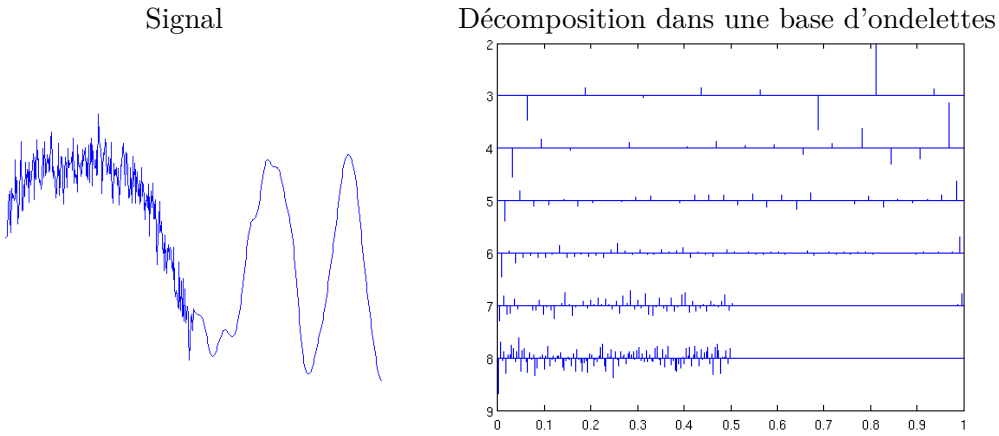


FIGURE 2.5 – Coefficients de détails du signal (présenté à gauche) dans la base d'ondelettes de Daubechies à 4 moments nuls. Les coefficients d'ondelettes sont représentés par échelle (chaque ligne de l'axe vertical correspond à une échelle différente). Ici, la transformée est effectuée sur 6 niveaux. Il y a 256 coefficients de détails à l'échelle $j = 8$. Les coefficients à chaque échelle ont été renormalisés par un facteur $2^{j/2}$.

La taille du support d'une ondelette est reliée à la taille du support de la fonction d'échelle (et du filtre h) :

Propriété 2.1 (*Support compact*)

La fonction d'échelle ϕ a un support compact si et seulement si le filtre h a un support compact (et dans ce cas leurs supports sont égaux). Si le support de ϕ est $[N_1; N_2]$ alors le support de la fonction d'ondelette ψ est $\left[\frac{(N_1 - N_2 + 1)}{2}; \frac{N_2 - N_1 + 1}{2}\right]$.

La propriété suivante, due à [Daubechies, 1988], exprime le lien existant entre la taille du support d'une ondelette et son nombre de moments nuls.

Propriété 2.2

Si une ondelette ψ a p moments nuls alors son support est supérieur ou égal à $2p - 1$. Les ondelettes de Daubechies ont un support de taille minimale (égal à $[-p + 1; p]$) pour un nombre de moments nuls p donné.

Il y a donc un compromis à faire entre le nombre de moments nuls d'une ondelette et la taille de son support. En augmentant le nombre de moments nuls on augmente le support sur lequel, localement, le signal doit être bien approché par un polynôme. Par conséquent, lorsque le signal présente de nombreuses discontinuités, augmenter le nombre de moments nuls de l'ondelette peut réduire le nombre de coefficients de faible amplitude. A contrario, si le signal présente peu de singularités, afin d'augmenter le nombre de coefficients de détails ayant une faible amplitude, il est conseillé de choisir une ondelette possédant de nombreux moments nuls.

2.2.4 Opérations de seuillage

Un intérêt des représentations dans des bases d'ondelettes (ou d'autre $X - let$) provient de la faculté d'écrire certains types de signaux de manière parcimonieuse dans ces espaces. La parcimonie observée (ou supposée) du signal peut être utilisée pour compresser l'information utile ou comme un *a priori* sur le *vrai* signal dans une phase d'analyse.

Parmi les nombreuses opérations réalisées sur les coefficients d'ondelettes, les opérations de seuillage sont sans doute les plus courantes.

Ces opérations consistent à "omettre" certains détails de la décomposition en ondelettes sans pour autant modifier "l'allure générale" du signal. L'idée consiste à se débarasser des détails considérés comme non discernables du bruit. Différentes manières de seuiller les coefficients existent : les plus connues sont le seuillage brut et le seuillage doux.

Seuillage brut

Le seuillage brut consiste simplement à forcer à 0les coefficients dont la norme ne dépasse pas un certain seuil τ. Il s'écrit sous la forme suivante :

$$\mathcal{T}(d^j[k]) = \begin{cases} d^j[k] & \text{si } |d^j[k]| > \tau \\ 0 \text{ sinon} \end{cases} \tag{2.36}$$

Seuillage doux

Le seuillage doux consiste à atténuer la discontinuité introduite par le seuillage précédent. Ce dernier est défini de la manière suivante :

$$\mathcal{T}(d^j[k]) = \begin{cases} d^j[k] - \tau \times \text{sign}(d^j[k]) & \text{si } |d^j[k]| > \tau \\ 0 \text{ sinon} \end{cases} \tag{2.37}$$

Remarque 2.11

Par exemple, lorsqu'on cherche à débruiter un signal y, le problème peut se résumer à trouver un signal $\hat{\theta}$ (dans un espace d'ondelettes) tel que

$$\hat{\theta} = \underset{\theta}{\text{argmin}} \, \|\mathbf{W}y - \theta\|^2 + p(\theta, \lambda) \tag{2.38}$$

où $\mathbf{W}y$ est la décomposition du signal original dans une base d'ondelettes et $p(\theta, \lambda)$ est une fonction de pénalisation (imposant par exemple une certaine régularité du signal).
Dans ce cadre, il existe un lien entre le seuillage des coefficients et le problème de minimisation précédent. Effectuer un seuillage doux avec τ comme seuil correspond à choisir dans l'équation (2.38) le terme de pénalisation suivant

$$p(\theta_i, \tau) = \tau|\theta_i|$$

pour chaque coefficient d'ondelette θ_i. Le lecteur intéressé par les relations liant seuillages et pénalisations peut se référer à [Antoniadis et al., 2007].

Exemple 2.2 (*Effet d'un seuillage*)

Afin de comprendre la différence entre ces deux procédures un seuillage des coefficients de l'exemple 2.1 est effectué. Le seuil τ est choisi de manière arbitraire et fixé à 0.9. Par conséquent, le vecteur représentant le signal dans l'espace d'ondelettes, i-e

$$\begin{pmatrix} \sqrt{2} & -\sqrt{2} & 1 & -1 & \frac{1}{\sqrt{2}} & \frac{-5}{\sqrt{2}} & \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} \end{pmatrix} \tag{2.39}$$

devient, lorsqu'un seuillage brut, est appliqué

$$\begin{pmatrix} \sqrt{2} & -\sqrt{2} & 1 & -1 & 0 & \frac{-5}{\sqrt{2}} & 0 & 0 \end{pmatrix} \tag{2.40}$$

ou

$$\begin{pmatrix} 0.51 & -0.51 & 0.1 & -0.1 & 0 & -2.63 & 0 & 0 \end{pmatrix} \tag{2.41}$$

lorsqu'un seuillage doux est appliqué.

> *La figure 2.6 présente le résultat de ces seuillages. Le signal original est présenté en rouge. La reconstruction à partir des coefficients seuillés de manière "brute" (2.40) est présentée en bleu tandis que celle obtenue à partir du seuillage doux (2.41) est présentée en vert.*

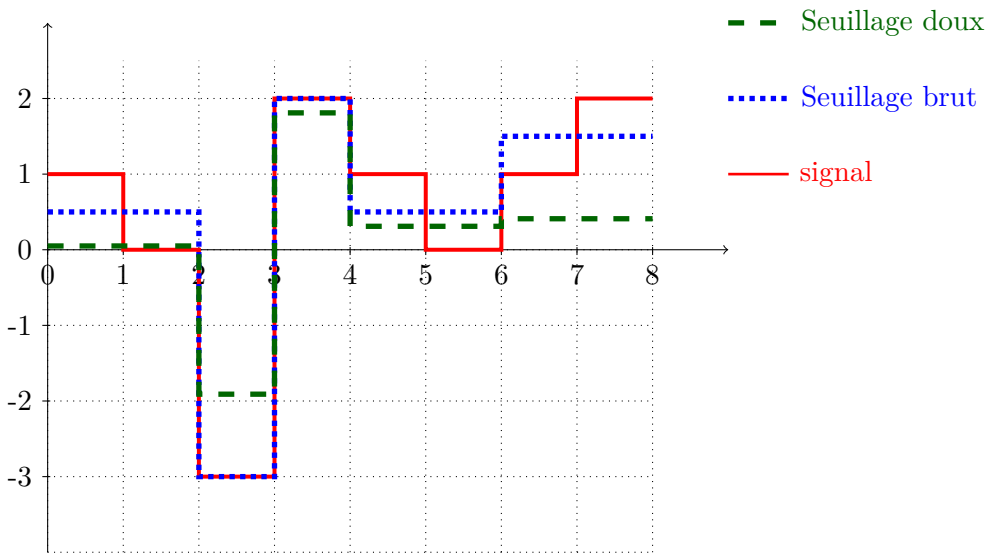


FIGURE 2.6 – Illustration du signal reconstruit après seuillage. En rouge est présenté le signal original (de l'exemple 2.1). En bleu (resp. en vert) est présenté le signal reconstruit après le seuillage brut (resp. seuillage doux) correspondant à l'exemple 2.2.

Il existe dans la littérature une multitude de fonctions de seuillage comme le seuillage SCAD [Antoniadis and Fan, 2001]. Ces seuillages définissent souvent un "compromis" entre seuillage brut et seuillage doux. D'autres fonctions de seuillage, dérivées des travaux de [Cai, 1999, Cai and Silverman, 2001], s'effectuent sur des blocs de coefficients afin de tenir compte de la structure spatiale des coefficients d'ondelettes. Le lecteur intéressé par les différents seuillages existants (et les raisons ayant conduit à les considérer) peut une nouvelle fois se tourner vers [Antoniadis et al., 2007]. De manière identique, il existe un grand nombre de manières de définir le seuil τ en dessous duquel un coefficient est considéré comme non important. Comme précédemment pour les fonctions de seuillage, la plupart de ces méthodes proviennent de travaux effectués en statistiques. Ils ont été inaugurés par les travaux de [Donoho and Johnstone, 1994] définissant (entre autres) le seuil "universel" suivant :

$$\tau = \sigma\sqrt{2\log M} \tag{2.42}$$

où M est le nombre de coefficients d'ondelettes et σ la variance d'un bruit blanc non corrélé dans l'espace des pixels. Le lecteur intéressé par la manière de définir ces seuils peut trouver une porte d'entrée à ce sujet dans le chapitre 11 de [Mallat, 2008].

2.3 Ondelettes séparables en dimension 2

Jusqu'à présent, seul un signal unidimensionnel a été considéré. Lorsque l'on travaille en dimension n, on peut, à partir de n'importe quelle base d'ondelettes de $L^2(\mathbb{R})$ générer une base d'ondelettes *séparable* de $L^2(\mathbb{R}^n)$ au moyen d'un produit tensoriel. Nous allons dans la suite de cette partie détailler quelque peu le cas où $n = 2$ puis nous expliquerons l'intérêt porté depuis le début des années 2000 à la recherche de nouvelles transformées multi-échelles.

2.3.1 Construction

Considérons une série d'espaces emboîtés $\{V_j\}_{j\in\mathbb{Z}}$ définissant une analyse multi-résolution mono-dimensionnelle de $L^2(\mathbb{R})$. Dans ce cadre, l'espace $\mathcal{V}_j = V_j \otimes V_j$ définit une analyse multi-résolution de $L^2(\mathbb{R}^2)$.

L'espace V_j pouvant s'écrire sous la forme $V_j = V_{j-1} \oplus W_{j-1}$, l'espace d'approximation en dimension 2 peut s'écrire :

$$\begin{aligned} \mathcal{V}_j &= (V_{j-1} \otimes V_{j-1}) \oplus (W_{j-1} \otimes V_{j-1}) \oplus (V_{j-1} \otimes W_{j-1}) \oplus (W_{j-1} \otimes W_{j-1}) \\ &= \mathcal{V}_{j-1} \oplus \mathcal{W}^1_{j-1} \oplus \mathcal{W}^2_{j-1} \oplus \mathcal{W}^3_{j-1} \end{aligned} \tag{2.43}$$

Chaque espace d'approximation $\mathcal{V}^j$ peut donc être décomposé en un espace d'approximation plus grossière $\mathcal{V}^{j-1}$ et trois espaces de détails $\mathcal{W}^1_{j-1}, \mathcal{W}^2_{j-1}, \mathcal{W}^3_{j-1}$.
De la même manière que l'espace V_j était engendré par les fonctions d'approximations ϕ, ces espaces sont engendrés par les fonctions suivantes :

$$\begin{aligned} \mathcal{V}_j &= \text{Vect}\left\{\Phi_{j,\mathbf{k}}(x,y) \quad \mathbf{k} \in \mathbb{Z}^2\right\} \\ \mathcal{W}^1_j &= \text{Vect}\left\{\Psi^1_{j,\mathbf{k}}(x,y) \quad \mathbf{k} \in \mathbb{Z}^2\right\} \\ \mathcal{W}^2_j &= \text{Vect}\left\{\Psi^2_{j,\mathbf{k}}(x,y) \quad \mathbf{k} \in \mathbb{Z}^2\right\} \\ \mathcal{W}^3_j &= \text{Vect}\left\{\Psi^3_{j,\mathbf{k}}(x,y) \quad \mathbf{k} \in \mathbb{Z}^2\right\} \end{aligned}$$

où

$$\begin{aligned} \Phi_{j,\mathbf{k}}(x,y) &= \phi_{j,k_1}(x)\phi_{j,k_2}(y) \\ \Psi^1_{j,\mathbf{k}}(x,y) &= \phi_{j,k_1}(x)\psi_{j,k_2}(y) \\ \Psi^2_{j,\mathbf{k}}(x,y) &= \psi_{j,k_1}(x)\phi_{j,k_2}(y) \\ \Psi^3_{j,\mathbf{k}}(x,y) &= \psi_{j,k_1}(x)\psi_{j,k_2}(y) \end{aligned}$$

Ces fonctions sont présentées en figure 2.7 pour les ondelettes de daubechies à 4 moments nuls.

Les fonctions Ψ^1 sont construites à l'aide d'une fonction d'échelle dans la direction x et d'une fonction d'ondelette dans la direction y. Elles ont donc tendance à "repérer" les *singularités* horizontales de l'image.

Remarque 2.12

Si la fonction d'ondelettes ψ a p moments nuls alors la fonction Ψ^1 possède p moments nuls dans la direction y. Ceci doit être compris dans le sens où Ψ^1 est orthogonale à toute fonction $f(x,y)$ polynomiale de degré $p-1$ dans la direction y lorsque x est fixé. Par conséquent, si sur le support de Ψ^1 la fonction f est bien approchée par un polynôme de degré $p-1$ dans la direction y le coefficient correspondant tendra à être faible. De manière similaire, Ψ^2 possède p moments nuls dans la direction x (et a donc tendance à repérer les singularités verticales d'une image). L'ondelette Ψ^3 possède p moments nuls dans la direction x et dans la direction y. Cette dernière n'est pas une ondelette directionnelle.

2.3.2 Obtention des coefficients

Pour obtenir les coefficients en deux dimensions, il suffit d'effectuer une étape de décomposition de la transformée en ondelettes sur chaque ligne de l'image puis d'effectuer une étape de décomposition sur chaque colonne de l'image. La figure 2.8 résume ces opérations.

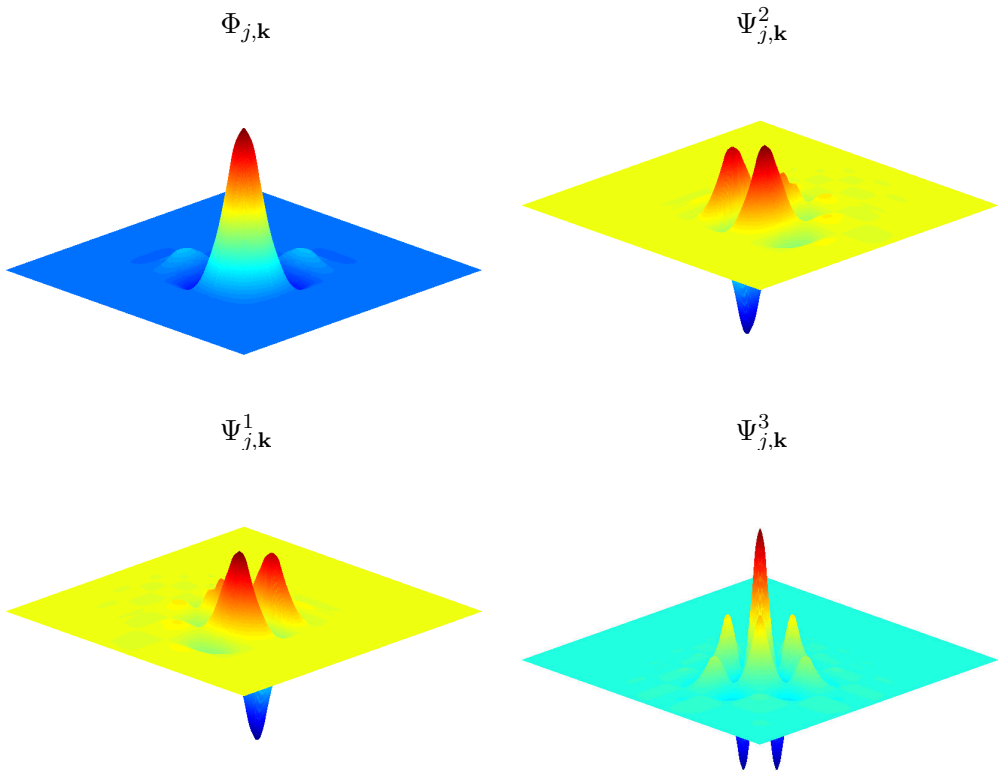


FIGURE 2.7 – Fonctions d'échelle et d'ondelettes dans les différentes directions en dimension 2 correspondant aux ondelettes de daubechies à 4 moments nuls.

L'obtention des coefficients d'approximation et de détail à l'échelle j en connaissant les coefficients d'approximation à l'échelle $j+1$ peut s'écrire ainsi :

$$\begin{aligned} C^j &= H_c\ H_r\ C^{j+1} \\ D_h^j &= H_c\ G_r\ C^{j+1} \\ D_v^j &= G_c\ H_r\ C^{j+1} \\ D_d^j &= G_c\ G_r\ C^{j+1} \end{aligned}$$

où H_r est l'opérateur modélisant le calcul des coefficients d'approximation sur les lignes, tandis que H_c modélise le calcul des coefficients d'approximation sur les colonnes.

Remarque 2.13

Lorsque dans la direction x et la direction y la même base d'ondelettes est considérée, l'opérateur H_c correspond à la transposée de l'application de H_r sur les colonnes de la matrice. Par conséquent, on a :

$$H_c H_r C^{j+1} = (H_r\ (H_r C^{j+1})^T)^T = H_r C^{j+1} H_r^T. \tag{2.44}$$

Il en va de même pour l'opérateur G_c (vis–à–vis de G_r).

Organisation des coefficients

Dans ces travaux, nous avons besoin de présenter le résultat d'une transformée en ondelettes d'un signal initial. Les coefficients seront disposés tels que présenté en figure 2.9.

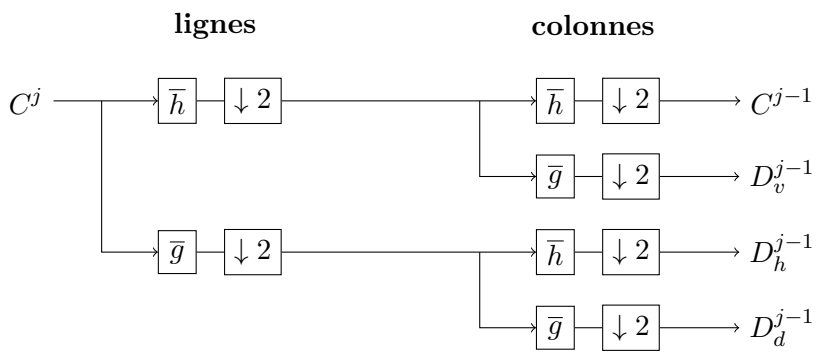


FIGURE 2.8 – Illustration de la transformée rapide en ondelettes séparables en dimension 2.

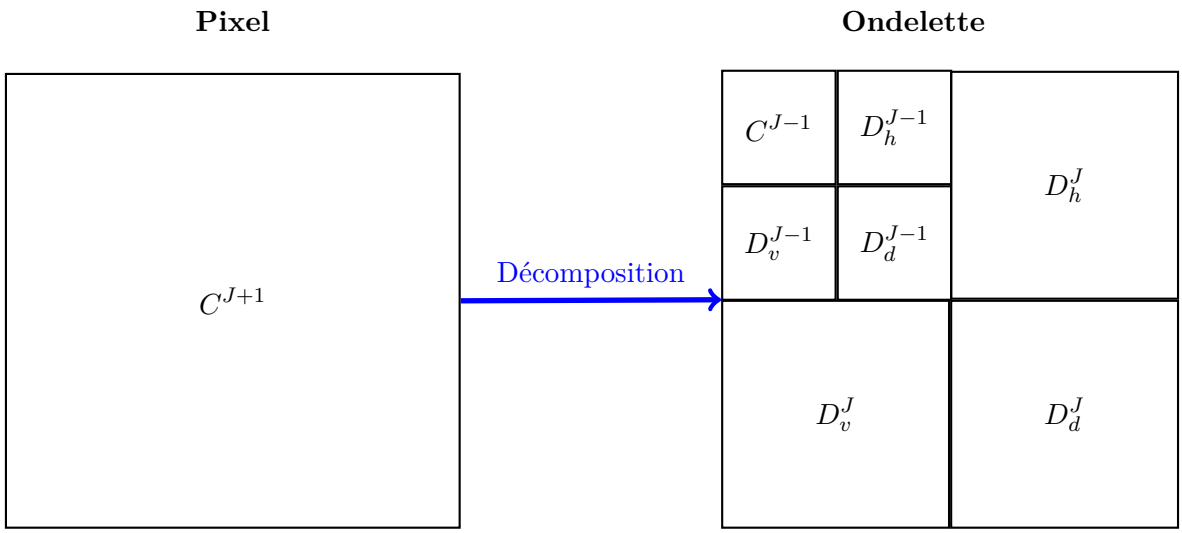


FIGURE 2.9 – Disposition des coefficients lors d'une décomposition en ondelettes 2D. Le bloc D_d^J (resp D_v^J et D_h^J) contient les coefficients de détails diagonaux (resp. verticaux et horizontaux) à l'échelle J. Le bloc C^{J-1} contient les coefficients d'approximation à l'échelle $J-1$. Dans cet exemple, deux niveaux de décomposition ont été utilisés.

2.3.3 Illustration sur un exemple du stockage de l'information dans un espace d'ondelettes

Afin de mieux appréhender comment est stockée l'information dans un espace d'ondelettes, nous allons dans cette partie l'illustrer sur un exemple. La figure 2.10 présente l'image originale (qui est un "classique" du traitement d'images) ainsi que sa décomposition dans une base d'ondelettes (celle de Daubechies à 4 moments nuls). L'image originale (de taille 512×512 pixels) est décomposé sur 3 niveaux d'ondelettes.

On observe que :

— les coefficients de détails importants sont principalement localisés (à petite échelle) au niveau du châle, du pantalon et de la nappe. Ceci est dû à la "texture" de ces objets (carreaux de petite taille).

— dans les zones où l'image est lisse, les coefficients de détails sont (presque) nuls.

— les pieds de la table engendrent des coefficients importants pour les détails horizontaux (situés en haut à droite dans la décomposition en ondelettes) : ceci est dû au fait que les ondelettes correspondant à ces coefficients "capturent" les singularités horizontales du signal.

— la proportion et la norme des coefficients significatifs augmente lorsque la résolution diminue.

FIGURE 2.10 – A gauche, l'image originale. A droite, la valeur absolue des coefficients d'ondelettes (allant de 0 (noir) à 100 (blanc)). L'échelle a été choisie afin de pouvoir visualiser les coefficients correspondants aux détails fins. Ici, la base d'ondelettes choisie est celle de Daubechies à 4 moments nuls.

Remarque 2.14 (*Ondelettes périodiques*)

On peut aussi voir l'effet du choix d'une ondelette périodique : il y a de nombreux coefficients non nuls à proximité des bords. Sur l'image, ils s'illustrent notamment par une bande blanche entre les coefficients de la plus petite échelle et ceux de l'échelle intermédiaire. Utiliser une ondelette périodique revient à périodiser l'image. Dès lors lorsque l'image périodisée présente des discontinuités au niveau des bords, de nombreux coefficients non nuls apparaissent à proximité de ceux-ci.

Information contenue aux différents niveaux d'approximations

La figure 2.11 présente l'information contenue aux différentes échelles de la décomposition en ondelettes.

L'image en haut à gauche est obtenue en utilisant lors de la recomposition uniquement les coefficients d'ondelettes correspondant à l'approximation grossière, c'est–à–dire en supposant que les coefficients de la décomposition en ondelettes sont :

$$c_{coarse} = \begin{pmatrix} C^6 & 0 & 0 & 0 \end{pmatrix}. \tag{2.45}$$

Remarque 2.15

Cela revient à effectuer un seuillage brut présenté en équation (2.36) sur l'ensemble des détails.

L'image en haut à droite est obtenue en utilisant lors de la recomposition uniquement les détails les plus grossiers (quelque soit l'orientation), c'est–à–dire en supposant que le vecteur de coefficients soit :

$$c_{det} = \begin{pmatrix} 0 & D^6 & 0 & 0 \end{pmatrix} \tag{2.46}$$

avec $D^6 = \begin{pmatrix} D_h^6 & D_v^6 & D_d^6 \end{pmatrix}$.

Sur cette image, on observe que

— Les détails à l'échelle la plus fine (sur l'image en bas à droite) portent principalement l'information concernant la "texture" de l'image. Ainsi une partie des oscillations de l'image dues au motif carrelé du châle, de la nappe et du pantalon, est bien capturée à cette

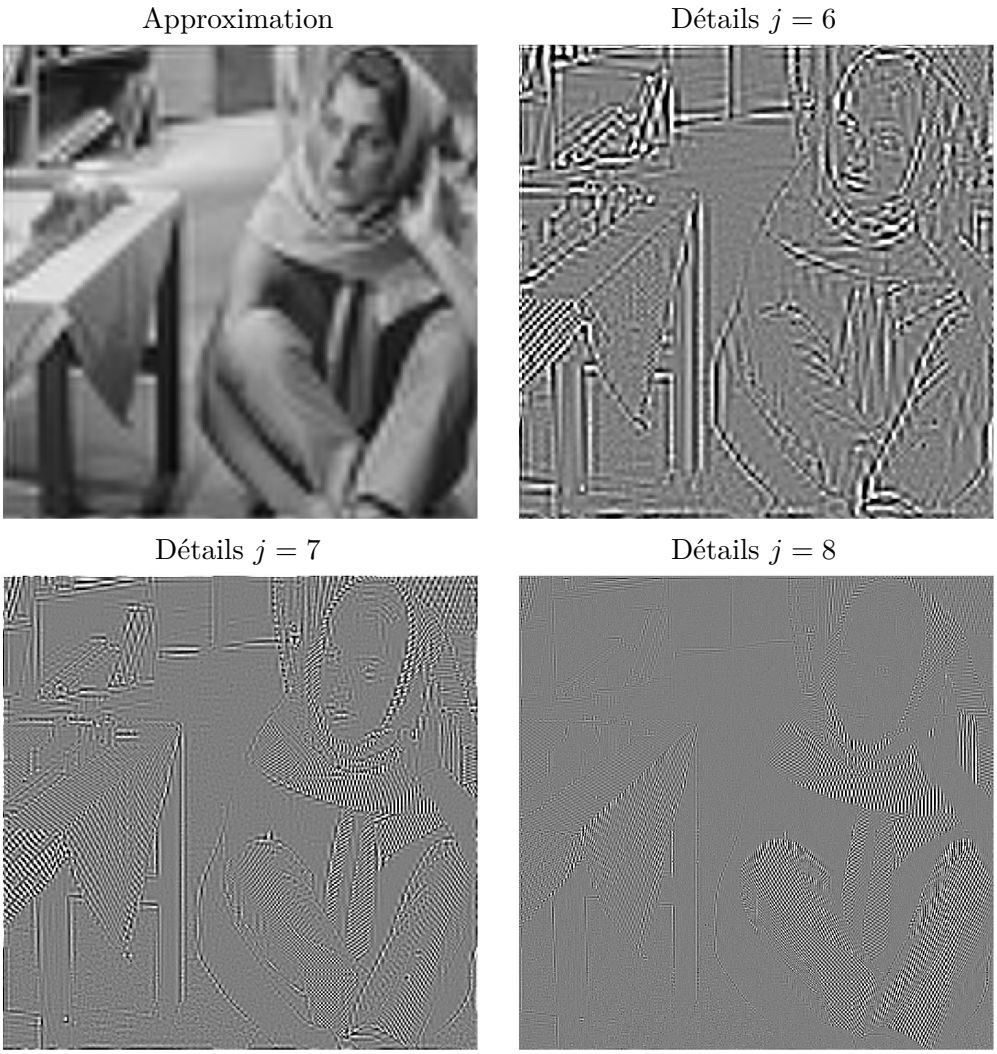


FIGURE 2.11 – Information contenue aux différents niveaux de décomposition. En haut à gauche, la recomposition obtenue à l'aide des coefficients d'approximation est présentée. Pour cette image, le niveau de gris est compris entre 0 (noir) et 256 (blanc). Les trois autres images présentent la recomposition obtenue lors de l'utilisation des détails à une échelle donnée. Pour ces images, le niveau de gris est compris entre -30 (noir) et 30 (blanc). .

échelle. Les contours des principaux objets sont par contre assez peu marqués en comparaison (bien qu'on distingue assez nettement les pieds de la table) ;

— Les détails un peu moins fins (figure en bas à gauche) comportent une partie de l'information sur la "texture" (motif carrelé). A cette échelle, les contours des "objets" (table, bras, armoire, livre) sont bien plus marqués qu'au niveau précédent. La texture des rideaux qui a peu été capturée par les ondelettes du niveau précédent est à cette échelle bien représentée ;

— Les détails les plus grossiers (image en haut à droite) de la décomposition considérée portent très peu d'informations sur la texture. En effet, seul le devant de la nappe, pour lesquels les carreaux sont de plus grande taille, est représentée à cette échelle. A contrario, on distingue de manière très claire les différents contours de l'image ;

— L'approximation à l'échelle grossière (figure en haut à gauche) est quant à elle relativement lisse. La nappe, le pantalon et le foulard ne comportent plus de texture à cette résolution. L'image fournie est "floue" et certaines oscillations parasites font leur apparition, notamment à proximité des bords où l'utilisation d'une ondelette périodique se fait

ressentir.
A titre d'information, cette image est reconstruite avec environ 1.5% des coefficients dans la base d'ondelettes.

2.3.4 Autre représentation en dimension 2

Les bases d'ondelettes séparables en plusieurs dimensions, bien que possédant certaines bonnes propriétés, ont un défaut majeur : beaucoup de coefficients sont nécessaires pour décrire une singularité non horizontale ou verticale. En effet, les fonctions d'ondelettes ainsi obtenues ont seulement deux directions privilégiées : la direction verticale et la direction horizontale. Par conséquent, quand l'image comporte une singularité non dirigée dans ces directions, un grand nombre de coefficients dans ces bases est nécessaire pour la représenter (voir figure 2.12).

Du fait (en partie) de cette limitation, de nombreuses transformées multi-échelles (dont le nom se termine invariablement par *-let*) ont vu le jour depuis le début des années 2000. C'est, entre autre, le cas des transformées d'ondelettes en arbre dual [Kingsbury, 1999], des curvelettes [Candes and Donoho, 2000], des contourlets [Do and Vetterli, 2002] ou des ridgelets [Do and Vetterli, 2003]. Aujourd'hui encore de nombreuses transformées dédiées à des applications particulières apparaissent. Laurent Duval sur son site http://www.laurent-duval.eu/siva-wits-where-is-the-starlet.html recense une partie de ces transformées multi-échelles (environ une centaine est décrite succintement). Un bon article de review sur les différentes transformées multi-échelles et leurs applications est [Jacques et al., 2011].

A noter que ces nouvelles transformées multi-échelles ne s'effectuent plus dans des bases mais dans des trames.

La notion de trame

La notion de trame (frame en anglais) a été introduite par [Duffin and Schaeffer, 1952]. Les trames sont des familles redondantes de vecteur dans un espace de Hilbert dans lesquelles n'importe quel vecteur possède une représentation naturelle mais admet aussi une infinité d'autres représentations [Christensen, 2003].

Théorème 2.1

Une famille de vecteurs $\{f_i\}_{i\in I}$ (avec I un ensemble fini ou infini d'indices) de $\mathcal{H}$ est une trame de $\mathcal{H}$ s'il existe deux constantes $0 < A \leq B < +\infty$ telles que pour tout vecteur $f \in \mathcal{H}$:

$$A\|f\|^2 \leq \sum_i |\langle f, f_i\rangle|^2 \leq B\|f\|^2 \tag{2.47}$$

*Une trame est dite **ajustée** lorsque $A = B$.*
*Lorsque $A = B = 1$ la trame est dite de **Parseval**.*

En pratique, la plupart des trames utilisées sont des trames ajustées (voir par exemple les travaux de [Casazza and Leonhard, 2008] détaillant différentes méthodes pour obtenir des trames ajustées à partir de trames quelconques).
Les coefficients du vecteur f dans la trame considérée sont obtenus via un produit scalaire avec les fonctions f_i formant ces espaces :

$$\{\langle f, f_i\rangle\}_{i\in I}\,. \tag{2.48}$$

La transformée en curvelettes

Dans ces travaux, nous nous intéressons à la transformée en curvelettes. Cette dernière introduite par [Candes and Donoho, 2000], dont la version discrète a été revisitée quelques années plus tard [Candes et al., 2006], a pour but de représenter avec peu de coefficients des objets présentant une discontinuité suivant une courbe $\mathcal{C}^2$ et réguliers partout ailleurs. Pour cela, les

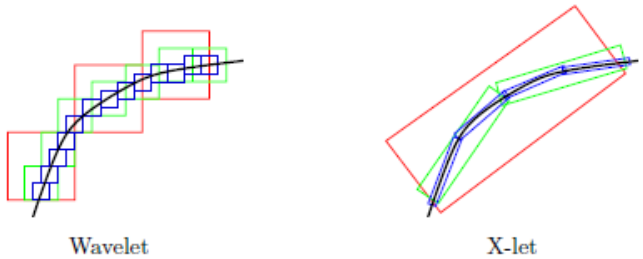


FIGURE 2.12 – Schéma illustrant les raffinements sucessifs d'un contour lisse par les ondelettes et par les nouvelles transformations multi-échelles recherchées. Lorsque le contour n'est pas vertical ou horizontal, de nombreux éléments de la base d'ondelettes sont nécessaires pour reconstruire ce contour. Source : [Do, 2001]

éléments de la trame possèdent chacun une orientation bien spécifique et ont un nombre de moments infinis dans cette direction [Candes et al., 2006].

La décomposition d'une image f dans une trame de curvelettes s'écrit de la manière suivante :

$$f = \sum_{j,l,\mathbf{k}} c_{j,l,\mathbf{k}} \varphi_{j,\mathbf{k},l}(x, y) \tag{2.49}$$

où les coefficients $c_{j,l,\mathbf{k}}$ sont obtenus via l'équation (2.48). Les vecteurs (ou fonctions) composant la trame de curvelettes possèdent, en plus du paramètre d'échelle j et du paramètre de position $\mathbf{k}$, un paramètre d'orientation l. La redondance de la transformée en curvelettes dépend du nombre d'orientations choisies et de la manière de calculer les coefficients à l'échelle la plus fine. Dans notre cas, cette redondance est de l'ordre de 7 ce qui signifie qu'on utilise 7 fois plus de coefficients pour coder le même signal.

Remarque 2.16

Contrairement à la transformée en ondelettes qui s'effectue en $O(N)$ opérations (où N est le nombre de pixels de l'image), la transformée en curvelettes s'effectue en $O(N \log(N))$ opérations. Il y a donc un coût supplémentaire induit par l'utilisation d'une telle transformée.

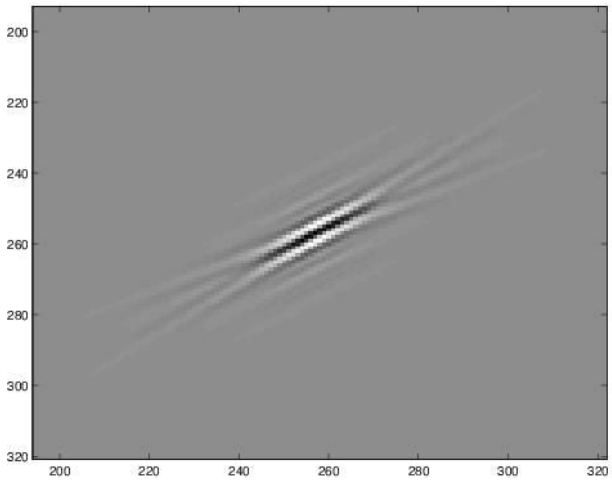


FIGURE 2.13 – Représentation d'un élément de la trame de curvelettes.
Source : http://www.curvelet.org/

3

Utilisation d'images en assimilation de données

Sommaire

3.1 Introduction **41**
3.1.1 Atmospheric motion vectors 42
3.1.2 Estimation du champ de vitesse par mesure du flot optique 43
3.1.3 Assimilation directe d'image 45
3.2 Méthode d'assimilation utilisée dans ce manuscrit **48**
3.2.1 Principe 48
3.2.2 Mesure de la distance entre deux images 49
3.3 Expériences jumelles avec des données parfaites **56**
3.3.1 Contexte expérimental 56
3.3.2 Diagnostic utilisé 59
3.3.3 Validation des opérateurs 61
3.3.4 Influence de la longueur de la fenêtre d'assimilation 64
3.3.5 Impact de la régularisation par l'ébauche 67
3.4 Conclusion **70**

3.1 Introduction

Depuis le début des années 1960 et le lancement de TIROS-1, de nombreux satellites ont été placés en orbite autour de la Terre dans le but d'accroître la connaissance des processus physiques gouvernant cette dernière. Depuis le début des années 1970, ces satellites contribuent activement à l'amélioration des prévisions météorologiques et océanographiques. Ils permettent d'accéder à des informations sur les processus se déroulant dans l'atmosphère et à la surface des océans. Actuellement, ces satellites fournissent la majeure partie des données utilisées dans les processus de prévision numérique du temps.

Parmi les informations récoltées, figurent des "photographies" de la planète Terre. Ces images, regroupées en séquences, permettent d'observer l'évolution dynamique d'"objets" météorologiques et océaniques identifiés : des fronts, des nuages, des tourbillons, etc $\cdots$ Par l'observation de telles séquences, l'oeil humain peut évaluer grossièrement le mouvement apparent des différents objets et ainsi accéder à une information sur la dynamique du système.

Cet aspect est favorisé par la densité spatiale et temporelle de ce type de données : toutes les 15 minutes, une image dont la résolution spatiale est de l'ordre du kilomètre (à la verticale du satellite pour le canal haute résolution) est fournie par le satellite METEOSAT. Avec ce type d'observations, il est donc possible d'accéder à une information aussi bien sur la dynamique à l'échelle synoptique qu'à méso–échelle.

Cette haute fréquence et haute résolution permet d'accéder à une information précise mais nécessite cependant de traiter une grande quantité d'information. Pour les processus d'assimilation de données, avoir accès à tant d'information peut être considérée comme un atout, mais induit aussi de nombreuses difficultés.

Une première difficulté provient de la masse d'information fournie en elle–même : intégrer l'ensemble de l'information présente dans ces images reviendrait à multiplier par quelques centaines la quantité d'informations utilisée au cours du processus d'assimilation.

Une seconde difficulté provient directement de la densité spatiale et temporelle de telles observations : l'hypothèse de décorrélation entre les erreurs d'observations n'est en effet plus valide ni en temps, ni en espace.

Actuellement, afin de contourner ces deux difficultés, seul un très faible pourcentage de la totalité des données satellites (à orbite polaire ou géostationnaires) est utilisé par les centres opérationnels. Les images sont en effet sous–échantillonnées (avec une distance d'échantillonnage de l'ordre de la centaine de kilomètres) pour éviter, entre autres, d'avoir à décrire les corrélations d'erreurs d'observations.
Pour mieux exploiter la densité des observations, plusieurs observations peuvent aussi être regroupées en une seule super–observation (en moyennant l'information récoltée sur un certain domaine spatial et/ou temporel) afin de diminuer l'erreur de mesure et espérer supprimer une partie de l'erreur de représentativité.

A l'heure actuelle, les données satellitaires sont principalement utilisées par les centres opérationnels via de la radiance mesurée par le satellite en chaque pixel de l'image. Ces radiances sont liées à des quantités physiques telles que la température de surface, la hauteur d'eau de l'océan, la concentration en chlorophylle, etc...

L'information sur la dynamique du système contenue dans les séquences d'images est aussi utilisée sous forme de pseudo-observations nommées AMVs (atmospheric motion vectors) en météorologie.

3.1.1 Atmospheric motion vectors

Les AMVs sont des champs de vitessse $\boldsymbol{w}(\boldsymbol{x})$ estimés à partir d'une séqence d'images (voir le paragraphe sur le flot optique plus bas). L'évaluation du mouvement apparent est effectuée en extrayant dans un premier temps les cibles, puis en estimant leur déplacement entre trois images successives [Holmlund, 2000] [Schmetz et al., 1993] par des techniques de corrélations.

Une fois ces champs de vitesse estimés, ils peuvent être intégrés comme observations dans le système d'assimilation.

Cependant, à cause de la nature indirecte et de la complexité de la phase de calcul de ces champs, décrire les erreurs associées à ces pseudo–observations n'est pas chose aisée. [Bormann et al., 2003] montrent que les erreurs d'observation associées à ces données présentent des corrélations spatiales importantes sur des échelles d'environ 300 km. Plus récemment, une étude menée par [Bormann et al., 2014] sur des AMVs obtenus à partir de séquences d'images synthétiques (voir figure 3.2) montre qu'en plus des corrélations spatiales, d'importantes corrélations temporelles sont présentes dans les champs d'erreurs. Ces corrélations ont une valeur non négligeable sur des échelles allant de 4 à 8 h. Les corrélations verticales (entre les estimations de ces champs à différentes altitudes) sont aussi importantes. Ces dernières sont non négligeables pour des différences de pression de l'ordre de 100 hPa.

A cause des corrélations (non prises en compte), ces observations subissent le même traitement ques les autres données satellitaires : un sous–échantillonnage spatial et temporel sévère pouvant être combiné avec la création de super–observation. Bien que les images utilisées pour dériver ces

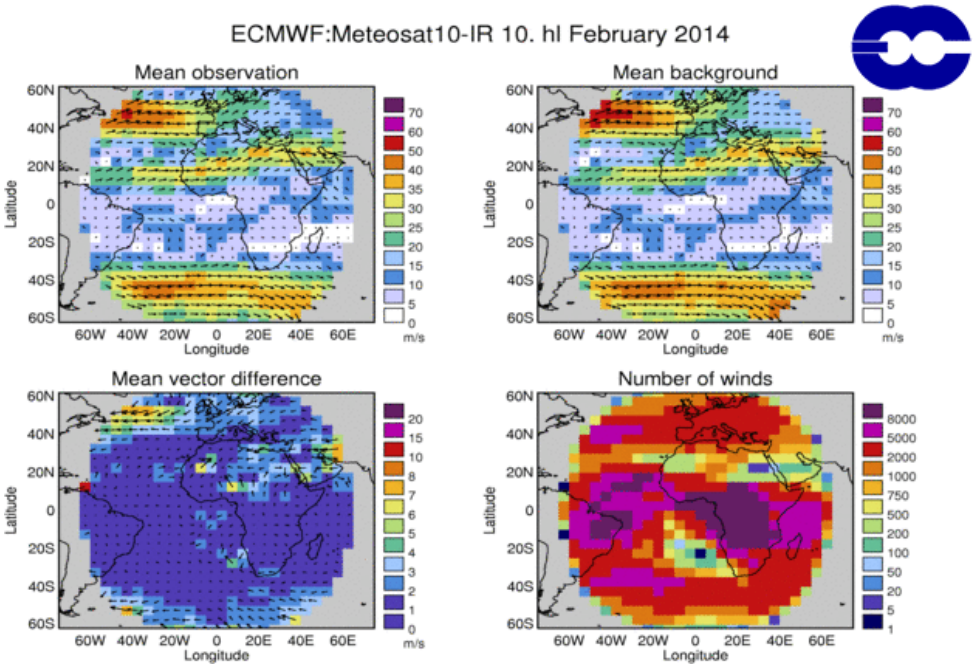


FIGURE 3.1 – Champ de vitesse moyen sur un mois utilisé à l'ECMWF. Source : MetOffice http://research.metoffice.gov.uk/research/interproj/nwpsaf/satwind_report/index.html

champs de vitesse aient une résolution de l'ordre de 4 km, l'échantillonnage effectuée consiste à conserver une information sur le champ de vitesse tout les 100 à 200 km selon les centres opérationnels [6].

Concernant l'échantillonnage temporel, ce dernier est moins important et est globalement d'environ 2 h [Forsythe et al., 2008] (quand une image est disponible toutes les 15 minutes). Les corrélations spatiales et temporelles au sein de ces pseudo–observations n'étant pas prises en compte dans le processus d'assimilation, ces observations sont assimilées avec un faible poids (leurs variances subissent une inflation).

En général les AMVs sont considérés comme très utiles, notamment car ils permettent d'obtenir une information sur les champs de vitesse atmosphériques au–dessus des océans (zones pauvres en observations conventionnelles). Cependant, comme montré dans [Cardinali, 2009], de telles observations peuvent, dans certains cas, avoir un impact négatif sur l'analyse, à cause notamment de la présence des corrélations et/ou de la difficulté à assigner la bonne altitude au champ de vitesse estimé.

En considérant le coût du système d'observations satellites (le coût de la fabrication du satellite METEOSAT de troisième génération est estimé aux alentours de 1.26 milliards d'euros) et des infrastructures nécessaires à la récolte des données, augmenter leur impact dans les systèmes de prévision numérique est un sujet de recherche important.

3.1.2 Estimation du champ de vitesse par mesure du flot optique

Il existe un nombre important de méthodes permettant l'estimation du mouvement apparent. Ces dernières sont connues sous le nom de méthodes de flot optique. Les méthodes par corrélation utilisées pour créer les AMVs en font partie.

Cependant, d'autres méthodes basées sur la minimisation de fonctionnelles peuvent être envisa-

6. Plus d'informations centre par centre, satellite par satellite, peuvent être trouvées sur http://research.metoffice.gov.uk/research/interproj/nwpsaf/satwind_report/index.html

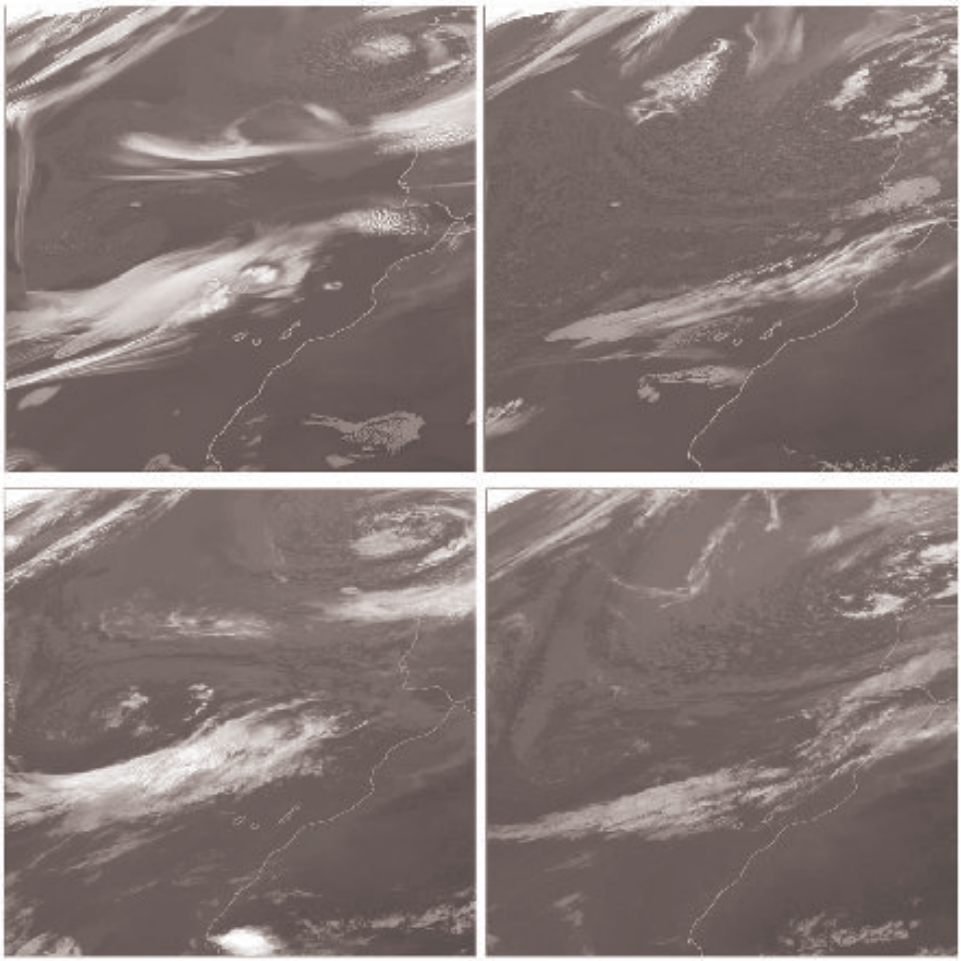


FIGURE 3.2 – En haut, l'image simulée en début de fenêtre et en fin de fenêtre. En bas, les images satellites correspondantes. Source : [Bormann et al., 2014].

gées. Ces méthodes sont souvent basées sur l'équation de contrainte du mouvement apparent :

$$\frac{dI(\boldsymbol{x},t)}{dt} = \frac{\partial I(\boldsymbol{x},t)}{\partial t} + \nabla I(x,t)\boldsymbol{w}(x,t) = 0. \tag{3.1}$$

où $\boldsymbol{w}(x)$ est le champ de vitesse expliquant le mouvement apparent entre deux images successives.

Remarque 3.1

On suppose ici que la luminance $I(\boldsymbol{x},t)$ évolue de façon continue dans le temps et dans l'espace.

Les observations étant discrètes en temps, une version discrétisée, nommée équation de contrainte du flot optique, est utilisée en pratique :

$$I(x,t+\Delta t) - I(x,t) + \boldsymbol{w}(x,t).\nabla I(x,t)\Delta t = 0 \tag{3.2}$$

Le problème rencontré lors de l'estimation du mouvement apparent réside dans le fait que le problème est sous–déterminé : il y a deux fois plus d'inconnues (deux composantes de la vitesse par pixel) que de données (une valeur par pixel). De plus, ce système d'équation fournit uniquement une information sur le mouvement dans la direction du gradient de l'image. Ce problème est connu comme étant le problème d'ouverture.

Remarque 3.2

Dans les zones de l'image pour lesquelles la luminance est constante, aucune information sur

le déplacement ne peut être acquise. Ceci n'est pas spécifique à cette méthode mais à l'information utilisée (image lisse), qui ne peut pas permettre d'évaluer l'impact du mouvement sous-jacent.

Afin de contourner ce problème d'ouverture, il est nécessaire d'apporter une information supplémentaire lors de l'estimation du champ de vitesse :

— On peut, comme ceci est fait pour les AMVs, découper l'image initiale et l'image finale en différentes petites zones puis chercher le mouvement permettant d'appareiller au mieux ces différentes zones.

— Une autre méthode consiste à ajouter une contrainte de régularité sur le champ de vitesse recherché. Historiquement, ce type de contraintes a été introduit par [Horn and Schunck, 1981] en ajoutant un terme pénalisant les variations de vitesse à l'équation (3.2). La fonctionnelle à minimiser est alors définie par :

$$J(\boldsymbol{w}) = \int_{\Omega} \|(\nabla I.\boldsymbol{w} + I_t)\|^2 dxdy + \alpha R(\boldsymbol{w}) \tag{3.3}$$

où R est le terme de régularisation imposé au champ de vitesse. Le terme de régularisation dépend du problème considéré. Définir le poids α est loin d'être une chose aisée. Une bonne review concernant les différents termes de régularisation utilisés est présentée dans [Auroux and Fehrenbach, 2011].

Remarque 3.3

D'autres fonctionnelles peuvent être minimisées pour définir le mouvement apparent. Dans [Auroux and Fehrenbach, 2011], une version intégrée de l'équation (3.2) est utilisée :

$$J(\boldsymbol{w}) = \int_{\Omega} \|(I_1((x,y) + \boldsymbol{w}(x,y)) - I_0(x,y))\|^2 dxdy + \alpha R(\boldsymbol{w}) \tag{3.4}$$

— La fonctionnelle peut aussi être minimisée sur des fenêtres locales [Lucas and Kanade, 1981] en supposant la vitesse constante sur chaque fenêtre. Dès lors, pour chaque pour chaque position $\boldsymbol{x}$, la fonctionnelle suivante est minimisée :

$$\forall \boldsymbol{x} \ \hat{\boldsymbol{w}}(\boldsymbol{x}) = \underset{\boldsymbol{w}}{\operatorname{argmin}} \int_{\Omega} g_{\lambda}(\boldsymbol{x} - \boldsymbol{x}') \left(\boldsymbol{w}.\nabla I(\boldsymbol{x}') + \frac{\partial I(\boldsymbol{x}')}{\partial t} \right)^2 d\boldsymbol{x}' \tag{3.5}$$

où g_{λ} est une fonction gaussienne d'écart type λ. Le réglage à effectuer avec cette technique consiste à estimer localement la bonne taille de la fenêtre à considérer.

Le mouvement estimé par ces différentes méthodes peut, comme pour les AMVs, être utilisé en tant que pseudo–observation comme proposé par exemple dans [Papadakis and Mémin, 2007].

Cependant, ces champs de vents estimés posent le même type de problème que les AMVs : décrire les corrélations au sein du champ d'erreurs associé à ces (pseudo)observations de la vitesse est loin d'être un exercice trivial à cause du processus ayant mené à l'estimation de ces derniers.

3.1.3 Assimilation directe d'image

Afin d'éviter les problèmes liés à l'emploi de pseudo–observation, des méthodes intégrant de manière directe l'information sur la dynamique du système contenue au sein des séquences d'images ont été proposées ces dernières années. Ces méthodes consistent à corriger directement le champ de vitesse durant la phase d'assimilation. Le pré-traitement des données est ainsi évité afin de ne pas introduire de corrélations difficilement modélisables.

L'équation de contrainte du flot optique comme opérateur d'observation

[Papadakis and Mémin, 2008] proposent une nouvelle approche pour prendre en compte ce type d'information. L'idée consiste à lier directement le champ de vitesse, utilisé dans le modèle, aux images observées. Ceci est fait au moyen de l'opérateur d'observation suivant :

$$\left\{\begin{array}{ll} \left.\begin{array}{l} \dfrac{\partial \boldsymbol{X}(\boldsymbol{x},t)}{\partial t} = \mathcal{M}(\boldsymbol{X}(\boldsymbol{x},t)) \\ \boldsymbol{X}(\boldsymbol{x},0) = \chi_0 \end{array}\right\} & \text{Modèle} \\ \\ \left.\|Y - \mathbb{H}(\boldsymbol{X})\|^2 = \left\| g_\lambda \star \left(\dfrac{\partial I^o(\boldsymbol{x},t)}{\partial t} + \boldsymbol{w}(\boldsymbol{x},t).\nabla I^o(\boldsymbol{x},t) \right) \right\|^2 \right\} & \text{Opérateur d'observation} \end{array}\right. \tag{3.6}$$

où la variable $\boldsymbol{X}$ représente le vecteur d'état (comprenant, entre autres, le champ de vitesse $\boldsymbol{w}$) et I^o est une image observée.

Remarque 3.4

Le terme d'écart aux observations est donc basé sur la formulation de type Lukas & Kanade du problème d'estimation du flot optique (équation (3.5)).

Les auteurs montrent que les champs analysés produits par cette assimilation directe s'avèrent de qualités bien supérieures à ceux dérivant de l'utilisation de pseudo–observation obtenue par la mesure du flot optique. Ce formalisme a aussi été utilisé avec succès dans [Corpetti et al., 2009] afin d'assimiler des cartes de différence de pressions avec un modèle atmosphérique simplifié.

Récemment, un autre opérateur d'observation a été utilisé dans [Beyou et al., 2013a]. Cet opérateur utilise une formulation stochastique multi-échelles proposé dans [Corpetti and Mémin, 2012]. Cette version, en plus de la possiblité de considérer un champ de vitesse sur plusieurs résolutions, permet d'incorporer un opérateur $\mathcal{F}(I)$ dans l'équation (3.5) afin de modéliser l'incertitude sur la position des pixels. Cette approche est utilisée avec un filtre de Kalman de transformation d'ensemble pondéré [Beyou et al., 2013b] et permet d'incoporer une incertitude sur le transport au sein de l'assimilation.

Pour de plus amples explications sur cette approche, le lecteur peut se réferer aux articles précédents ou aux thèses de [Papadakis, 2007, Beyou, 2013].

Evolution de la quantité observée décrite dans le modèle

Dans [Vidard et al., 2008], les auteurs proposent d'utiliser une autre approche en ajoutant un champ afin de générer une séquence d'images synthétiques comparables aux images observées.

$$\left\{\begin{array}{ll} \left.\begin{array}{l} \dfrac{\partial \boldsymbol{X}(\boldsymbol{x},t)}{\partial t} = \mathcal{M}(\boldsymbol{X}(\boldsymbol{x},t)) \\ \boldsymbol{X}(\boldsymbol{x},0) = \boldsymbol{X}_0 \end{array}\right\} & \text{Modèle} \\ \\ \left.\begin{array}{l} I^{synth}(\boldsymbol{x},0) = I_0^{synth} \\ \dfrac{\partial I^{synth}(\boldsymbol{x},t)}{\partial t} + \boldsymbol{w}(\boldsymbol{x},t).\nabla I^{synth}(\boldsymbol{x},t) = 0 \\ \\ \|Y(\boldsymbol{x},t) - H(\boldsymbol{X}(\boldsymbol{x},t))\|^2 = \|I^{synth}(\boldsymbol{x},t) - I^{obs}(\boldsymbol{x},t)\|^2 \end{array}\right\} & \text{Opérateur d'observation} \end{array}\right. \tag{3.7}$$

Ici, seul le vecteur d'état $\boldsymbol{X}$ comprenant notamment le champ de vitesse $\boldsymbol{w}$ est contrôlé. L'image synthétique initiale I_0^{synth} transportée par la dynamique du modèle ne l'est pas.

L'idée d'effectuer la comparaison entre l'image observée et son équivalent modèle dans un autre espace est introduite dans cet article. Par la suite, cette idée est développée plus en détails dans [Titaud et al., 2010, Souopgui, 2010].

Le même type d'approche a été employé dans les travaux de [Huot et al., 2010, Herlin et al., 2012]. Ces auteurs montrent, entre autres, qu'il est possible d'introduire des erreurs modèles afin de gérer la non–consistance en temps du mouvement au sein des images observées.

Remarque 3.5

Plus récemment, les auteurs de [Lepoittevin et al., 2013] proposent d'ajouter une carte de distance ϕ_s aux contours d'un objet suivi au cours d'une séquence. Cette carte de distance est, comme l'image synthétique, transportée par la dynamique du système. Un terme supplémentaire est alors ajouté à l'opérateur d'observation $\|\phi_s - \phi_{obs}\|^2$ basé sur la distance entre la carte synthétique et une carte extraite à l'instant d'observation. Cette technique semble permettre de suivre convenablement des objets au sein de séquences au prix d'une extraction de leur contour à chaque observation.

Comparaison entre image et dynamique

Les deux premiers types d'approches développés plus haut sont basés sur le concept du flot optique. Cependant, un autre type d'approche comparant directement les images observées à la dynamique du système a été proposé dans [Titaud et al., 2011] et poursuivi dans [Gaultier et al., 2012, Gaultier, 2013].

L'idée consiste ici à extraire certains éléments d'une carte de la dynamique puis à les comparer directement avec les images observées.

En effet, en océanographie, en observant un certain type de cartes décrivant la dynamique du système et des observations de température de l'eau ou de cholorophylle, on remarque des structures "communes" (voir figure 3.3).
La carte à droite est une carte, composée d'exposants de Lyapunov en temps fini (FTLE) ([D'Ovidio et al., 2009]), obtenue par une intégration de la position de particules ajoutées au modèle. Pour calculer ces cartes, l'idée consiste à perturber la position des particules, puis à mesurer le déplacement et l'amplitude de la déviation après une intégration finie en temps.
L'image de gauche est quant à elle une image de la SST.

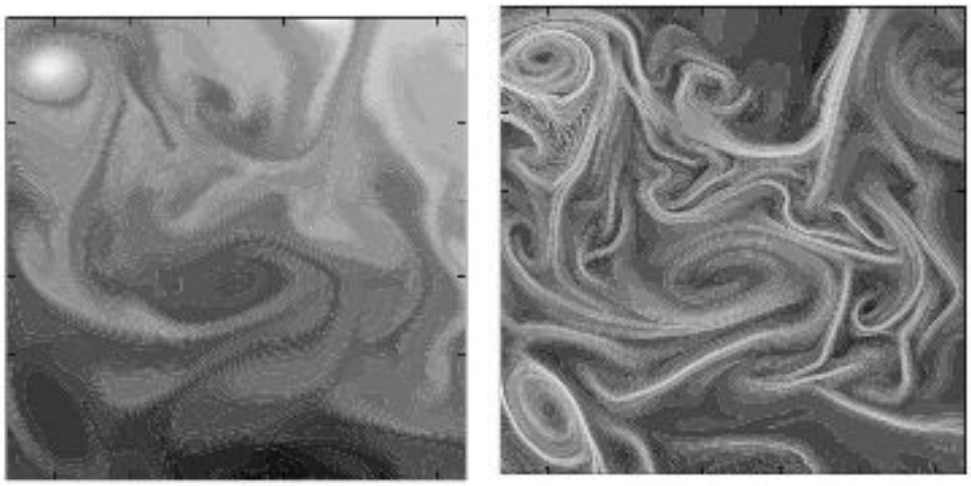


FIGURE 3.3 – A gauche, une image de la SST, à droite une image de FTLE. L'idée développée dans [Titaud et al., 2011] est d'effectuer directement la comparaison entre ces deux quantités.

Les structures de ces deux quantités se ressemblant fortement, l'idée est de les comparer afin d'incorporer de l'information sur la dynamique du système grâce à l'utilisation d'images de SST ou de chlorophylle.

Cependant, ces quantités ne décrivant pas les mêmes variables et n'étant pas directement comparables, le processus d'assimilation repose sur la binarisation de la carte des exposants de Lyapunov et des gradients de l'image : la position des principaux gradients est comparée à la position des principaux exposants de Lyapunov.

Cette binarisation rend l'intégration de cette technique dans un système d'assimilation de données complexe car elle implique un processus non différentiable et dégrade fortement l'information utilisé. De plus, associer des statistiques d'erreurs d'observations à la binarisation des gradients d'une image n'est pas forcément une chose aisée. A vrai dire, certains pourraient même en faire des cauchemars tant les statistiques d'erreurs associées à un tel opérateur d'observation ne semblent pas pouvoir être obtenues de manière triviale.

Autres usages d'images en assimilation de données

D'autres idées ont été développées en assimilation d'images. Par exemple, dans [Makris and Papadakis, 2012], une assimilation en deux temps est réalisée. Une distinction est faite entre erreurs de positionnement et erreurs d'intensités en introduisant de nouvelles variables représentant les erreurs de positions de l'état. Cette méthode consiste à recaler dans un premier temps les structures de l'état par rapport aux observations puis à corriger l'intensité des variables avec un processus classique d'assimilation.

Cependant, l'assimilation d'images n'est pas utilisée uniquement dans le but de corriger le champ de vitesse initial d'un système. Ces méthodes peuvent aussi servir à suivre certaines structures présentes dans les images [Papadakis, 2007, Thomas et al., 2010]. L'utilisation de la connaissance de la position de ces structures en tant que pseudo-observation se révèle cependant une information complexe à intégrer dans les systèmes de prévision numérique du temps [Michel, 2011].

3.2 Méthode d'assimilation utilisée dans ce manuscrit

3.2.1 Principe

Le principe de la méthode d'assimilation d'image utilisée est similaire à celle développée dans [Titaud et al., 2010, Souopgui, 2010]. Il repose sur une comparaison **directe** (il n'y a pas création de pseudo observations) entre l'image observée I^o et l'état du modèle $\boldsymbol{x}$. Comme la variable observée n'est pas une variable du vecteur d'état du modèle, la comparaison entre modèle et observations passe par la création d'une séquence d'images synthétiques I^{synth}. Dans ce manuscrit, on suppose disposer d'un opérateur d'observation, noté $\mathcal{H}$, permettant de créer une telle séquence à partir des sorties du modèle.

La figure 3.4 présente un cas concret où l'hypothèse précédente est respectée. A droite est présenté un exemple d'image synthétique étant l'équivalent modèle d'une image observée issue du satellite METEOSAT 9 (à gauche). On observe que bien que se ressemblant fortement, ces deux images ne sont pas identiques (en dehors des zones non observées). Par exemple, la localisation des structures principales présentes au sein de ces images est proche mais ne concorde pas en tout point.

Bien que visuellement la différence entre ces deux images soit facilement perçue, la quantification de l'écart n'est pas chose aisée. En section 3.2.2 différentes manières de le mesurer sont définies.

Remarque 3.6

Bien que présentées uniquement dans le cadre de l'assimilation de données variationnelle, ces

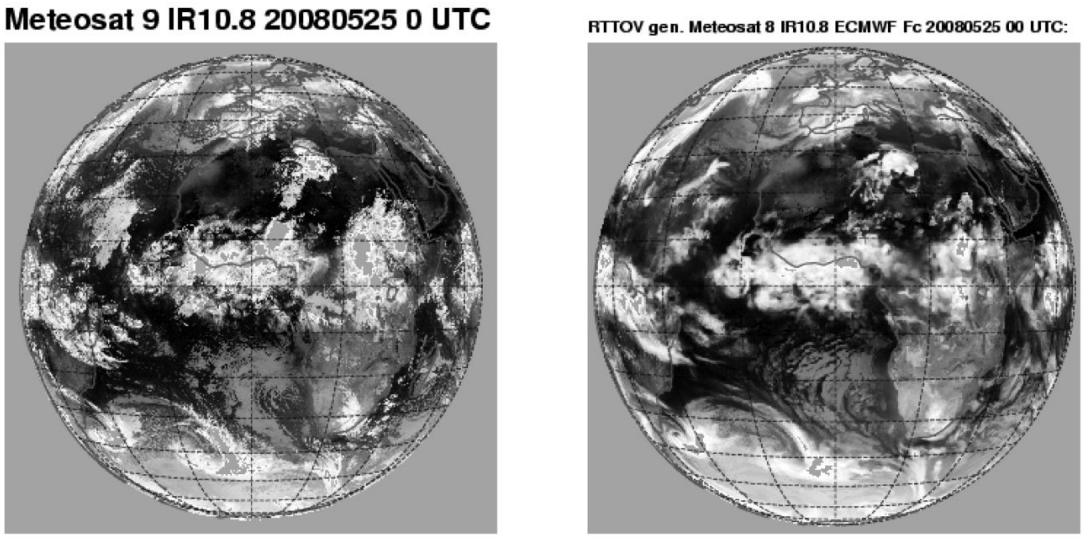


FIGURE 3.4 – A gauche une image issue du satellite METEOSAT 9, à droite un équivalent modèle construit à l'ECMWF de cette image satellite.

manières de définir l'écart entre deux images peuvent être utilisées dans d'autres contextes d'assimilation de données sans modification.

3.2.2 Mesure de la distance entre deux images

Comme nous l'avons vu en introduction de ce chapitre, différentes manières de définir l'écart entre deux images peuvent être mises en oeuvre afin d'utiliser l'information pertinente présente dans les images. Dans cette partie, nous exposons les termes d'écart aux observations utilisés dans l'ensemble du manuscrit.

Distance dans L^2

Cette notion d'écart considère une image sous sa forme brute, à savoir dans l'espace des pixels (chaque pixel fournissant une information dans une région donnée). L'information est utilisée sous une forme non-structurée. Le terme d'écart entre deux images est le suivant :

$$J^o_{L^2}(X_0) = \frac{1}{2}\sum_{t_0}^{t_f} \|I^{synth}(t_i) - I^o\|^2_{\mathbf{R}_{pix}} \tag{3.8}$$

où $\mathbf{R}_{pix}$ est la matrice de covariance dans l'espace des pixels et

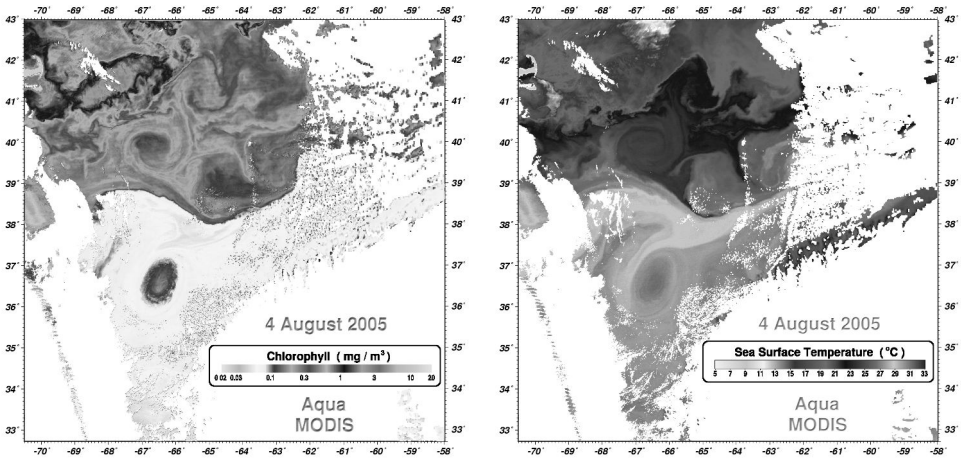
$$\|x\|^2_R = x^T R^{-1} x. \tag{3.9}$$

Lors de l'utilisation de cette distance, chaque pixel observé est comparé à son équivalent modèle, la matrice $\mathbf{R}_{pix}$ décrit les liens existants entre les erreurs aux différentes localisations.

Cette distance possède la très grande qualité d'être aisément mise en place. Cependant, elle ne prend pas en compte que les objets comparés sont structurés. Lorsque la matrice de covariance d'erreur d'observation $\mathbf{R}_{pix}$ **n'est pas diagonale**, la connaissance de la structure au sein du champ d'erreurs d'observation est utilisée.

Remarque 3.7

En pratique, au vu du coût induit par l'inversion des matrices de grande taille, les matrices diagonales sont souvent utilisées dans des cas inappropriés privant de fait le minimiseur de la connaissance de la distribution des erreurs.



(a) Concentration en chlorophylle (b) Température de surface

FIGURE 3.5 – Vue du Gulf Stream en la date du 4 Août 2005 par le satellite Aqua (doté du radiomètre MODIS). Les structures observées sur les deux images sont presque identiques malgré la différence de canal d'acquisition (quantités physiques observées). Source : NASA - http://oceancolor.gsfc.nasa.gov.

Pseudométrique dans H^1

Bien que les variables observées ne soient pas les mêmes, les images de la figure 3.5 présentent, au niveau global, des structures proches. Visuellement l'important sur ces images, afin de déduire des informations sur la dynamique du système, n'est pas la valeur prise en chaque pixel mais la position des discontinuités (ou des lignes de fronts) caractérisées par de forts gradients. En considèrant que l'information pertinente au sein de la séquence d'images concernant la dynamique du système est localisée au niveau des discontinuités, une idée consiste à comparer uniquement les gradients 2D de l'image observée $\nabla = [\partial_x \quad \partial_y]^T$ et de l'image synthétique.

Mémo 1 – Norme et seminorme

La fonction p est une norme si elle vérifie les conditions suivantes :

$$p(\alpha u) = |\alpha| p(u) \tag{3.10}$$

$$p(u+v) \leq p(u) + p(v) \tag{3.11}$$

$$p(u) = 0 \Leftrightarrow u = 0 \tag{3.12}$$

Une application p vérifiant uniquement les conditions (3.10) et (3.11) est une semi-norme.

Cette idée conduit à définir la notion d'écart suivante :

$$J^o_\nabla(X_0) = \frac{1}{2} \sum_{t_0}^{t_f} \left\| \nabla I^{synth}(t_i) - \nabla I^o(t_i) \right\|^2_{R_\nabla} \tag{3.13}$$

où R_∇ est la matrice de covariance relative aux erreurs d'observations dans l'espace des gradients.

Avec cette définition de l'écart aux observations, on utilise un espace moins riche en information : la valeur du signal en chaque pixel jugée superflue n'est pas prise en compte. Ne pas utiliser cette information entraîne l'existence d'une infinité de champs de traceurs donnant la même information dans l'espace des gradients. L'écart aux observations défini en équation (3.13) n'est pas une fonction de distance mais une fonction pseudométrique (voir le mémo 2) basée sur une semi-norme (voir le mémo 1).

Remarque 3.8

L'utilisation de fonction pseudométrique (au mieux) pour définir l'écart entre observation et modèle apparait dès lors qu'il est choisi d'utiliser une partie seulement de l'information contenue dans les observations. Utiliser une fonction pseudométrique plutôt qu'une distance n'est pas nécessairement un problème en soi (tant qu'on est conscient qu'une partie de l'information disponible n'est pas utilisée).

L'opérateur linéaire tangent peut être obtenu pour une petite pertubation γ autour de I^{synth} par :

$$\frac{\partial H_\nabla}{\partial \gamma}(q) = \nabla\gamma \tag{3.14}$$

tandis que l'opérateur adjoint appliqué à un vecteur ρ est :

$$(\partial_q H_\nabla)^* \rho = -\nabla.\rho \tag{3.15}$$

Remarque 3.9

La semi-norme dans l'espace des gradients, peut être combinée à celle dans l'espace des pixels donnant la fonctionnelle suivante à minimiser :

$$J^o_{H^1}(X_0) = \frac{1}{2}\sum_{t_0}^{t_f}\left(\left\|\nabla I^{synth} - \nabla I^o\right\|^2_{\mathbf{R}_\nabla} + \sum_{t_0}^{t_f}\left\|I^{synth} - I^o\right\|^2_{\mathbf{R}_{pix}}\right)$$

pouvant être perçue comme la distance de Mahalanobis entre deux images dans H^1. Un défaut de cette définition de l'écart entre deux images est la non prise en compte de la redondance des erreurs d'observation (les erreurs dans l'espace des gradients sont par construction liées aux erreurs dans l'espace des pixels).

Mémo 2 – Distance et fonction pseudométrique

On appelle distance sur V une fonction $d : V \times V \mapsto \mathbb{R}$ vérifiant les propriétés suivantes :

$$d(u,u) = 0 \tag{3.16}$$
$$d(u,v) = d(v,u) \tag{3.17}$$
$$d(u,v) \leq d(u,z) + d(z,u) \tag{3.18}$$
$$d(u,v) = 0 \Leftrightarrow u = v \tag{3.19}$$

Une fonction d ne vérifiant pas la dernière condition est nommée fonction pseudométrique.

Pseudométrique entre gradients normalisés

Il est aussi possible de se concentrer uniquement sur la direction du gradient des structures (aux endroits où les gradients ont une importance), ce qui conduit à éliminer une partie de l'information contenue dans l'espace des gradients,.
Cette idée est développée en comparant les gradients 2D normalisés $\nabla = [\partial_x \quad \partial_y]^T$ de la séquence observée et de la séquence synthétique. La différence en un point de l'image est alors calculée par :

$$\frac{\nabla I^{synth}}{||\nabla I^{synth}||} - \frac{\nabla I^o}{||\nabla I^o||},$$

où $||\nabla q|| = \sqrt{q_x^2 + q_y^2}$.
L'idée est d'espérer que les lignes de contours des traceurs soient conservées dans l'image observées I^o et ce, même si le processus d'acquisition est hautement non-linéaire (ce qui est le cas en pratique dans les images satellites). De tels processus posent de nombreux problèmes quant à la définition des erreurs de mesure. En utilisant uniquement une information "pauvre" (i-e la direction des gradients) on peut espérer contourner ce problème.

Remarque 3.10

Cette distance peut aussi être utilisée afin de mesurer la distance entre deux choses non comparables (par ex. concentration en chlorophylle et température de l'eau) dont les structures ont (supposément) des caractéritiques communes (dont la position et l'orientation). Elle pourrait par exemple être employée pour mesurer la distance entre les images de chlorophylle et les images de FTLE extraites des champs de vitesses (voir figure 3.3).

Dans le but d'éviter un problème mal posé pour les vecteurs nuls, deux idées s'offrent à nous :

1. Définir l'écart aux observations par :

$$\frac{\nabla I^{synth}}{||\nabla I^{synth}||_\epsilon} - \frac{\nabla I^o}{||\nabla I^o||_\epsilon},$$

où $||\nabla q||_\epsilon = \sqrt{q_x^2 + q_y^2 + \epsilon^2}$ et $\epsilon > 0$ est choisi tel que $\epsilon = ||\nabla I^o||^+/10$, avec :

$$||\nabla I^o||^+ = \min_{\substack{\boldsymbol{x} \in \Omega \\ t \in [t_0, t_f]}} \Big\{ ||\nabla I^o(\boldsymbol{x}, t)||, \text{ tel que } ||\nabla I^o(\boldsymbol{x}, t)|| > 0 \Big\}. \tag{3.20}$$

Cela conduit à l'inégalité suivante quand ∇I^o n'est pas nul :

$$\frac{0.995}{\|\nabla I^o\|} \leq \frac{1}{\|\nabla I^o\|_\epsilon} \leq \frac{1}{\|\nabla I^o\|}.$$

Par conséquent $\dfrac{\nabla I^o}{||\nabla I^o||_\epsilon}$ est soit nul, soit presque égal au vecteur unitaire $\dfrac{\nabla I^o}{||\nabla I^o||}$.

2. Utiliser l'information sur la direction des gradients de l'image uniquement lorsque la valeur du gradient est suffisament importante.

En fin de compte, la fonction d'écart aux observations est définie par :

$$J^o_{\mathbf{Ang}}(X_0) = \frac{1}{2} \sum_{t_0}^{t_f} \left|\left| \frac{\nabla I}{||\nabla I||_\epsilon} - \frac{\nabla I^{synth}}{||\nabla I^{synth}||_\epsilon} \right|\right|^2_{\mathbf{R_{Ang}}} \tag{3.21}$$

où $\mathbf{R_{Ang}}$ est la matrice de covariance relative aux erreurs d'observations. L'opérateur linéaire tangent $\partial_{I^{synth}} \mathcal{H}_{\mathbf{Ang}}$ peut être obtenu pour une petite perturbation γ autour de I^{synth} par :

$$\partial_{I^{synth}} \mathcal{H}_{\mathbf{Ang}} \gamma = \frac{1}{||\nabla I^{synth}||^3_\epsilon} \underbrace{\begin{bmatrix} \left(I_y^{synth}\right)^2 + \epsilon^2 & -I_x^{synth} I_y^{synth} \\ -I_x^{synth} I_y^{synth} & \left(I_x^{synth}\right)^2 + \epsilon^2 \end{bmatrix}}_{C(I^{synth})} \nabla \gamma. \tag{3.22}$$

Les valeurs propres de la matrice $C(I^{synth})$ sont ϵ^2 et $||\nabla I^{synth}||^2 + \epsilon^2$, respectivement associées aux vecteurs propres ∇I^{synth} et $\nabla^\perp I^{synth}$. Par conséquent, si $||\nabla I^{synth}||$ n'est pas nul, la majeure partie de l'information est transmise dans la direction orthogonale au gradient $\nabla^\perp I^{synth}$. Dans l'autre cas, cet opérateur réalise simplement une diffusion isotrope.

L'opérateur adjoint appliqué à un vecteur u s'écrit :

$$(\partial_{I^{synth}}\mathcal{H}_{\text{Ang}})^* u = -\nabla \cdot \left(\frac{1}{||\nabla I^{synth}||_\epsilon^3} \begin{bmatrix} \left(I_y^{synth}\right)^2 + \epsilon^2 & -I_x^{synth} I_y^{synth} \\ -I_x^{synth} I_y^{synth} & \left(I_x^{synth}\right)^2 + \epsilon^2 \end{bmatrix} u \right). \tag{3.23}$$

Une fois encore, il existe une infinité de champs de traceurs dont les gradients partagent la même direction. Cependant, la fonction $d(u,v) = \left\| \frac{\nabla u}{\|\nabla\|_\epsilon} - \frac{\nabla v}{\|\nabla\|_\epsilon} \right\|$ est bien une pseudométrique. En effet :

$$\begin{aligned}
d(u,v)^2 &= \left\langle \frac{\nabla u}{\|\nabla u\|_\epsilon} - \frac{\nabla v}{\|\nabla v\|_\epsilon} , \frac{\nabla u}{\|\nabla u\|_\epsilon} - \frac{\nabla v}{\|\nabla v\|_\epsilon} \right\rangle \\
&= \left\langle \frac{\nabla u}{\|\nabla u\|_\epsilon} - \frac{\nabla z}{\|\nabla z\|_\epsilon} + \frac{\nabla z}{\|\nabla z\|_\epsilon} - \frac{\nabla v}{\|\nabla\|_\epsilon} , \frac{\nabla u}{\|\nabla\|_\epsilon} - \frac{\nabla z}{\|\nabla z\|_\epsilon} + \frac{\nabla z}{\|\nabla z\|_\epsilon} - \frac{\nabla v}{\|\nabla\|_\epsilon} \right\rangle \\
&= d(u,z)^2 + d(z,v)^2 + 2\left\langle \frac{\nabla u}{\|\nabla u\|_\epsilon} - \frac{\nabla z}{\|\nabla z\|_\epsilon} , \frac{\nabla z}{\|\nabla z\|_\epsilon} - \frac{\nabla v}{\|\nabla\|_\epsilon} \right\rangle \\
&\leq d(u,z)^2 + d(z,v)^2 + 2\left| \left\langle \frac{\nabla u}{\|\nabla u\|_\epsilon} - \frac{\nabla z}{\|\nabla z\|_\epsilon} , \frac{\nabla z}{\|\nabla z\|_\epsilon} - \frac{\nabla v}{\|\nabla\|_\epsilon} \right\rangle \right| \\
&\leq d(u,z)^2 + d(z,v)^2 + 2d(u,z)d(z,v) = (d(u,z) + d(v,z))^2.
\end{aligned}$$

Remarque 3.11

Cet opérateur permet, en théorie, de corriger uniquement l'orientation des discontinuités du traceur. Afin d'avoir un problème un peu mieux posé, il est envisageable de combiner cette information sur les lignes de niveau avec une information de type valeur du pixel. Dans ce cadre là, l'utilisation des gradients normalisés peut être considérée comme un apport permettant de contraindre plus fortement le terme d'écart aux observations au niveau des discontinuités majeures de l'image observée.

Mémo 3 – Notations pour les matrices de covariance

Espace	Dans l'espace	Dans un sous-espace
Pixels	$\mathbf{R}_{pix}$	
Gradients	$\mathbf{R}_{\nabla}$	
Gradients Normalisés	$\mathbf{R}_{\mathbf{GN}}$	
Transformée multi-échelles	$\mathbf{R}_{\mathbf{F}_{sc}}$	$\mathbf{R}_{\mathcal{T}(\mathbf{F}_{sc})}$

Distances et pseudométriques dans des espaces multi-échelles

Afin de pouvoir manipuler plus aisément l'information présente au sein de l'image, une idée développée dans [Titaud et al., 2010] consiste à considérer l'information non plus au niveau élémentaire (i-e du pixel) mais dans un espace structuré. Les auteurs proposent d'utiliser la transformée en curvelettes (définie dans [Candes et al., 2006]) afin d'extraire les structures principales présentes au sein du signal observé. L'extraction est effectuée via une opération de seuillage telle celle définie en équation (2.36). Cette opération permet de représenter de manière parcimonieuse (seulement un faible pourcentage des fonctions dans la base des structures sont conservées) l'information importante contenue au sein de l'image et de manipuler directement une information structurée et non plus des pixels pris isolément les uns des autres.

Ce processus de décomposition de l'image dans un espace structuré peut être mis en place dans d'autres espaces multi-échelles. Dans ce manuscrit, le choix a été fait, en plus de la décomposition dans une trame ajustée de curvelettes (voir aide 4 et section 2.3.4 pour la notion de trame), de considérer la décomposition dans des bases d'ondelettes orthonormales.

Mémo 4 – Trame

La notion de trame est une généralisation de la notion de base. Soit une séquence de vecteurs $\{e_k\} \in V$. Lorsque l'ensemble des vecteurs $\{e_k\}$ couvre V, on peut exprimer un vecteur $u \in V$ comme une combinaison linéaire de ces vecteurs : $u = \sum_k c_k e_k$. Lorsque l'ensemble $\{e_k\}$ est de plus linéairement indépendant, il forme une base de V. Lorsque cet ensemble est linéairement **dépendant**, il est possible de calculer ces coefficients "naturels" de manière simple lorsqu'il existe A et B tel que $0 < A \leq B < +\infty$ et que $\forall u \in V$

$$A\|u\|^2 \leq \sum_{k \in I} |\langle u|e_k\rangle|^2 \leq B\|u\|^2. \tag{3.24}$$

Dans ce cadre là, l'ensemble $\{e_k\}_{k\in I}$ est appelé une trame. Lorsque $A = B$, ceci est une trame ajustée. Lorsque $A = B = 1$, malgré la redondance d'information due à la représentation du signal dans une trame, la norme L^2 est conservée et la trame est dite de Parseval. C'est le cas pour la transformée en curvelettes utilisée.

Remarque 3.12

Les ondelettes ont l'avantage de former des bases , ce qui est agréable afin de comprendre la différence entre la représentation en ondelettes et en pixels d'une image. La simplicité des espaces d'ondelettes, leurs "bonnes propriétés" (voir section 2.2.3) combinées à un coût numérique faible (en O(N) ou N est le nombre de pixels), en ont fait un objet mathématique particulièrement étudié et utilisé (dans un nombre de domaines important) ces 30 dernières années (ce qui est un avantage en comparaison des curvelettes).

Comme présenté dans le chapitre 2, ces transformées reposent sur le fait qu'une image I peut être décomposée de la manière suivante :

$$I = \sum_{j,\mathbf{k},l} c_{j,\mathbf{k},l}\varphi_{j,\mathbf{k},l}$$

où les fonctions $\{\varphi_{j,\mathbf{k},l}\}_{j,\mathbf{k},l}$ sont des éléments de la trame de curvelettes (ou base d'ondelettes) choisie et $c_{j,\mathbf{k},l}$ les coefficients corresondants. Les paramètres $j, \mathbf{k}, l$ spécifient respectivement l'échelle, la position spatiale (où $\mathbf{k}$ possède une composante par dimension) et l'orientation de chacune des fonctions. Les coefficients sont calculés de la manière suivante :

$$c_{j,\mathbf{k},l} =< I, \varphi_{j,\mathbf{k},l} > . \tag{3.25}$$

Afin d'évaluer la distance entre une image observée et une image synthétique, on peut comparer leur décomposition multi-échelles en utilisant la fonction coût suivante :

$$J^o_{\mathbf{F}_{sc}}(X_0) \quad = \quad \sum_{t_0}^{t_f} ||\boldsymbol{c}^{synth}(t_i) - \boldsymbol{c}^{obs}(t_i)||^2_{\mathbf{R}_{\mathbf{F}_{sc}}}. \tag{3.26}$$

où $(\boldsymbol{c}^{synth} - \boldsymbol{c}^{obs})$ est le vecteur d'écart aux observations entre les coefficients observés $(\boldsymbol{c}^{obs})$ et synthétiques $(\boldsymbol{c}^{synth})$. $\mathbf{R}_{\mathbf{F}_{sc}}$ est la matrice de covariance d'erreur d'observation dans l'espace multi-échelles considéré.

Remarque 3.13

Pour toute image I, la norme L^2 est conservée dans une base d'ondelettes ou dans une trame de curvelettes :

$$\sum_{j,\mathbf{k},l} \|c_{j,\mathbf{k},l}\|^2 = \sum_{i,j} \|I(i,j)\|^2. \tag{3.27}$$

Il y a donc une équivalence entre la distance définie dans L^2 (en équation (3.8)) et celles définies dans ces espaces multi-échelles lorsque lorsque $\mathbf{R}_{pix} = \sigma^2 \mathbf{I_n}$ et $\mathbf{R}_{\mathbf{F}_{sc}} = \sigma^2 \mathbf{I_m}$ (où m est le nombre de fonctions $\varphi_{j,\mathbf{k},l}$).
Lorsque la transformée utilisée est redondante (comme pour les curvelettes), fixer $\mathbf{R}_{\mathbf{F}_{sc}} = \sigma^2 \mathbf{I_m}$ est un choix. En effet, la matrice de covariance dans cet espace n'est en théorie pas inversible (et ne peut donc pas être proportionnelle à l'identité). Plus de détails sur ce choix sont donnés au chapitre 4 en section 4.2.3.

Pour ces choix de matrice de covariance, si l'ensemble de l'information présente dans l'espace multi-échelles est conservé, il n'y a aucune différence (en théorie et en pratique) avec la distance pixel à pixel. Par conséquent, au vu du coût numérique supplémentaire induit par la représentation multi-échelles, il n'y a aucun intérêt à utiliser ce type de représentation dans ces cas là.

Le seuillage comme sélecteur d'information Afin de profiter du système d'écriture parcimonieux dans un espace multi-échelles (lorsque le signal à une échelle donnée est suffisamment lisse), une idée consiste à restreindre la taille de l'espace d'observation en utilisant des techniques de seuillage. Dans ce cadre là, il est considéré que seuls les coefficients ayant une amplitude importante (par rapport à l'amplitude du bruit) sont porteurs d'information.

Ces opérations de seuillage sont particulièrement utilisées en traitement d'images afin de restreindre l'analyse aux coefficients portants la majeure partie du signal. Un coefficient observé entaché d'une erreur $\epsilon_{j,\mathbf{k},l}$ (ie $c^{obs}_{j,\mathbf{k},l} = c^{t}_{j,\mathbf{k},l} + \epsilon_{j,\mathbf{k},l}$) est considéré significatif s'il comporte peu de bruit :

$$|c^{obs}_{j,\mathbf{k},l}| \gg \sigma \tag{3.28}$$

où σ^2 est la variance de $\epsilon_{j,\mathbf{k},l}$. A contrario, si le coefficient est de faible amplitude devant σ, il est considéré que le signal ne peut pas être distingué du bruit. Par conséquent, lors d'une analyse du signal (visant par exemple à extraire les contours d'une image), on peut se dispenser d'utiliser cette information.

Remarque 3.14

Cette opération de seuillage permet dès lors de débruiter partiellement une image (en utilisant seulement les coefficients pour lesquels on est certain de distinguer un signal intéressant) ou tout au moins de compresser l'information "utile" dans un espace de taille réduite.

Pour tenir compte uniquement de l'information pertinente, il est possible de conserver une partie seulement des observations en utilisant le seuillage dur :

$$\mathcal{T}(c^{obs}_{j,\mathbf{k},l}) = \begin{cases} c^{obs}_{j,\mathbf{k},l} & \text{si } |c^{obs}_{j,\mathbf{k},l}| > \tau_{j,k,l} \\ 0 & \text{sinon.} \end{cases} \tag{3.29}$$

Afin de comparer des choses comparables, l'équivalent modèle (i-e le vecteur $\boldsymbol{c}^{synth}$ composé des coefficients $\left\{c^{synth}_{j,\mathbf{k},l}\right\}_{j,\mathbf{k},l}$) est lui aussi considéré dans le sous-espace des coefficients d'observation conservé :

$$\mathbf{P}(c^{synth}_{j,\mathbf{k},l}) = \begin{cases} c^{synth}_{j,\mathbf{k},l} & \text{si } |c^{obs}_{j,\mathbf{k},l}| > 0 \\ 0 & \text{sinon} \end{cases} \tag{3.30}$$

où l'opérateur $\mathbf{P}$ est un opérateur de projection. Cet équation traduit le fait qu'un coefficient synthétique $c^{synth}_{j,\mathbf{k},l}$ est conservé si et seulement si son équivalent observé l'est aussi.

Par ce biais, il est possible de construire la pseudométrique suivante dans un espace multi-échelles en ne comparant que les coefficients conservés par le seuillage :

$$J^{o}_{\mathcal{T}(\mathbf{F}_{sc})}(X_0) = \sum_{t_0}^{t_n} \|\mathbf{P}(\boldsymbol{c}^{synth}) - \mathcal{T}(\boldsymbol{c}^{obs})\|^2_{\mathbf{R}_{\mathcal{T}\mathbf{F}_{sc}}} \tag{3.31}$$

Dans ce cadre là, $\mathbf{R}_{\mathcal{T}\mathbf{F}_{sc}}$ est la matrice de covariance d'erreur d'observation dans le sous-espace défini par le seuillage $\mathcal{T}$.

Remarque 3.15

Lorsque τ est choisi assez petit, ce seuillage permet de détruire peu d'informations nécessaires à la reconstruction de l'image observée, tout en réduisant de manière importante le nombre de coefficients porteurs d'informations (dans le sens où ils servent à reconstruire l'image de départ). On défini ainsi une notion d'écart aux observations presque équivalente à celle dans l'espace des pixels mais l'information est compressée sur un petit nombre de coefficients.

D'autres manières de sélectionner l'information peuvent être envisagées dans des bases d'ondelettes. A titre d'exemple on peut, lorsque les données sont à une plus haute résolution que le modèle, choisir de ne conserver que les coefficients à partir d'une certaine échelle (en évitant de ce fait le problème de définir les erreurs liées à l'utilisation d'un opérateur d'observation $\mathcal{H}$ produisant de l'information non disponible).

Remarque 3.16

Le seuillage de coefficients peut aussi jouer un rôle important lorsqu'on cherche à inverser la matrice de covariance d'erreur d'observation. La taille de la matrice à inverser peut être drastiquement réduite dans un espace multi-échelles permettant dès lors d'envisager une inversion directe de cette dernière.

3.3 Expériences jumelles avec des données parfaites

Après avoir introduit le contexte expérimental de ces expériences en section 3.3.1, la validation des différents termes d'écart aux observations introduits aux équations (3.8), (3.13), (3.21) et (3.31) dans un cadre idéalisé est abordée en section 3.3.3. Cette validation est effectuée à l'aide d'expériences jumelles dont le but est de retrouver l'état initial d'un système à l'aide d'observations synthétiques non bruitées et d'un modèle parfait. Après avoir évoqué l'influence de la longueur de la fenêtre d'assimilation en section 3.3.4, l'impact du choix de la régularisation par l'ébauche sur le processus de minimisation est étudié en section 3.3.5.

Remarque 3.17

Cette partie a pour but, avant de passer à des cas plus complexes, de mettre en lumière certains processus et limitations propres au système étudié.

3.3.1 Contexte expérimental

Les expériences effectuées dans cette thèse ont pour cadre général l'étude de la dérive d'un vortex. Elles ont pour source une expérience effectuée à la plateforme CORIOLIS [7] (permettant de simuler en laboratoire la dérive d'un tourbillon). Le nom du dispositif provient du fait qu'il permet de recréer l'effet de la force de Coriolis dans une couche d'eau peu profonde.

Dans l'expérience étudiée, un tour complet de la plateforme dure 60 secondes et simule une rotation de la Terre. Le tourbillon est créé en aspirant l'eau et rendu visible par l'ajout d'un traceur passif (la fluorescine). Pour plus de détails sur la mise en oeuvre et les motivations originales de l'expérience voir [Flór and Eames, 2002].

7. http://www.legi.grenoble-inp.fr/web/spip.php?rubrique175

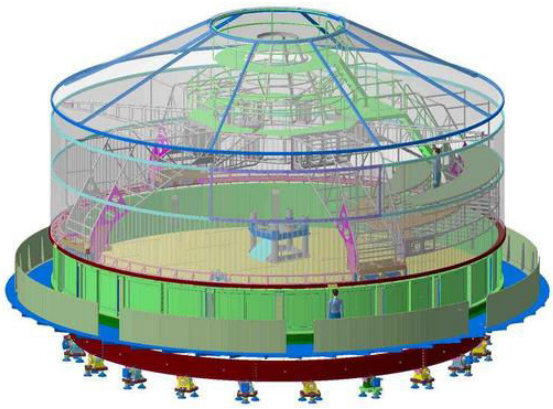


FIGURE 3.6 – Image synthétique de la plateforme Coriolis qui, après 50 années de service, est en cours de reconstruction.

Le modèle

Dans ce cadre–là, l'écoulement, ayant lieu dans une mince couche de fluide, peut être modélisé par les équations de Saint-Venant :

$$\begin{cases} \partial_t u - (f+\zeta)v + \partial_x B &= -ru + \kappa\Delta u \\ \partial_t v + (f+\zeta)u + \partial_y B &= -rv + \kappa\Delta v \\ \partial_t h + \partial_x(hu) + \partial_y(hv) &= 0. \end{cases} \tag{3.32}$$

La variable d'état du modèle est $X(t) = \begin{pmatrix} u(t) & v(t) & h(t) \end{pmatrix}$ où u et v représentent les composantes zonale et méridionale de la vitesse du courant et h représente l'élévation de la surface libre. La vorticité relative ζ est définie par :

$$\zeta = \partial_x v - \partial_y u \tag{3.33}$$

et le potentiel de Bernoulli par :

$$B = gh + \frac{u^2+v^2}{2} \tag{3.34}$$

où g est la gravité. Le paramètre de Coriolis sur le β–plan est donné par :

$$f = f_0 + \beta y. \tag{3.35}$$

Le coefficient de diffusion est noté κ tandis que r est le coefficient de frottement sur le fond.

Dans nos expériences numériques, les valeurs des différents paramètres sont : $r = 9.10^{-7}s^{-1}$, $\kappa = 0m^2s^{-1}$, $f_0 = 0.25s^{-1}$, $g = 9.81m.s^{-2}$ et $\beta = 0.0406m^{-1}.s^{-1}$.

La simulation numérique est effectuée sur un domaine rectangulaire $[0, L]\times[0, H]$ représentant un sous domaine de la plaque tournante de dimension $L = H = 2.525$ m. Le domaine est discrétisé sur une grille carrée (de taille 128×128) uniforme de type C d'Arakawa. Un schéma de différences finies est utilisé pour la discrétisation en espace. L'intégration temporelle est effectuée au moyen d'un schéma de Runge–Kutta d'ordre 4. Le pas de temps considéré est de 0.01 s, ce qui correspond à 14.4 s dans l'atmosphère.

Le code numérique associé au modèle (3.32) et à son adjoint est emprunté à la bibliothèque de code de l'équipe MOISE[8]. Provenant initialement des travaux de thèse d'Arthur Vidard [Vidard, 2001] ces codes ont été utilisés (et enrichis) au cours des travaux de thèse d'Innocent Souopgui [Souopgui, 2010] et du postdoctorat d'Olivier Titaud [Titaud et al., 2010].

8. Equipe projet INRIA au sein de laquelle cette thèse a été préparée.

Méthode de création des images synthétiques

Dans le contexte expérimental décrit précédemment, le traceur est uniquement transporté et diffusé par la dynamique du système. Son évolution peut donc être décrite par l'équation d'advection-diffusion suivante :

$$\begin{cases} \partial_t q + \nabla q \cdot \boldsymbol{w} - \nu_T \Delta q = 0 \\ q(t_0) = q_0. \end{cases} \tag{3.36}$$

où $\boldsymbol{w} = (u, v)$ est le champ de vitesse, ν_T le coefficient de diffusion du traceur et $q(t_0)$ est l'état du traceur au temps initial t_0.

A partir de la concentration initiale du traceur passif $q(t_0)$, il est donc possible de générer une séquence d'images synthétiques, notée I^{synth}, en "couplant" l'équation d'avection–diffusion (3.36) aux équations de Saint-Venant (3.32).

Remarque 3.18

Dans les expériences jumelles effectuées au cours de ce manuscrit, la connaissance parfaite de l'état initial du traceur passif est utilisée. En pratique cette connaissance ne peut être atteinte. Il sera nécessaire de trouver une méthode afin d'initialiser correctement ce traceur passif. Cela peut par exemple passer par le contrôle de la condition initiale du traceur q_0.

Création du scénario

Dans le but d'étudier la qualité des différents opérateurs d'observation nous nous plaçons dans un contexte où la vérité terrain est connue. Cette dernière se compose du champ de vitesse (u_0^t, v_0^t), de la surface libre (h_0^t) ainsi que de la concentration en traceur passif (q_0^t) en début de fenêtre d'assimilation.

La vérité utilisée dans les expériences qui vont suivre a été générée de la manière suivante :

1. Le champ de vitesse a tout d'abord été initialisé via le calcul du déplacement perçu entre 2 images des expériences réalisées à la plate-forme Coriolis. La méthode de flot optique, introduite dans [Souopgui, 2010], a été utilisée pour effectuer cette estimation du mouvement.
2. La concentration de traceur passif est initialisée à l'aide de la première image de traceur tandis que la surface libre est supposée plane (initialisée par une constante $h_{mean} = 0.3553$ en accord avec les paramètres de l'expérience).
3. Une première simulation est réalisée via l'intégration du modèle direct à partir de ces champs. Le coefficient de frottement (r) étant très faible et le coefficient de diffusion, (κ), étant nul les ondes parasites (résultant de l'utilisation d'un état non équilibré) ne sont pas atténuées. Afin de les atténuer, des conditions de frontières ouvertes sont utilisées pour cette simulation.
4. La vérité est définie par l'état équilibré obtenu en fin de simulation.

Cette manière de procéder permet de définir un scénario cohérent (dont les différentes composantes sont présentées en figure 3.7) avec l'initialisation du traceur passif utilisée.

Observations utilisées

La vérité terrain obtenue est utilisée afin de générer une séquence d'images de traceur passif. Ces images constituent les seules observations du système assimilées lors des expériences. L'information sur la dynamique du système est donc obtenue de manière indirecte (via l'équation d'advection– diffusion (3.36)). La figure 3.8 présente un aperçu de ces observations aux temps

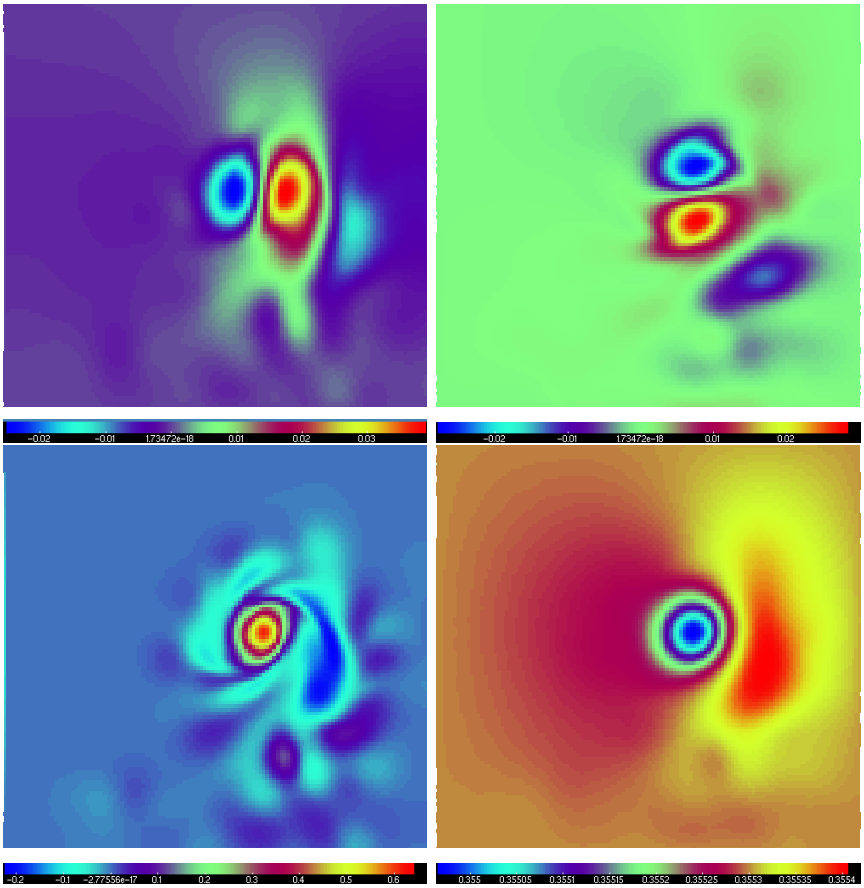


FIGURE 3.7 – Présentation de la vitesse méridionale (en haut à gauche), zonale (en haut à droite) de la vorticité relative (en bas à gauche) et de la hauteur d'eau (en bas à droite) initiale. La vitesse méridionale est comprise entre $-0.025\ m.s^{-1}$ et $0.039\ m.s^{-1}$, la vitesse zonale entre $-0.028\ m.s^{-1}$ et $0.028\ m.s^{-1}$, la hauteur d'eau est presque plate (comprise entre $0.3550\ m$ et $0.3554\ m$) et la vorticité entre $-0.23\ s^{-1}$ et $0.64\ s^{-1}$.

$t = \{0s, 6s, 15s, 30s\}$ (correspondant respectivement à $t = \{0, 2h24, 6h, 12h\}$ en équivalent atmosphère).

On observe qu'au fil du déplacement du vortex vers le nord, les images deviennent de plus en plus lisses (le coefficient de diffusion employé dans l'équation d'advection-diffusion est de $\nu_T = 10^{-5}$).

3.3.2 Diagnostic utilisé

Une fois les expériences réalisées celles-ci doivent être analysées.
Pour ce faire, en plus de l'approche qualitative, nous utilisons comme diagnostic la racine carrrée de l'erreur quadratique moyenne normalisée (par sa valeur pour l'ébauche). Ce diagnostic s'écrit pour une variable $\boldsymbol{x}$ quelconque sous la forme :

$$NRMSE(\boldsymbol{x}) = \frac{RMSE(x)}{RMSE(x^b)} = \sqrt{\frac{\sum_{i,j}(\boldsymbol{x}^t(i,j) - \boldsymbol{x}(i,j))^2}{\sum_{i,j}(\boldsymbol{x}^t(i,j) - \boldsymbol{x}^b(i,j))^2}} \tag{3.37}$$

où x^t est le véritable état de la variable et x^b l'ébauche. Ce dernier est calculé aussi bien pour les composantes zonale (u) et méridionale (v) du champ de vitesse que pour la vorticité (définie en équation (3.33)). Un diagnostic concernant l'erreur angulaire du champ de vitesse (définie dans

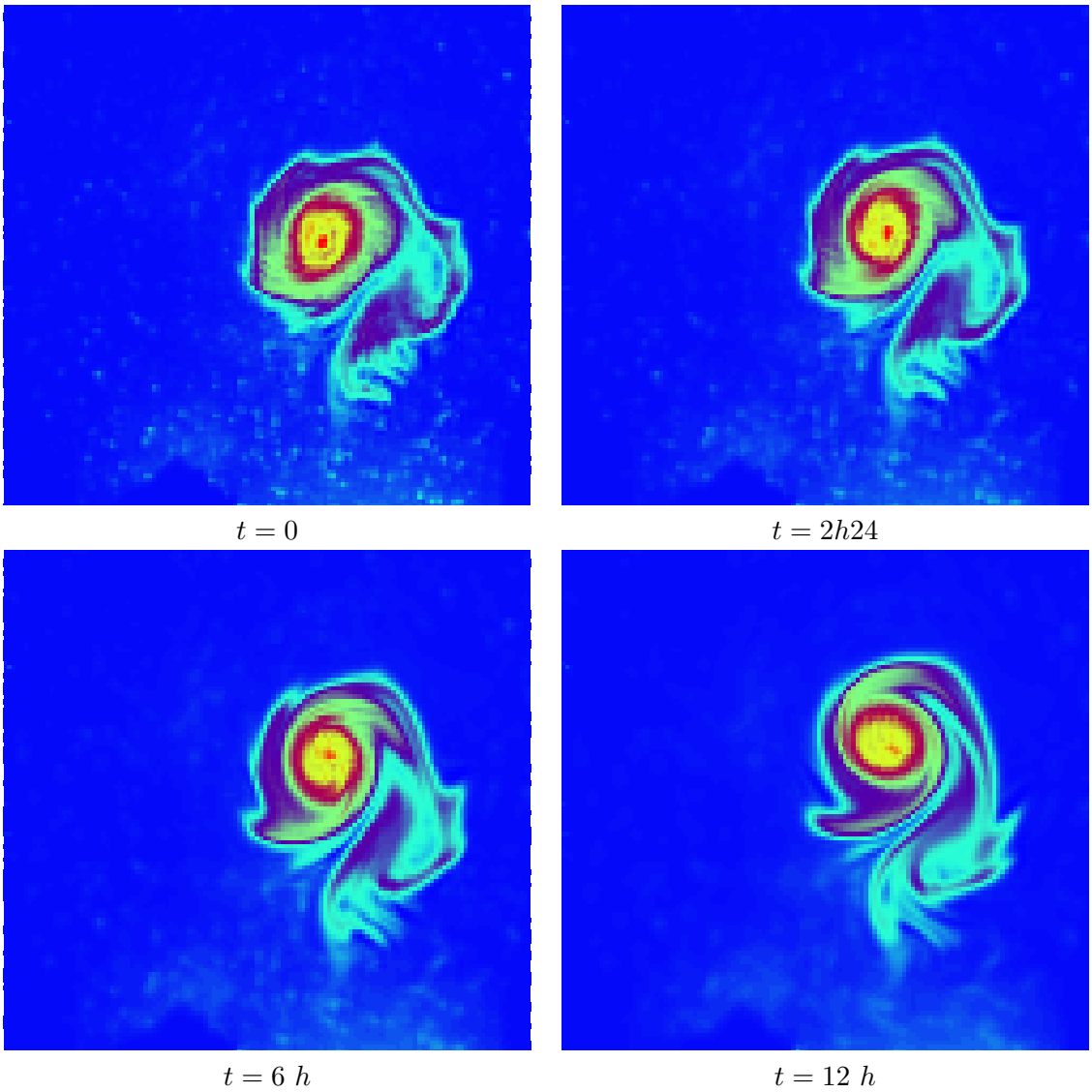


FIGURE 3.8 – Quelques observations de la séquence d'images. La concentration de traceur passif est comprise entre 0.13 (en bleu) et 0.86 (en rouge).

le mémo 5) qui s'écrit :

$$NRMSE(ae) = \sqrt{\frac{\sum_{i,j} \alpha(i,j)^2}{\sum_{i,j} \alpha^b(i,j)^2}}. \tag{3.38}$$

où α^b est l'erreur angulaire de l'ébauche sur le champ de vitesse.

Remarque 3.19

A noter que la RMSE normalisée bien que ne permettant pas de caractériser l'amplitude des erreurs, est pratique à interpréter :

- $NRMSE(\boldsymbol{x}) > 1 \Rightarrow$ *Le champ analysé est moins bon que l'ébauche (selon le diagnostic* $\boldsymbol{x}$*). La phase d'analyse a donc un effet négatif sur le champ* $\boldsymbol{x}$*.*
- $NRMSE(\boldsymbol{x}) \simeq 1 \Rightarrow$ *Le champ analysé est globalement de la même qualité que l'ébauche. La phase d'analyse n'a globalement aucun effet (bien que les erreurs au sein du champ* $\boldsymbol{x}$ *puissent être distribuées d'une manière différente).*
- $NRMSE(\boldsymbol{x}) < 1 \Rightarrow$ *Le champ analysé est globalement meilleur que le champ initial. La phase d'analyse a donc eu un effet positif sur le champ* $\boldsymbol{x}$*.*

Mémo 5 – Erreur angulaire

L'erreur angulaire est définie de la manière suivante :

$$\alpha = \arccos\left(\frac{\langle U_\epsilon, U_\epsilon^t\rangle}{||U_\epsilon||.||U_\epsilon^*||}\right), \tag{3.39}$$

avec $U_\epsilon = [u, v, \epsilon]$, $U^t(\boldsymbol{x}) = [u^t, v^t, \epsilon]$ avec $\epsilon = 5.10^{-4} \max(\sqrt{u^2+v^2})$. Ce terme α correspond à une modification de l'erreur angulaire de [Barron et al., 1994] comme proposé dans [Souopgui, 2010] afin de moins prendre en compte les erreurs dans les zones où la vitesse est très faible par rapport à la vitesse globale de l'écoulement.

Dans les expériences qui vont suivre, certains champs (u,v, vr) sont initialisés par des valeurs nulles (i-e $\boldsymbol{x}^b = 0$). Dans ce cas, le carré du ratio précédent, à savoir :

$$\frac{\sum_{i,j}(\boldsymbol{x}^t(i,j) - \boldsymbol{x}(i,j))^2}{\sum_{i,j}(\boldsymbol{x}^t(i,j))^2} \tag{3.40}$$

s'interprète comme le pourcentage (en terme d'énergie) de l'erreur présente dans le champ $\boldsymbol{x}$.

3.3.3 Validation des opérateurs

Le point de départ du problème de minimisation est un état au repos :

$$(u^b, v^b, h^b) = (0, 0, h_{mean}). \tag{3.41}$$

N'ayant que peu de confiance dans cette ébauche, un terme de poids est introduit dans la fonction coût :

$$J = J^o + \alpha_b J^b \tag{3.42}$$

Choisir le terme α_b assez petit assure que le terme d'écart à l'ébauche, J^b, agisse uniquement en tant que terme de régularisation (et comme préconditionneur) et non comme une fonction de rappel importante vers l'ébauche. Dans les expériences effectuées, $\alpha_b = 10^{-5}$.

But des expériences

Dans un premier temps, nous souhaitons savoir si l'utilisation d'observations parfaites permet de contrôler l'état initial du système. Nous cherchons de plus à comparer le comportement des nouveaux termes d'écart aux observations introduits dans ces travaux (à savoir l'écart entre gradient (3.13), entre gradient normalisé (3.21) et dans un sous-espace d'ondelettes (3.31)) à ceux précédement utilisés dans [Souopgui, 2010, Titaud et al., 2010] (i-e comparaison pixels à pixels et dans un sous-espace de curvelettes).

Observations utilisées

Dans un premier jeu d'expériences, nous nous plaçons sur une fenêtre d'assimilation de 30 s (ce qui correspond à 12 h en équivalent atmosphère et est donc relativement long) et observons une image de traceur passif toutes les 0.25 s (ce qui correspond à une image toutes les 6 minutes en équivalent atmosphère). Au final, une séquence de 120 images parfaites est assimilée. La première (resp. la dernière) est présentée en haut à gauche (resp. en bas à droite) de la figure 3.8. Le nombre d'itérations de minimisation (effectuée à l'aide du solveur N1QN3 [Gilbert and Lemaréchal, 1989]) est limité à 100.

Resultats obtenus

Le tableau 3.1 présente le rapport $\frac{RMSE(x^a)}{RMSE(x^b)}$ pour les différents champs analysés. Ces résultats permettent de quantifier l'amélioration sur chacune des composantes du système imputable à l'assimilation de la séquence d'images de traceurs passifs [9].

Table 3.1 – RMSE normalisées de l'état analysé après 100 itérations de minimisation pour une fenêtre d'assimilation de 12 h. Une image de la concentration du traceur passif est utilisée toutes les 6 minutes.

	u	v	ζ	h	ae
Pixels	5.7%	4.1%	12.1%	87.4%	29.6%
Gradient	6.9%	5.6%	14.6%	90%	31.5%
Gradient Normalisé	15.5%	11.6%	27.3%	90%	32.5%
Ondelettes Seuillées	8.4%	7.0%	15.5%	87.4%	38.3%
Curvelettes Seuillées	7.6%	6.0%	14%	87%	35.7%

De ce tableau, nous pouvons tirer plusieurs enseignements développés dans les paragraphes suivants.

Pertinence de l'approche Le premier point à observer est que, quelque soit le terme d'écart aux observations, l'impact de l'analyse est positif. Cet impact dépend cependant du champ concerné et de la fonction d'écart aux observations considérée. Par exemple, l'erreur commise sur le champ de vitesse est faible en fin d'assimilation (par rapport à l'erreur sur l'ébauche) tandis que celle sur la surface libre reste élevée.
Afin d'avoir une idée sur la signification de ces pourcentages, la figure 3.9 présente la vitesse zonale, la vorticité et la surface libre analysées. Les résultats obtenus lors d'une comparaison entre pixels de l'image sont présentés à gauche, ceux découlant de la comparaison entre gradients normalisés au centre. Afin d'avoir un élément de comparaison, la vérité est rappelée à droite.

Caractéristique de l'analyse dans l'espace des pixels L'utilisation de la distance entre pixels permet de reconstruire le champ de vitesse à la bonne localisation. L'amplitude de la vitesse zonale est bonne.
Il est donc possible d'extraire de ces séquences d'images lors de la phase d'analyse suffisamment d'information pour contrôler de manière correcte le champ de vitesse.
Cependant le vortex reconstruit est quelque peu trop lisse (par rapport au véritable vortex). Ce phénomène est amplifié sur la vorticité analysée. Concernant la surface libre analysée, on observe que cette dernière est de piètre qualité (et le tableau 3.1 nous confirme que sa qualité est équivalente à celle de l'ébauche). Les informations présentes dans ces séquences d'images ne sont donc pas suffisantes (ou tout au moins pas suffisamment bien exploitées) pour permettre à la surface libre analysée d'avoir une qualité acceptable. Plus de détails sur ce fait sont fournis en annexe A.

Comportement de l'écart entre gradients normalisés Concernant la distance entre gradients normalisés, bien que la vitesse zonale reconstruite soit localisée à la bonne position (figure 3.9) son amplitude est sous-estimée. Ce phénomène est aussi valable pour la vitesse méridionale (voir table 3.1). Cette sous-estimation de la vitesse entraîne des variations spatiales de vitesses moins importantes proche du centre du vortex et donc une bien moins bonne vorticité analysée en cet endroit (car cette dernière dépend des variations spatiales de la vitesse, voir (3.33)).

9. On rappelle au passage que cette séquence comporte *indirectement* de l'information sur la dynamique du système.

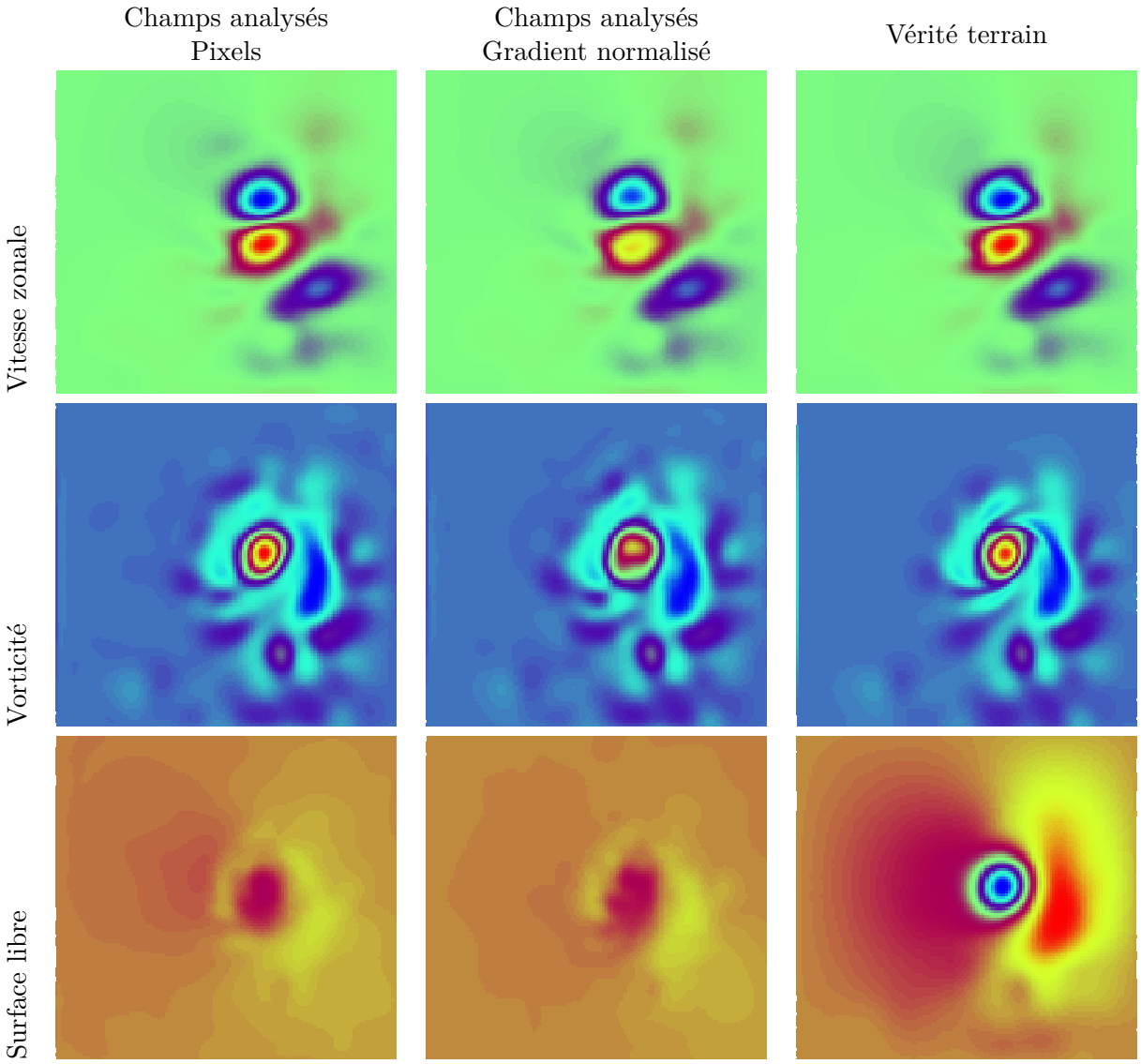


FIGURE 3.9 – Diverses composantes analysées. La dernière colonne présente la vérité.

L'erreur angulaire du champ de vitesse est cependant du même ordre de grandeur que pour les autres termes d'écart aux observations. Ainsi, la direction du champ de vitesse analysé est de qualité identique à celle obtenue avec les autres opérateurs. Seule l'amplitude de la vitesse est donc de moins bonne qualité.

Plusieurs explications peuvent être apportées pour expliquer ces "mauvais" résultats (on arrive tout de même à extraire de l'information avec cet opérateur d'observations). Nous y reviendrons en section 3.3.4.

Impact de la sélection d'information (ou seuillage) On observe dans le tableau 3.1 que la sélection d'information effectuée dans une base d'ondelettes (ou dans une trame de curvelettes) a un faible impact (négatif) sur la qualité de l'analyse. Ainsi, lorsque les données sont parfaites, ne pas prendre en compte certaines informations (comme le fait que le champ de traceur soit lisse en certains endroits) impacte peu l'analyse. Malgré la compression de l'information dans un plus petit espace (la taille du vecteur d'observation est divisé par 10 dans le cas des ondelettes et par 1.5 pour les curvelettes) la qualité du champ de vitesse estimé reste de bonne. Cette capacité de compresser l'information utile sur un jeu de coefficient de petite taille peut ainsi être envisagée pour utiliser (dans des cas réalistes) des images dont la résolution et la fréquence

d'acquisition sont en constante croissance sans saturer d'observations le système d'assimilation de données.

Comparaison entre gradients Les champs analysés reconstruits en utilisant la comparaison entre gradients sont à peine moins bons que ceux obtenus lors de la comparaison entre pixels. Ainsi, ne tenir compte que des discontinuités des images, et non de la valeur en chaque pixel, ne pénalise pas le système d'assimilation lors de l'utilisation de données parfaites.

Bilan 3.1

Un premier bilan de ces expériences montre qu'il est envisageable de contrôler le champ de vitesse et ce, quel que soit le terme d'écart aux observations utilisé (bien que l'écart entre gradients normalisés fournisse de moins bon résultats) à l'aide d'une séquence d'images. Avec des données parfaites et sur des expériences assez longues (mais bien observées) les résultats obtenus sur le champ de vitesse sont de bonne qualité. Cependant la hauteur d'eau a été très peu corrigée.

3.3.4 Influence de la longueur de la fenêtre d'assimilation

But des expériences

Le jeu de données utilisé dans les expériences précédentes, résumées dans le tableau 3.1, suppose l'utilisation d'une fenêtre d'assimilation relativement longue (12 h en équivalent atmosphère) et une fréquence d'observation relativement élevée (correspondant à une observation toutes les 6 minutes). Dans cette partie nous allons regarder les effets du choix de fenêtres d'assimilations plus courtes (2 h et 6 h) et d'une fréquence d'observation moins élevée (une observation toutes les 30 minutes). L'impact de ces choix sur l'analyse effectuée et sur la rapidité de la convergence de la minimisation est étudié dans cette section.

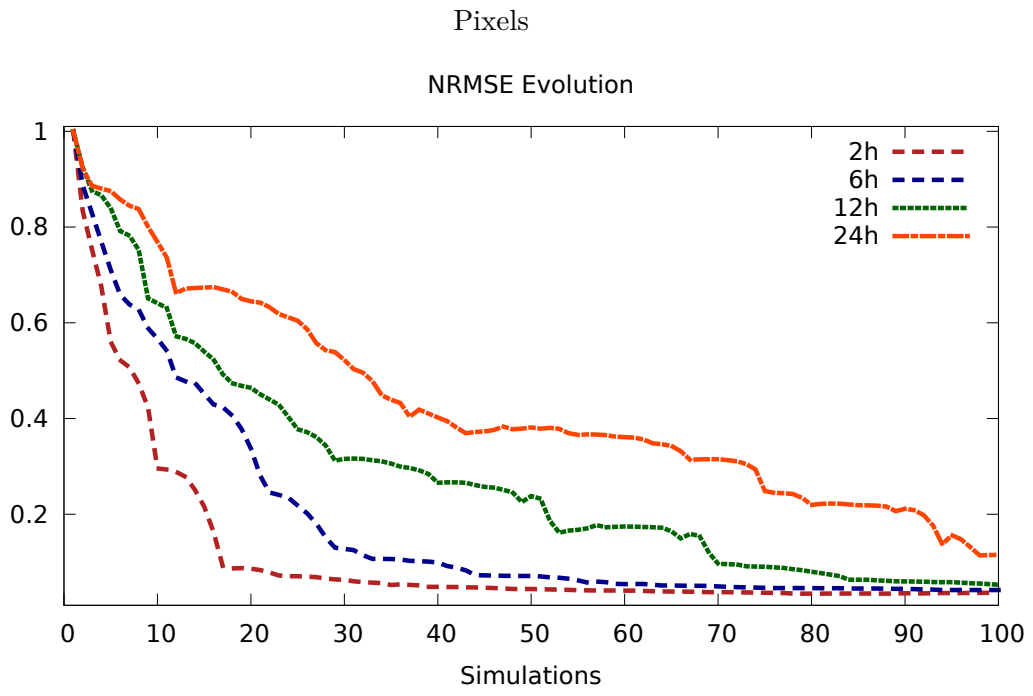


FIGURE 3.10 – Evolution du pourcentage d'erreur résiduelle au fil des itérations pour des fenêtres de 2 h, 6 h, 12 h et 24 h. La variable d'intérêt est la composante méridionale de la vitesse. La distance est calculée dans l'espace des pixels. Le traceur est observé toutes les 30 minutes.

Résultats obtenus

Les résultats obtenus avec ces jeux de données sont présentés uniquement lorsque l'écart aux observations est calculé entre pixels (équation (3.8)) et entre gradients normalisés (équation (3.21)). En effet, les résultats obtenus lors de la comparaison entre gradients (équation (3.13)) et dans un sous-espace d'ondelettes ou de curvelettes (équation (3.31)) sont *similaires*[10]. à ceux utilisant la distance entre pixels.

Analyse pour la distance pixels à pixels

La figure 3.10 présente l'évolution du ratio $\frac{RMSE(v)}{RMSE(v^b)}$ (où v est la composante méridionale du champ de vitesse) au cours des itérations de minimisation pour différentes tailles de fenêtre d'assimilation. La courbe marron correspond à une fenêtre d'assimilation de $2h$ (4 observations sont utilisées), la courbe bleue à $6h$ (12 observations), la courbe verte à 12 h (24 observations) et la courbe orange correspond à une fenêtre de $24h$ (48 observations).

Lorsque la longueur de la fenêtre d'assimilation augmente, bien que le nombre d'observations prises en compte augmente (et donc l'information sur le système étudié augmente), la rapidité de la convergence diminue fortement. Ainsi, lorsqu'une fenêtre de $2h$ est considérée la convergence s'effectue en environ 30 itérations alors que 60 (resp. 80) itérations sont nécessaires pour une fenêtre de $6h$ (resp. $12h$). Concernant la fenêtre d'assimilation de $24h$, au bout de 100 itérations nous ne sommes pas encore arrivés à convergence.

Remarque 3.20

Il a tout d'abord été vérifié que le problème ne vient pas d'un trop grand nombre d'observations fourni au système. Pour cela, des expériences sur des fenêtres d'assimilations de 12 h et de 24 h ont été effectuées en utilisant d'autres fréquences d'observations (une observation toutes les 6 minutes, toutes les heures et toutes les deux heures). Les résultats de ces expériences sont similaires à ceux présentés précédemment.

L'intégration de nouvelles données (parfaites) dans la phase d'analyse dégrade donc cette dernière (ou tout au moins la convergence observée). Ce phénomène provient des non-linéarités présentes dans la fonction coût (dues dans ce cas à la non-linéarité du modèle). De fait, plus la longueur de la fenêtre augmente, plus la fonction coût à minimiser devient non linéaire et plus il est complexe de trouver son minimum. Ainsi, bien qu'ayant fourni plus d'information au processus de minimisation, il est moins facile de trouver un minimum (local ?) de la fonctionnelle.

Analyse pour la comparaison entre gradients normalisés :

Les résultats des expériences réalisées en utilisant la comparaison entre gradients normalisés sur les fenêtres de $2h$, $6h$ et $12h$ sont présentés en figure 3.11.

Dans la section précédente, les résultats obtenus sur des expériences de $12h$ étaient bien moins bons que ceux obtenus avec d'autres opérateurs. Avec le nouveau jeu d'observations, ce phénomène est de nouveau observé pour des expériences de $12h$.

Remarque 3.21

A convergence pour les expériences de $12h$, le ratio $\frac{RMSE(v^a)}{RMSE(v^b)} \simeq 22\%$ ce qui est bien supérieur au 11.6% obtenus avec plus d'observations (le nouveau jeu de données comporte seulement une observation sur 5 du jeu de données précédent). Supprimer des observations a donc un effet négatif lorsque la fenêtre d'assimilation est relativement longue avec ce terme d'écart aux observations.

10. Par similarité, il est entendu que la minimisation a un comportement identique (donc que les courbes obtenues ont la même allure) bien que la valeur à convergence dépende (assez faiblement) de la notion d'écart aux observations utilisées (voir par exemple le tableau 3.1)

Gradients Normalisés

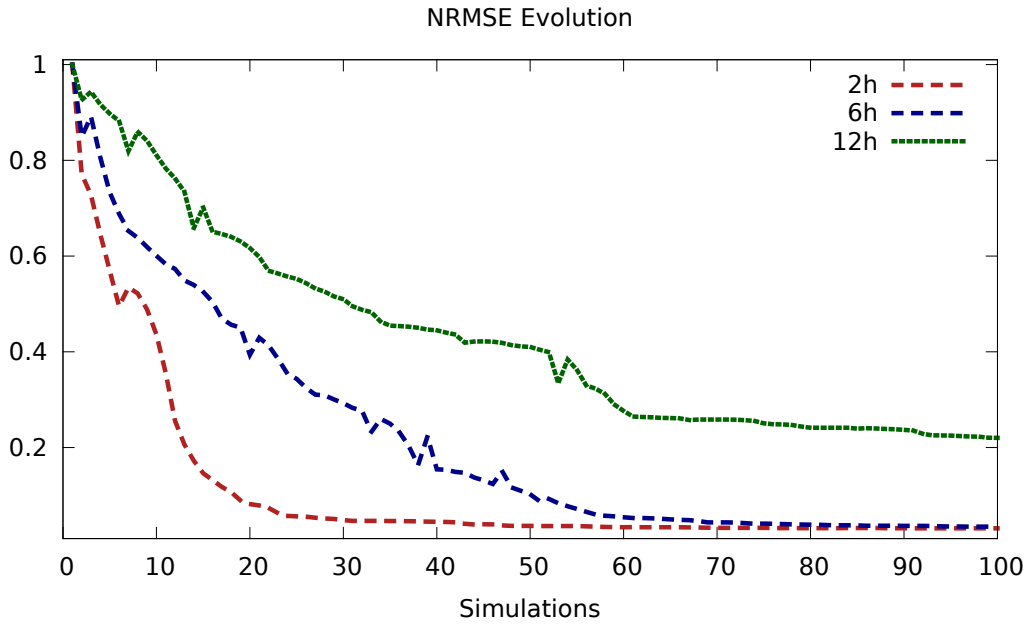


FIGURE 3.11 – Evolution du pourcentage d'erreur résiduelle au fil des itérations pour des fenêtres de $2h$, $6h$, $12h$ et $24h$. La variable d'intérêt est la composante méridionale de la vitesse. L'écart entre le modèle et les observations est calculé entre gradients normalisés. Le traceur est observé toutes les 30 minutes.

Cependant, lorsque les expériences s'effectuent sur des fenêtres d'assimilations bien plus courtes ($2h$ et $6h$) les résultats à convergence sont comparables (voire même un peu meilleurs) que ceux obtenus avec la distance entre pixels.

Les mauvais résultats lorsque la comparaison est effectuée entre gradients normalisés (sur des fenêtres de $12h$) ne proviennent donc pas d'un manque d'information intrinsèque à cette représentation des champs de traceur passifs. Si tel était le cas, effectuer cette comparaison avec moins d'information ne devrait pas conduire à des résultats similaires à ceux obtenus dans l'espace des pixels.

Une explication possible aux mauvais résultats observés lors d'expériences d'assimilation assez longues réside dans la non linéarité de la transformation appliquée au champ de traceurs. Cette non linéarité peut amplifier les non linéarités de la fonction coût, rendant dès lors encore plus complexe la minimisation lorsque la fenêtre d'assimilation est longue.

Bilan 3.2

En résumé nous avons observé que :

- *Les non-linéarités du modèle entraînent une convergence plus lente de la minimisation lorsque la fenêtre d'assimilation grandit. Ce phénomène est amplifié lors de l'utilisation du terme d'écart entre gradients normalisés.*
- *Lorsque les expériences s'effectuent sur des fenêtres de $2h$ ou de $6h$ les résultats obtenus avec les gradients normalisés peuvent concurrencer ceux obtenus avec les autres termes d'écarts aux observations,*
- *la diminution de la fréquence d'observation n'entraîne pas (avec des observations parfaites) de déterioration de l'analyse (sauf pour les gradients normalisés sur la fenêtre de $12h$).*

3.3.5 Impact de la régularisation par l'ébauche

En section 3.3.3 nous avons présenté les résultats obtenus pour des expériences d'assimilation sur des fenêtres de $12h$ en précisant que les champs de vitesse analysés étaient un peu trop lisses par rapport au véritable champ de vitesse (voir figure 3.9).
Ce phénomène provient du terme d'écart à l'ébauche choisi. Dans cette section nous allons voir l'influence de l'incorporation d'une ébauche sur la minimisation.

Quelle est l'importance de l'ébauche ? Comme précisé en début de section l'ébauche étant de piètre qualité, le terme d'écart à l'ébauche dans la fonction coût a été pondéré par un poids $\alpha_b = 10^{-5}$. Sa contribution à la valeur de la fonction coût, même lorsque les observations sont parfaites reste donc très faible $\left(\text{le ratio } \dfrac{J^o}{J^b} > 10^7 \text{ en fin de minimisation}\right)$.

Remarque 3.22

On peut dès lors se demander pourquoi lors d'expériences relativement bien observées ce terme de régularisation est ajouté (et surtout pourquoi il est important). Son importance ne vient ni de l'information apportée par l'ébauche (on sait que cette information n'est pas bonne) ni par le fait que les champs analysés doivent être de faible amplitude (la contribution à la fonction coût du terme d'écart à l'ébauche est relativement faible). Son intérêt réside principalement dans la matrice de covariance d'erreur d'ébauche utilisée et plus précisément du changement de variable effectué (voir section 1.3). Dans les faits, lors des expériences effectuées précédemment, son introduction est cruciale.

Les expériences présentées jusqu'ici ont été réalisées à l'aide de la matrice de covariance introduite dans [Weaver and Courtier, 2001] modélisée par un opérateur de diffusion (voir section 1.5). La longueur de corrélation de l'erreur d'ébauche utilisée jusqu'à présent est de 5.4 pixels. Dans la suite, l'influence du choix de cette longueur de corrélation d'erreur d'ébauche est présenté.

Impact du choix de la matrice B

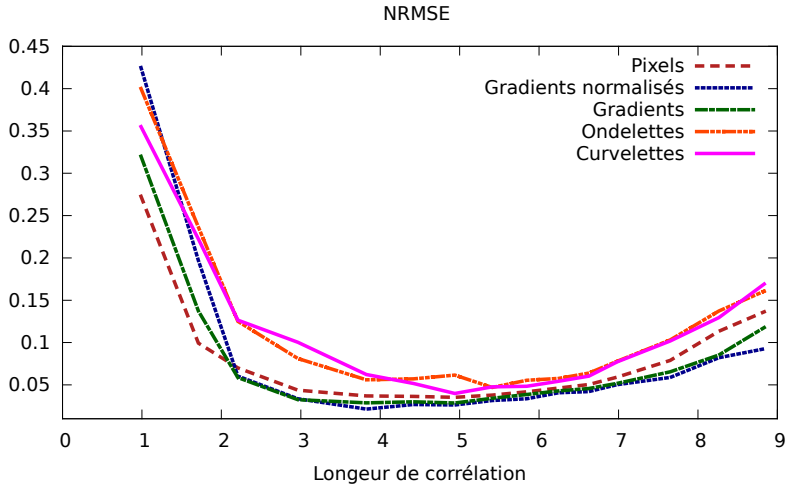


FIGURE 3.12 – Le ratio $\dfrac{RMSE(v^a)}{RMSE(v^b)}$ pour les différents termes d'écart aux observations est présenté en ordonné en fonction de la longueur de corrélation d'erreur d'ébauche utilisée.

Afin d'évaluer plus précisément l'impact de ce choix, replaçons-nous dans le cas jusqu'à présent le plus favorable à tous les opérateurs, à savoir le cas où la fenêtre d'assimilation est relativement courte ($2h$) en utilisant une observation toute les demi-heure (par conséquent nous

utilisons 4 images sur la période de temps concernée). Les expériences réalisées consistent, pour tous les opérateurs d'observations, à changer la longueur de corrélation d'erreur d'ébauche utilisée lors de la création de la matrice **B**. Ainsi, en comparant les résultats obtenus nous pouvons appréhender l'impact de ces choix sur le processus d'assimilation.

Description des expériences

But : Evaluer l'impact du changement de modélisation de la matrice de covariance d'erreur d'ébauche.

Observations utilisées : Une séquence de 4 images parfaites. Le traceur est observé toute les 30 minutes.

Longueur de corrélation d'erreur d'ébauche : Elle varie entre 1 et 9 (pixels).

Minimisation : 50 itérations de minimisations sont utilisées.

Allure générale

La figure 3.12 présente, pour tous les termes d'écarts aux observations, le pourcentage d'erreur résiduelle de la composante méridionale du champ de vitesse en fin de minimisation. Ce pourcentage d'erreur (en ordonné) est présenté en fonction de la longeur de corrélation de l'erreur d'ébauche utilisée lors de la génération de la matrice **B**.

L'allure des courbes dépend peu de la définition du terme d'écart aux observations : il y a d'abord une phase de décroissance rapide, suivit d'une phase de croissance lente.

- Lorsque la longueur de corrélation est faible, la minimisation est complexe. Les solutions ne sont pas assez contraintes pour obtenir des résultats de qualité convaincante (voir figure 3.14)
- En augmentant cette longueur de corrélation, on rend la fonctionnelle à minimiser plus lisse ce qui permet d'obtenir de bien meilleurs résultats.
- Cependant lorsque la longueur de corrélation d'erreur d'ébauche devient trop importante, on réduit trop fortement l'espace de recherche ce qui a pour conséquence de déteriorer les résultats obtenus.

Différence entre opérateurs

Toujours en figure 3.12 on observe que l'utilisation de la pseudométrique dans un sous-espace d'ondelettes (ou de curvelettes) dégrade (un peu) l'analyse. Ce phénomène est observé quel que soit la longueur de corrélation de l'erreur d'ébauche spécifiée. Cependant, contrairement aux analyses effectuées sur des fenêtres de tailles plus longues, la comparaison entre gradients normalisés permet d'obtenir des résultats similaires à la distance entre pixels (pour la variable v).

La figure 3.13 présente le pourcentage d'erreur angulaire restant en fin d'assimilation. On observe ici une différence entre termes d'écart aux observations plus importante. Les meilleurs résultats sont obtenus lors de la comparaison entre gradients normalisés lorsque la longueur de corrélation est comprise entre 3 et 4 pixels. La moins bonne performance des ondelettes et des curvelettes est amplifiée pour cet indicateur (environ 7% d'erreur supplémentaire par rapport aux autre termes d'écart aux observations et ce quel que soit la longueur de corrélation). La suppression d'une partie de l'information (consistant essentiellement au fait que le champ de traceurs est lisse dans certaines zones) entraîne donc une perte sur la qualité de la direction du champ de vitesse analysé.

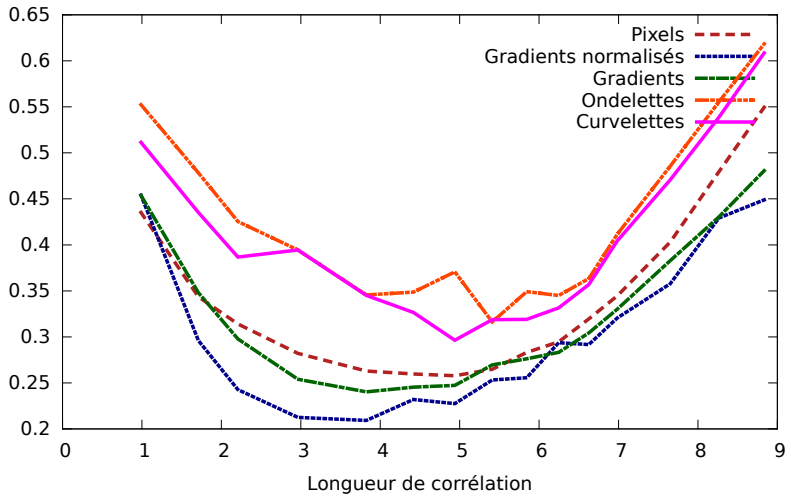


FIGURE 3.13 – Le ratio $\frac{RMSE(\alpha^a)}{RMSE(\alpha^b)}$ (où α est l'erreur angulaire) pour différents termes d'écart aux observations est présenté en ordonnée en fonction de la longueur de corrélation de l'erreur d'ébauche.

Quel phénomène est à l'oeuvre ?

Après avoir décrit de manière quantitative les résultats (par la comparaison de l'erreur initiale et l'erreur sur les champs analysés) intéressons-nous à une description qualitative afin de comprendre le phénomène sous-jacent.

La figure 3.14 présente la composante zonale de la vitesse analysée lorsque la longueur de corrélation d'erreur d'ébauche utilisée pour créer la matrice **B** est de 1, 3.8 et 8.8 pixels. Les résultats sont uniquement présentés lorsque l'écart aux observations est calculé dans l'espace des pixels [11].

Lorsque la longueur de corrélation est de 1 pixel (figure en haut à droite), nous observons que le champ u analysé n'est pas lisse (bien qu'il ressemble vaguement au véritable champ u). Au contraire, quand cette longueur est de 8.8 pixels, le champ analysé ne peut pas représenter les petites discontinuités au niveau du vortex.

Ce phénomène est lié au gradient de la fonction coût (présenté en figure 3.15) qui, lorsque la longueur de corrélation d'erreur d'ébauche employée est petite, est fortement localisé (ou au contraire extrêmement lisse lorsque cette dernière est importante). Lorsque la longueur de corrélation est petite, la localisation observée est liée au fait que les observations utilisées sont des observations indirectes. Une majeure partie du coût (dû aux observations) est portée par relativement peu de pixels (car le mauvais placement à quelques pixels près du tourbillon entraîne la localisation de l'erreur au niveau de la frontière de ce dernier). Dès lors, bien que l'intégration rétrograde diffuse quelque peu l'information, celle-ci reste fortement localisée en ces endroits. La diffusion de l'information est dans ce cas assurée par l'utilisation d'une matrice **B** non diagonale en tant que préconditionneur.

Cependant cette diffusion a un coût (et pas que numérique) : on oblige l'incrément à vivre dans un espace lisse (ou tout au moins, on le contraint très fortement à être lisse). Dès lors, il devient compliqué de recréer certaines discontinuités au sein du champ de vitesse.

11. les résultats obtenus sont "comparables" avec les autres définitions de l'écart aux observations

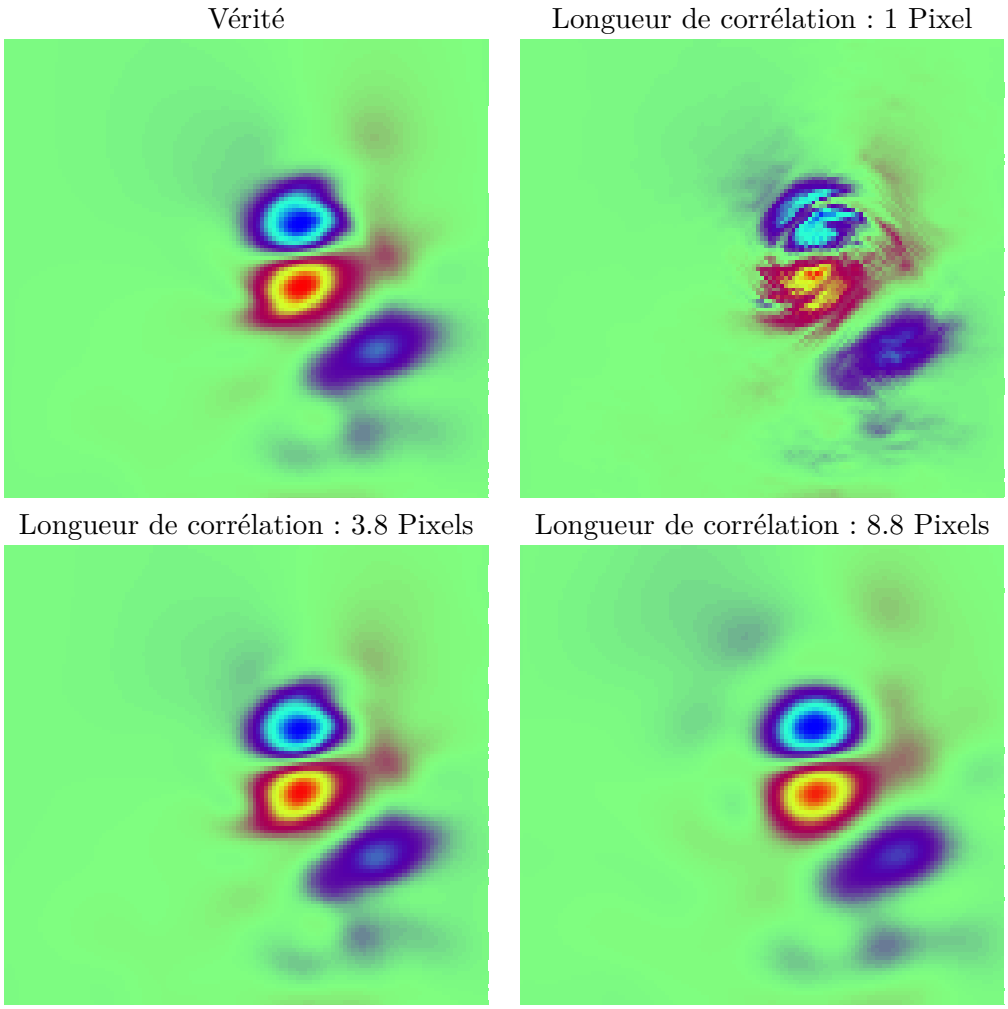


FIGURE 3.14 – En haut à gauche, la véritable vitesse zonale. Les autres figures présentent cette composante analysée en fonction de la longeur de corrélation de l'erreur d'ébauche employée.

3.4 Conclusion

Dans ce chapitre, nous avons introduit les différents termes d'écarts aux observations utilisés dans l'ensemble du manuscrit ainsi que le contexte expérimental servant de support à l'ensemble des expériences réalisées dans la suite.

En section 3.3 nous avons vu qu'il était possible avec des données parfaites (et sur des expériences assez courtes pour les gradients normalisés) d'extraire suffisamment d'information de séquences d'images de traceurs passifs afin de contrôler le champ de vitesse initial du système. Nous avons aussi observé certaines limites du processus d'assimilation :

— En absence d'opérateur d'équilibre, la condition initiale sur la surface libre n'est que très peu corrigée. Même pour des expériences relativement longues ($24h$ équivalent atmosphère) l'intégration du modèle rétrograde ne permet pas d'effectuer une correction importante sur cette variable. Ce problème est présent pour l'ensemble des expériences effectuées dans ce manuscrit mais n'interfère pas avec les conclusions que l'on peut tirer de ces dernières. En effet, cette mauvaise correction de la surface libre, n'affecte pas notre capacité à extraire de l'information exploitable sur le champ de vitesse. Dans ce manuscrit, c'est la détermination de ce champ de vitesse qui nous intéresse avant tout. En pratique, dans des modèles opérationnels, on peut espérer avoir accès à des opérateurs d'équilibre plus sophistiqués que l'équilibre géostrophique afin de pouvoir retransmettre l'amélioration sur le champ de vitesse vers les autres variables durant la phase d'assimi-

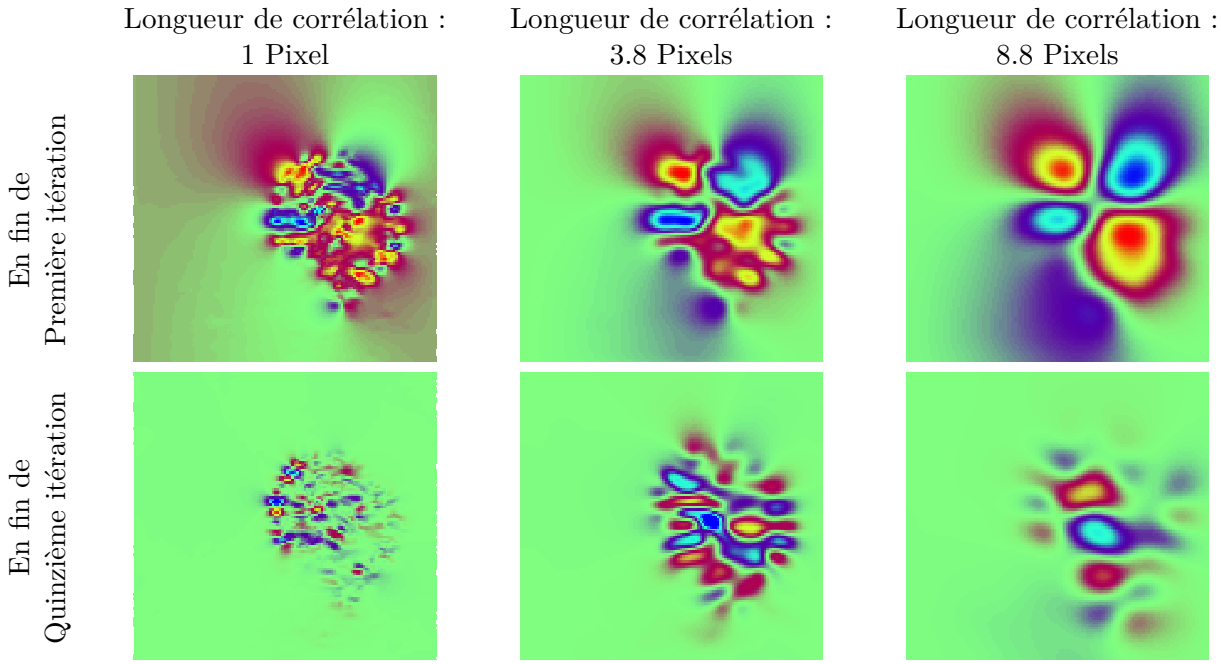


FIGURE 3.15 – Gradient de la fonction coût par rapport à la composante zonale de la vitesse. A noter que l'échelle varie selon l'image regardée (mais ici seule la distribution spatiale du gradient nous intéresse).

lation ;

— lorsque la taille de la fenêtre d'assimilation augmente, la vitesse de convergence diminue fortement (à cause des non linéarités du modèle). Pour cette raison (et aussi pour des raisons de temps de calculs), dans la suite les expériences seront réalisées sur des fenêtres relativement courtes (entre $2h24$ et $3h$ équivalent atmosphère) ;

— Enfin, nous avons observé le rôle crucial joué par la matrice $\mathbf{B}$ via la contrainte imposée sur la régularité de l'incrément par l'étape de préconditionnement. Nous reviendrons de temps à autre sur ce rôle dans les chapitres qui vont suivre.

4

Impact d'un bruit d'observation non corrélé en espace

Sommaire

4.1 Introduction **73**
4.1.1 Rappel 73
4.1.2 Notations 74
4.1.3 Matrice de covariance ou matrice de corrélation ? 74
4.2 Matrice de covariance dans différents espaces **75**
4.2.1 Distance entre gradients 76
4.2.2 Dans une base d'ondelettes orthonormées 79
4.2.3 Distance dans une trame de curvelettes 79
4.2.4 Distance entre les gradients normalisés 82
4.3 Expériences effectuées **85**
4.3.1 Impact du niveau de bruit 86
4.3.2 Intérêt de la perte d'information en présence de biais 88
4.3.3 Impact de la fréquence d'observation 90

4.1 Introduction

Dans le chapitre précédent, des observations parfaites ont été utilisées afin d'évaluer le degré de faisabilité de la méthode d'assimilation employée. Malheureusement, en réalité, les observations ne sont pas parfaites mais entachées d'erreurs dont la présence doit être prise en compte lors de l'utilisation des données. Dans cette partie, nous étudions l'impact d'un bruit décorrélé (en espace et en temps) dans l'espace des pixels. Les différentes matrices de covariances utilisées sont présentées dans une première partie pour chacune des notions d'écarts aux observations introduites en section 3.2.2 du chapitre 3. Par la suite, en section 4.3, le résultat d'expériences jumelles réalisées en présence de données corrompues par un bruit additif gaussien est présenté afin de percevoir l'effet d'un tel bruit sur les différents opérateurs d'observation.

4.1.1 Rappel

L'effet d'une transformation linéaire, notée de manière générique $\mathbf{A}$, sur des observations entachées d'erreurs est rappelé dans un premier temps.
Soit des données entachées d'erreurs définies par $\mathbf{y^o}$:

$$\mathbf{y^o} = \mathbf{y^t} + \boldsymbol{\epsilon} \tag{4.1}$$

où $\mathbf{y^t}$ représente le vrai vecteur de données et où le bruit $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{R}_{\mathcal{O}_1})$. Soit une application linéaire $\mathbf{A} : \mathcal{O}_1 \mapsto \mathcal{O}_2$ appliquée aux données, alors

$$\begin{aligned} \mathbf{Ay^o} &= \mathbf{Ay^t} + \mathbf{A}\boldsymbol{\epsilon} \\ &= \mathbf{Ay^t} + \boldsymbol{\epsilon}_2 \end{aligned}$$

avec $\boldsymbol{\epsilon}_2 \sim \mathcal{N}(0, \mathbf{R}_{\mathcal{O}_2})$ où

$$\mathbf{R}_{\mathcal{O}_2} = \mathbf{A}\mathbf{R}_{\mathcal{O}_1}\mathbf{A}^T. \tag{4.2}$$

La formule (4.2) va régulièrement nous servir dans la suite à construire des matrices de covariances équivalentes dans l'espace image de $\mathbf{A}$ à la matrice de covariance de départ (lorsque l'application est linéaire).

4.1.2 Notations

Les matrices de covariance d'erreur d'observation intervenant dans la fonction coût sont étudiées de manière approfondie :

$$J(X) = \|x^b - X\|^2_{\mathbf{B}} + \sum_{t_i=t_0}^{t_f} \|\mathbf{A}I^o(t_i) - \mathbf{A}I^{synth}(t_i)\|^2_{\mathbf{R_A}} \tag{4.3}$$

La véritable matrice de covariance d'erreur d'observation est notée (quelque soit le bruit considéré) :

$$\mathbf{R_A} \tag{4.4}$$

où $\mathbf{A}$ dénote l'espace dans lequel les données sont analysées. Le tableau 4.1 présente les différents noms associés aux véritables matrices de covariances selon l'espace de travail, où le sous-espace défini par le seuillage présenté en équation (3.29).

Table 4.1 – Notations concernant les véritables matrices de covariances

Espace	Matrice pleine	Matrice dans un sous espace associée au seuillage $\mathcal{T}$
Pixels	$\mathbf{R}_{pix}$	
Gradients	$\mathbf{R}_{\nabla}$	
Ondelettes	$\mathbf{R_W}$	$\mathbf{R}_{\mathcal{T}\mathbf{W}}$
Curvelettes	$\mathbf{R_C}$	$\mathbf{R}_{\mathcal{T}\mathbf{C}}$
Gradients Normalisés	$\mathbf{R_{GN}}$	

Cependant, nous allons bien souvent travailler non pas avec ces matrices de covariances mais avec des approximations de ces matrices. Ces approximations sont notées de manière générique :

$$\tilde{\mathbf{R}}^a_{b,c} \tag{4.5}$$

où a représente la distribution du bruit considéré dans l'espace des pixels (par exemple indépendant et identiquement distribué), b l'espace dans lequel les données sont analysées et c le type d'approximation effectuée dans cet espace. Le tableau 4.2 présente les différentes notations utilisées.

4.1.3 Matrice de covariance ou matrice de corrélation ?

Dans ce chapitre et dans les suivants, à partir d'une certaine matrice de covariance dans l'espace des pixels, des approximations de cette matrice dans d'autres espaces sont introduites. Cependant, seul le cas où la diagonale de $\mathbf{R}_{pix}$ est proportionnelle à la matrice identité est abordé.

TABLE 4.2 – Notations concernant les matrices d'approximations

Type de bruit	Notation associée
Bruit indépendant et identiquement distribué	iid
Bruit corrélé en espace	cor

Type d'approximation	Notation associée	
Approximation scalaire	scal	$\tilde{\mathbf{R}}_{,\text{scal}} = \alpha \mathbf{I_n}$
Approximation diagonale	diag	$\tilde{\mathbf{R}}_{,\text{diag}} = \text{Diag}(\mathbf{R})$
Approximation par blocs	bloc	

Bien que paraissant à première vue très restrictif, ce cas englobe l'ensemble des matrices de covariances.

Pour cela, remarquons que toute matrice de covariance dans l'espace des pixels peut s'écrire sous la forme :

$$\mathbf{R}_{pix} = \mathbf{\Sigma}_{pix}^{1/2} C_{pix} \mathbf{\Sigma}_{pix}^{1/2} \tag{4.6}$$

avec C_{pix} la matrice de corrélation dans l'espace des pixels et $\mathbf{\Sigma}_{pix}^{1/2}$ la matrice diagonale des écarts types.

Lorsque l'application linéaire $\mathbf{A}$ est appliquée aux données, le terme d'écart aux observations s'écrit dans cet espace par :

$$\begin{aligned} J_{\mathbf{A}}^{o} &= \left(\mathbf{A}(I^{synth} - I^{o})\right)^{T} (\mathbf{R_A})^{-1} \left(\mathbf{A}(I^{synth} - I^{o})\right) \\ &= \left(I^{synth} - I^{o}\right)^{T} \mathbf{A}^{T} (\mathbf{R_A})^{-1} \mathbf{A} \left(I^{synth} - I^{o}\right) \end{aligned}$$

Dès lors que l'approximation effectué dans cet espace $\tilde{\mathbf{R}}_{\mathbf{A},}$ permet de définir une bonne approximation de l'inverse de la matrice de corrélation C_{pix} :

$$C_{pix}^{-1} \simeq \mathbf{A}^{T} (\tilde{\mathbf{R}}_{\mathbf{A},})^{-1} \mathbf{A} \tag{4.7}$$

alors ceci se généralise à la matrice de covariance R_{pix} :

$$\mathbf{\Sigma}_{pix}^{-T/2} C_{pix}^{-1} \mathbf{\Sigma}_{pix}^{-1/2} \simeq \mathbf{\Sigma}_{pix}^{-T/2} \mathbf{A}^{T} (\tilde{\mathbf{R}}_{\mathbf{A},})^{-1} \mathbf{A} \mathbf{\Sigma}_{pix}^{-1/2}. \tag{4.8}$$

Par conséquent, si on peut accéder à une bonne approximation de l'inverse de la matrice de corrélation dans un autre espace, cette dernière peut être employée même lorsqu'on cherche à approcher une matrice de covariance d'erreur d'observation. Dans ce cas, le terme d'écart aux observations s'écrit :

$$J_{\mathbf{A}}^{o} = \left(\mathbf{A} \mathbf{\Sigma}_{pix}^{-1/2} (I^{synth} - I^{o})\right)^{T} (\mathbf{R_A})^{-1} \left(\mathbf{A} \mathbf{\Sigma}_{pix}^{-1/2} (I^{synth} - I^{o})\right)$$

Bilan 4.1

Pour cette raison seul le cas où la matrice de covariance possède une diagonale proportionelle à l'identité est étudiée dans par la suite. Cependant, nous appelerons toujours cette matrice une matrice de covariance.

4.2 Matrice de covariance dans différents espaces

Dans cette section, sont décrites brièvement les matrices de covariances utilisées dans les expériences de la section 4.3. Le bruit est supposé indépendant, identiquement distribué et de variance connue (σ^2) dans l'espace des pixels. Sa matrice de covariance s'écrit :

$$\mathbf{R}_{pix} = \sigma^2 \mathbf{I_n}. \tag{4.9}$$

4.2.1 Distance entre gradients

Lorsque la fonction calculant l'écart aux observations est bâtie dans l'espace des gradients, d'après (4.2), la matrice de covariance est définie par :

$$\begin{aligned}\mathbf{R}_{\nabla} &= \nabla \mathbf{R}_{pix} \nabla^T \\ &= \sigma^2 \nabla \nabla^T.\end{aligned}$$

Le gradient d'une image bruitée s'écrit sous la forme :

$$\nabla I^o = \nabla I^t + \nabla \boldsymbol{\epsilon} \tag{4.10}$$

Le choix fait dans ce manuscrit est d'utiliser un schéma de différences finies centré d'ordre 2 afin de calculer l'opérateur gradient $\nabla \epsilon = [\epsilon_x, \epsilon_y]^T$ en un pixel (i, j) :

$$\epsilon_x(i,j) = \frac{\epsilon(i,j+1) - \epsilon(i,j-1)}{2} \text{ si } 1 < j < H \text{ et } 1 \leq i \leq L \tag{4.11}$$

et

$$\epsilon_y(i,j) = \frac{\epsilon(i+1,j) - \epsilon(i-1,j)}{2} \text{ si } 1 < j \leq H \text{ et } 1 \leq i < L \tag{4.12}$$

où L et H représentent le nombre de lignes et de colonnes.

Remarque 4.1

Il faut noter qu'avec cette définition le gradient n'est pas calculé sur les bords de l'image. Cependant l'information contenue sur les pixels du bord est utilisée au travers du calcul des gradients des points proches.

La matrice de covariance $\mathbf{R}_{\nabla}$ comporte différentes composantes extra–diagonale non nulles et ce même si le bruit est gaussien indépendant et identiquement distribué (ie $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I_n})$). Ces composantes sont liées au choix du schéma pour calculer le gradient. Les différents termes composants une ligne de cette matrice sont explicités dans les paragraphes suivants.

La variance de chacun des éléments est :

$$\begin{aligned}\text{var}\left(\epsilon_x(i,j)\right) &= \text{var}\left(\frac{\epsilon(i,j+1) - \epsilon(i,j-1)}{2}\right) \\ &= \frac{\text{var}\left(\epsilon(i,j+1)\right) + \text{var}\left(\epsilon(i,j-1)\right) - 2\,\text{cov}\left(\epsilon(i,j+1), \epsilon(i,j-1)\right)}{4} \\ &= \frac{\sigma^2}{2} \quad \text{car } \boldsymbol{\epsilon} \text{ est un bruit iid.}\end{aligned}$$

Deux éléments peuvent être corrélés uniquement si leurs calculs (via les schémas (4.11) et (4.12)) mettent en jeu l'information contenue en un même pixel. C'est par exemple le cas de $\epsilon_x(i+2,j)$ et $\epsilon_x(i,j)$ qui comportent chacun une part d'erreur due à la mesure du pixel $I(i+1,j)$. La covariance entre les erreurs dans le calcul des dérivées $I_x(i+2,j)$ et $I_x(i,j)$ se calcule de la manière suivante :

$$\begin{aligned}\text{cov}\left(\epsilon_x(i+2,j), \epsilon_x(i,j)\right) &= \text{cov}\left(\frac{\epsilon(i+3,j) - \epsilon(i+1,j)}{2}, \frac{\epsilon(i+1,j) - \epsilon(i-1,j)}{2}\right) \\ &= -\frac{\text{var}\left(\epsilon(i+1,j)\right)}{4} \quad \text{car } \boldsymbol{\epsilon} \text{ est un bruit iid} \\ &= -\frac{\sigma^2}{4}.\end{aligned}$$

D'autres éléments sont corrélés avec $\epsilon_x(i,j)$: ce sont les éléménts $\epsilon_x(i+2,j)$, $\epsilon_x(i-2,j)$, $\epsilon_y(i-1,j-1)$, $\epsilon_y(i+1,j-1)$, $\epsilon_y(i-1,j+1)$ et $\epsilon_y(i,j-1)$. Un résumé des corrélations entre $\epsilon_x(i,j)$ et les autres composantes du bruit est présenté en figure 4.1.

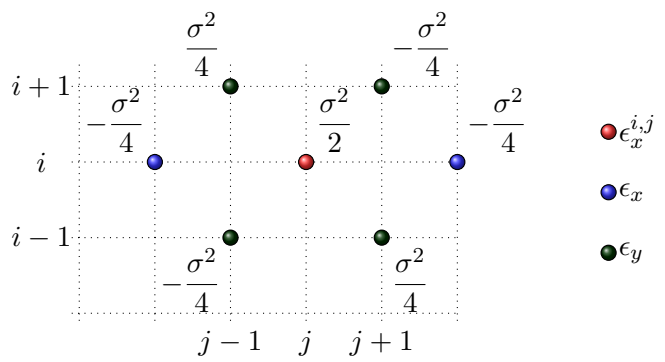


FIGURE 4.1 – La covariance entre $\epsilon_x(i,j)$ (en rouge) et les autres dérivées partielles (en vert et en bleu) pour un bruit gaussien indépendant et identiquement distribué dans l'espace des pixels. Les points en bleu (resp. en vert) représentent les dérivées dans la direction x (resp. y) corrélées à $\epsilon_x(i,j)$. L'ensemble des covariances entre $\epsilon_x(i,j)$ et les autres éléments est représenté sur ce schéma.

La matrice de covariance provenant de cette construction s'écrit de la manière suivante :

$$\mathbf{R}_\nabla = \begin{pmatrix} C_{\boldsymbol{\epsilon}_x\boldsymbol{\epsilon}_x} & C_{\boldsymbol{\epsilon}_x\boldsymbol{\epsilon}_y} \\ C_{\boldsymbol{\epsilon}_y\boldsymbol{\epsilon}_x} & C_{\boldsymbol{\epsilon}_y\boldsymbol{\epsilon}_y}. \end{pmatrix} \tag{4.13}$$

où $C_{\boldsymbol{\epsilon}_x\boldsymbol{\epsilon}_x}$ est un bloc comprenant les covariances entre éléments du vecteur $\boldsymbol{\epsilon}_x$, $C_{\boldsymbol{\epsilon}_y\boldsymbol{\epsilon}_y}$ entre éléments du vecteur $\boldsymbol{\epsilon}_y$, $C_{\boldsymbol{\epsilon}_x\boldsymbol{\epsilon}_y}$ et $C_{\boldsymbol{\epsilon}_y\boldsymbol{\epsilon}_x}$ entre les éléments du vecteur $\boldsymbol{\epsilon}_x$ et les éléments du vecteur $\boldsymbol{\epsilon}_y$.

Remarque 4.2

Cette matrice de covariance n'est pas inversible : elle est au mieux de rang $L \times H$ quand son nombre de lignes et de colonnes est de $(L-1)\times H + (H-1)\times L$ (avec L et H le nombre de lignes et de colonnes de l'image). Dans la suite de cette section, deux approximations différentes sont envisagées : une approximation diagonale et une approximation par blocs diagonaux.

Approximation diagonale

Une première approximation consiste à considérer que, dans l'espace des gradients, le bruit est décorrélé et donc que toutes les covariances calculées précédemment sont le fruit de notre imagination. La matrice de covariance obtenue après cette approximation s'écrit sous la forme :

$$\begin{aligned} \tilde{\mathbf{R}}^{iid}_{\nabla,\text{scal}} &= \text{diag}(\mathbf{R}_\nabla) \\ &= \frac{\sigma^2}{2}\mathbf{I} \end{aligned} \tag{4.14}$$

Approximation par blocs

Une manière de simplifier la matrice de covariance $\mathbf{R}_\nabla$ consiste à considérer qu'il n'existe pas de corrélations croisées entre les vecteurs $\boldsymbol{\epsilon}_x$ et $\boldsymbol{\epsilon}_y$ (ce qui revient à dire que $C_{\epsilon_x\epsilon_y} = 0$). Cette simplification est illustrée en figure 4.2. On observe, qu'en plus d'engendrer une matrice inversible, cette approximation a le bon goût de considérer que $\epsilon_x(i,j)$ est simplement corrélé à des éléments de la même ligne que lui (de même $\epsilon_y(i,j)$ est corrélé uniquement avec des éléments de la même colonne que lui).

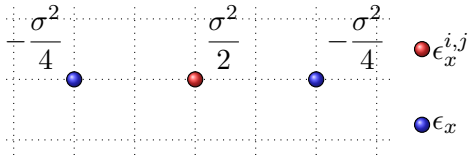


FIGURE 4.2 – Illustration des corrélations prises en compte lors de la simplification par blocs diagonaux de la matrice de covariance d'erreur d'observation. Seules celles entre les dérivées partielles dans la même direction sont considérées.

La matrice par blocs résultant de cette approximation peut s'écrire :

$$\tilde{\mathbf{R}}^{iid}_{\nabla,\text{bloc}} = \begin{pmatrix} C_{\epsilon_x^{i=1}\epsilon_x^{i=1}} & 0 & \cdots & 0 \\ 0 & C_{\epsilon_x^{i=2}\epsilon_x^{i=2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & C_{\epsilon_y^{j=H}\epsilon_y^{j=H}} \end{pmatrix} \tag{4.15}$$

où $C_{\epsilon_x^{i=1}\epsilon_x^{i=1}}$ est le bloc correspondant à la covariance entre les éléments du vecteur $\boldsymbol{\epsilon}_x$ correspondants à la première ligne de l'image.

Inversion de $\tilde{\mathbf{R}}^{iid}_{\nabla,\text{bloc}}$ Comme on peut s'en apercevoir sur la figure 4.2, sur chaque ligne les éléments des colonnes paires sont uniquement corrélés entre eux (de même pour les éléments des colonnes impaires). Chaque bloc $C_{\epsilon_x^{i=k}\epsilon_x^{i=k}}$ de la matrice $\tilde{\mathbf{R}}^{iid}_{\nabla,\text{bloc}}$ peut être séparé en deux blocs (dépendant de la parité de j) :

$$C_{\epsilon_x^{i=k}\epsilon_x^{i=k}} = \begin{pmatrix} C_{\epsilon_x^{j=odd}\epsilon_x^{j=odd}} & 0 \\ 0 & C_{\epsilon_x^{j=even}\epsilon_x^{j=even}} \end{pmatrix}$$

Pour chacune des lignes (resp. colonnes) ces deux blocs sont identiques. On doit donc, au maximum, inverser quatre matrices tridiagonales $C_{\epsilon_x^{j=odd}\epsilon_x^{j=odd}}$, $C_{\epsilon_x^{j=even}\epsilon_x^{j=even}}$, $C_{\epsilon_y^{i=odd}\epsilon_y^{i=odd}}$, $C_{\epsilon_y^{i=even}\epsilon_y^{i=even}}$ afin de calculer $\tilde{\mathbf{R}}^{iid}_{\nabla,\text{bloc}}$.

Dans le cas où $L = H = 2N$ on a égalité entre ces 4 matrices. Par conséquent, l'inversion de $C_{\epsilon_x^{odd}\epsilon_x^{odd}}$ $\left(\text{matrice de taille } \left(\frac{L}{2}-1\right) \times \left(\frac{L}{2}-1\right)\right)$ est suffisante pour construire la matrice de covariance $\tilde{\mathbf{R}}^{iid}_{\nabla,\text{bloc}}$. Cette matrice est constituée de la sorte :

$$C_{\epsilon_x^{i=odd}\epsilon_x^{i=odd}} = \sigma^2 \begin{pmatrix} 0.5 & -0.25 & 0 & \cdots & 0 \\ -0.25 & 0.5 & -0.25 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & 0 & -0.25 & 0.5 & -0.25 \\ \vdots & \vdots & 0 & -0.25 & 0.5 \end{pmatrix} \tag{4.16}$$

Son inverse existe et s'écrit de la manière suivante :

$$C^{-1}_{\epsilon_x^{i=odd}\epsilon_x^{i=odd}} = \frac{4}{N\sigma^2} F$$

avec

$$F = \begin{pmatrix} N-1 & N-2 & N-3 & \cdots & 1 \\ N-2 & 2*(N-2) & 2*(N-3) & \cdots & 2 \\ N-3 & 2*(N-3) & 3*(N-3) & \cdots & 3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2 & 3 & \cdots & N-1 \end{pmatrix}$$

Bilan 4.2

Ainsi, il est possible, en ne tenant pas compte des corrélations existantes entre les dérivées directionnelles x et y, de construire une matrice par blocs aisément inversible dans le cas où le bruit dans l'espace des pixels est indépendant et identiquement distribué en espace. Dans la section 4.3, nous comparerons les résultats obtenus avec l'approximation scalaire (4.14) et l'approximation par blocs diagonaux (4.15) dans l'espace des gradients afin de voir ce que nous apporte cette différence de modélisation.

4.2.2 Dans une base d'ondelettes orthonormées

Les décompositions en ondelettes étant linéaires, la matrice de covariance d'un bruit indépendant et identiquement distribué dans l'espace des pixels, s'écrit en utilisant l'équation (4.2) :

$$\mathbf{R_W} = \sigma^2 \mathbf{W}\mathbf{W}^T. \tag{4.17}$$

où $\mathbf{W}$ est la décompostion en ondelettes considérée. Comme seules des transformées en ondelettes orthonormales sont utilisées, $\mathbf{W}^T = \mathbf{W}^{-1}$. Par conséquent, la matrice de covariance d'erreur d'observation, dans l'espace des ondelettes, est proportionnelle à l'identité :

$$\mathbf{R_W} = \sigma^2 \mathbf{I}. \tag{4.18}$$

Comme précisé en remarque 3.13, dans ce cas là, les notions d'écart aux observations dans l'espace des ondelettes (définie en équation (3.26)) et dans l'espace des pixels (équation (3.8)) sont identiques.

Remarque 4.3

Ceci n'est plus vrai dans le cas d'utilisation de bases d'ondelettes biorthogonales (largement répandues dans la littérature) car dans ce cas là $W^T \neq W^{-1}$. Des corrélations entre les éléments de base apparaissent lors de l'utilisation d'une telle transformée.

Par conséquent, lorsque le bruit est indépendant et identiquement distribué, seule la pseudométrique entre deux images dans un sous-espace d'ondelettes définie via l'opération de seuillage présentée en équation (3.31) est considérée. La matrice de covariance correspondante est :

$$\mathbf{R}_{\mathcal{T}\mathbf{W}} = \sigma^2 \mathbf{P}\mathbf{I}\mathbf{P}^T \tag{4.19}$$

où $\mathbf{P}$ correspond à la projection sur le sous-espace des coefficients non modifiés par le seuillage (3.29).

La comparaison entre les résultats obtenus avec l'opération de seuillage dans une base d'ondelettes et ceux dans l'espace des pixels permet dès lors de quantifier l'impact de la sélection d'information sur la minimisation.

Remarque 4.4

Dans ce cadre là, l'ensemble de l'information sur le signal n'est pas utilisé. Cependant, comme pour l'écart au niveau pixels, les véritables statistiques portant sur les erreurs d'observations sont utilisées (pour l'information conservée après seuillage).

4.2.3 Distance dans une trame de curvelettes

De manière analogue à ce qui est fait pour les ondelettes (et qui correspond à l'équation (4.2)), on peut définir la matrice de covariance dans l'espace des curvelettes par :

$$\mathbf{R_C} = \mathbf{C}\mathbf{R}_{Pix}\mathbf{C}^T = \sigma^2 \mathbf{C}\mathbf{C}^T \tag{4.20}$$

Cependant les matrices $\mathbf{C}$, $\mathbf{R}_{pix}$ et $\mathbf{C}^T$ étant de rang n (où n est le nombre de pixels de l'image), la matrice $\mathbf{R}_\mathbf{C}$ ne peut être de rang plein. En effet les espaces de curvelettes sont des espaces redondants. Les curvelettes utilisées dans ces travaux ont une redondance de 7.2. Cette redondance dépend des éléments à l'échelle la plus fine (curvelettes ou ondelettes), du nombre d'orientations à cette échelle et du type de transformée en curvelette utilisée. La matrice $\mathbf{R}_\mathbf{C}$ (de taille $m \times m$ avec $m \simeq 7.2n$) n'est donc pas inversible et ne peut être utilisée de manière brute dans (4.3).

Contrairement à ce qui a été fait pour le gradients, nous ne cherchons pas à construire explicitement la vraie matrice de covariance d'erreur d'observation dans un repère de curvelettes (pour après coup construire une approximation inversible de cette matrice). Il existe différentes raisons nous poussant à cela :

1. La première (et pas nécessairement la meilleure) réside dans le fait que cette matrice est loin d'être creuse et qu'il est ainsi compliqué d'en extraire une bonne approximation (excepté l'approximation diagonale sur laquelle on reviendra en section 5.6.2). La figure 4.3 nous présente la matrice de corrélation dans l'espace des curvelettes correspondant à une matrice diagonale dans l'espace des pixels (pour une image 32×32 pixels). On observe que de nombreux éléments sont corrélés entre eux. Trouver une approximation inversible de cette matrice n'apparait pas dès lors comme une tâche aisée.

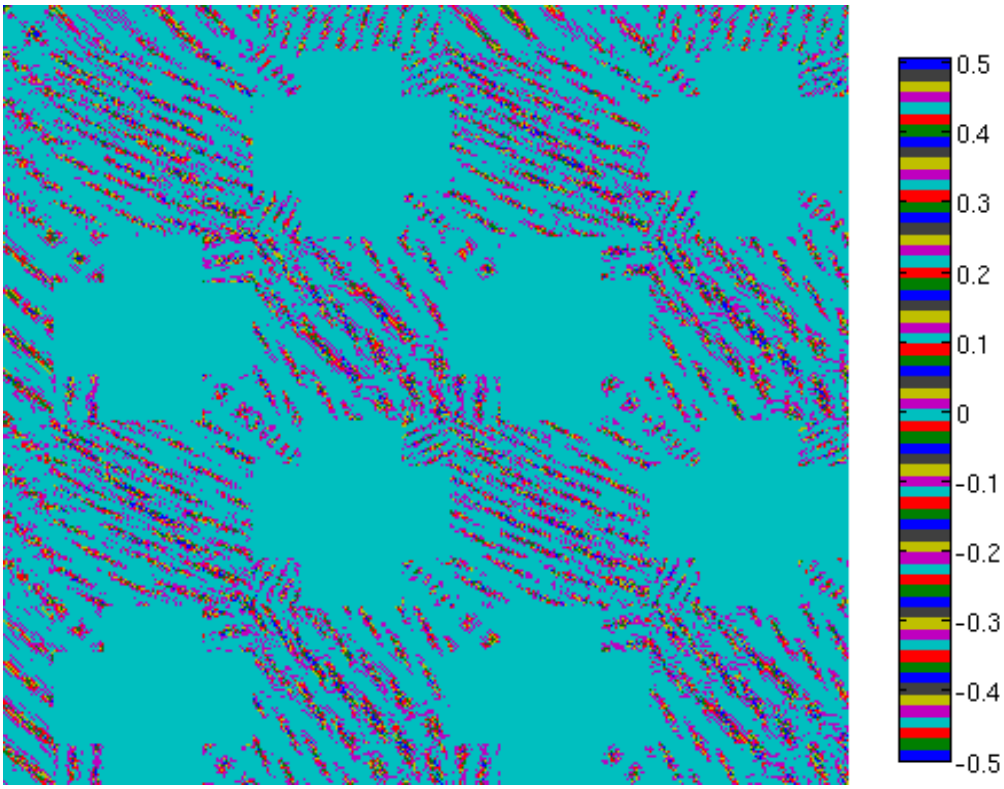


FIGURE 4.3 – Matrice de corrélation dans l'espace des curvlettes pour une matrice diagonale dans l'espace des pixels. La taille de l'image considérée est de 32×32 pixels ce qui conduit à travailler avec 7433 coefficients de curvelettes. Bien qu'il soit compliqué de distinguer la totalité des coefficients non nuls de la matrice, on observe que celle-ci est loin d'être diagonale.

2. La seconde (et certainement la meilleure) est qu'en réalité nous n'avons pas besoin de construire cette matrice de covariance. En effet, le problème que l'on souhaite résoudre est la minimisation de l'écart aux observations dans l'espace des pixels. Cette fonction

coût s'écrit (lorsque $\mathbf{R}_{pix} = \sigma \mathbf{I_n}$) de la manière suivante :

$$\begin{aligned} J^o_{pix} &= \sigma^{-2}(I^{synth} - I^o)^T(I^{synth} - I^o) \\ &= \sigma^{-2}(I^{synth} - I^o)^T\mathbf{C}^T\mathbf{C}(I^{synth} - I^o) \quad \text{car } \mathbf{C}^T\mathbf{C} = \mathbf{I_n} \\ &= \sigma^{-2}(\mathbf{C}I^{synth} - \mathbf{C}I^o)^T(\mathbf{C}I^{synth} - \mathbf{C}I^o) \\ &= (\mathbf{C}I^{synth} - \mathbf{C}I^o)^T\mathbf{K_C}(\mathbf{C}I^{synth} - \mathbf{C}I^o) \\ &= J^o_{\mathbf{C}}. \end{aligned} \tag{4.21}$$

où $\mathbf{K_C}$ est une matrice de *pondération* et non plus l'inverse de la matrice de covariance dans l'espace des curvelettes.

En effectuant ce développement on remarque qu'une possibilité est de choisir, dans l'espace des curvelettes, une matrice de pondération diagonale notée :

$$\mathbf{K_C} = \frac{1}{\sigma^2}\mathbf{I_m}. \tag{4.22}$$

Impact de la sélection d'informations

Lorsque la matrice de pondération $\mathbf{K_C}$ définie en équation (4.22) est utilisée, on a donc égalité entre l'écart aux observations calculé dans l'espace des pixels et celui calculé dans la trame de curvelettes :

$$\|\mathbf{C}I^{synth} - \mathbf{C}I^o\|^2_{\mathbf{K}_\mathbf{C}^{-1}} = \|I^{synth} - I^o\|^2_{\mathbf{R}_{pix}}. \tag{4.23}$$

Remarque 4.5

Il y a dans l'équation (4.23) un abus de notation concernant la matrice de pondération. L'écart aux observations défini dans l'espace des curvelettes ne nécessite pas de travailler avec une matrice de pondération inversible : $\|x\|_{\mathbf{K}_\mathbf{C}^{-1}} = x^T\mathbf{K_C}x$. *La notation* $\mathbf{K}_\mathbf{C}^{-1}$ *ne signifie donc pas que la matrice* $\mathbf{K_C}$ *est une matrice inversible mais que nous allons utiliser la matrice de pondération pour* modéliser *l'inverse d'une matrice de covariance.*

Remarque 4.6

La manière de définir $\mathbf{K_C}$ *de manière à satisfaire (4.21) n'est pas unique. Par exemple, la matrice de pondération*

$$\mathbf{K_C} = \frac{1}{\sigma^2}\mathbf{C}\mathbf{C}^T \tag{4.24}$$

satisfait aussi cette égalité. En effet dans ce cas,

$$\begin{aligned} J^o_{\mathbf{C}} &= (\mathbf{C}I^{synth} - \mathbf{C}I^o)^T\mathbf{K_C}(\mathbf{C}I^{synth} - \mathbf{C}I^o) \\ &= \frac{1}{\sigma^2}(I^{synth} - I^o)^T\mathbf{C}^T\mathbf{C}\mathbf{C}^T\mathbf{C}(I^{synth} - I^o) \\ &= J^o_{pix} \end{aligned}$$

car la trame des curvelettes est une trame de Parseval impliquant que l'opérateur adjoint C^T *est aussi une pseudo-inverse de* C.

Par conséquent il est possible d'évaluer l'impact de la sélection d'information effectuée dans la trame de curvelettes en comparant les résultats obtenus dans l'espace des pixels (3.8) et ceux obtenus en utilisant l'écart aux observations suivant :

$$\|\mathbf{P}(\mathbf{C}I^{synth}) - \mathcal{T}(\mathbf{C}I^o)\|^2_{\mathbf{K}^{-1}_{\mathcal{T}\mathbf{C}}} \tag{4.25}$$

où $\mathcal{T}$ correspond à l'opération de seuillage des observations définie en équation (3.29), $\mathbf{P}$ est la projection de l'équivalent modèle dans le sous-espace des coefficients conservés et $\mathbf{K}_{\mathcal{T}\mathbf{C}}$ est la projection de la matrice $\mathbf{K_C}$ sur ce sous-espace.

Bilan 4.3

Dans cette partie nous avons vu comment définir, dans le cas où les observations ne sont pas corrélées dans l'espace des pixels, une matrice de pondération permettant d'avoir l'égalité entre la fonction coût dans l'espace des pixels (3.8) et dans une trame de curvelettes.

Le second terme d'écart aux observations n'est pas obtenu via l'utilisation d'une matrice de covariance *mais d'une* matrice de pondération. *Cette matrice de pondération n'est ni unique (4.22) et (4.24) ni nécessairement inversible (4.24).*

On commence à entrevoir dans cette partie un des avantages (et une des difficultés) lié au fait que le travail s'effectue dans une trame (et non dans une base). On a ainsi plus de liberté pour définir une "bonne" matrice de pondération, dans le sens où l'on résout le même problème que dans l'espace des pixels.
La difficulté consiste à savoir comment trouver une approximation acceptable de cette "bonne" matrice $\mathbf{K_C}$ dans un cas où $\mathbf{R}_{pix}$ n'est pas une matrice diagonale. Cette difficulté est abordée en section 5.6.

4.2.4 Distance entre les gradients normalisés

Dans l'espace des gradients normalisés, les variances et les moyennes ne sont pas calculables analytiquement. En un pixel donné, les variances dépendent de la valeur des pixels environnants (en plus du bruit).

Comment allons nous procéder ?
Dans cette partie, nous allons dans un premier temps voir comment approcher la variance des gradients normalisés de l'image observée lorsque l'amplitude du bruit est négligeable devant le signal ($\|\nabla I^t\|^2 \gg \sigma^2$) ou lorsque le signal est lisse par rapport au bruit ($\|\nabla I^t\|^2 \ll \sigma^2$). Dans un second temps, nous allons définir, selon les valeurs de $\|\nabla I^o\|$ (le signal observé) si on peut considérer que le signal est prépondérant vis à vis du bruit (ou inversement). Enfin, cela nous mènera ,à partir des valeurs prises par $\|\nabla I^o\|$ et σ, à définir la variance des gradients normalisés.

Première étape : calcul des valeurs asymptotiques

1. Lorsque $\|\nabla I^t\| \gg \sigma$, c'est à dire que le bruit est négligeable devant l'ampleur des véritables gradients, on peut faire l'approximation suivante :

$$\begin{aligned} \mathrm{var}\left(\frac{I_x^o}{\|\nabla I^o\|}\right) &\simeq \mathrm{var}\left(\frac{(I_x^t + \epsilon_x)}{\|\nabla I^t\|}\right) \\ &= \frac{\mathrm{var}\,(\epsilon_x)}{\|\nabla I^t\|^2} \\ &= \frac{\sigma^2}{2\|\nabla I^t\|^2} \ll 1 \end{aligned} \tag{4.26}$$

quand le schéma utilisé pour le calcul du gradient est le schéma centré présenté en équation (4.11).

2. Lorsque $\|\nabla I^t\| \ll \sigma$, on peut faire l'approximation suivante :

$$\begin{aligned}
\mathrm{var}\left(\frac{I_x^o}{\|\nabla I^o\|}\right) \simeq \mathrm{var}\left(\frac{\epsilon_x}{\sqrt{\epsilon_x^2+\epsilon_y^2}}\right) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{\infty} \frac{\epsilon_x^2}{\epsilon_x^2+\epsilon_y^2} \exp^{-\frac{\epsilon_x^2+\epsilon_y^2}{2\sigma^2}} \frac{\mathrm{d}\epsilon_x \mathrm{d}\epsilon_y}{2\pi\sigma^2} \\
&= \int_0^{+\infty}\int_0^{2\pi} R\cos^2(\theta) \exp^{-\frac{R^2}{2\sigma^2}} \frac{\mathrm{d}\theta \mathrm{d}R}{2\pi\sigma^2} \\
&= \int_0^{+\infty} R \exp^{-\frac{R^2}{2\sigma^2}} \frac{\mathrm{d}R}{2\sigma^2} \\
&= \frac{1}{2}
\end{aligned} \tag{4.27}$$

Ce second développement suppose donc de savoir dans quel cas on peut considérer le gradient du signal négligeable devant le bruit.

Comment appliquer en pratique les calculs précédents ?

On cherche maintenant à savoir, connaissant σ^2, quelles sont les conditions sur $\|\nabla I^o\|$ nous menant à considérer que $\|\nabla I^{synth}\| \gg \sigma$.

Pour cela nous allons utiliser l'inégalité suivante (provenant de l'inégalité triangulaire) :

$$\|\nabla I^o\| \leq \|\nabla I^t\| + \|\nabla \boldsymbol{\epsilon}\| \tag{4.28}$$

Le problème que l'on souhaite résoudre peut être formulé de la manière suivante :

$$\begin{cases} \text{Quelles sont les conditions sur } \|\nabla I^o\| \text{ nous assurant d'avoir} \\ \mathbb{P}(\|\nabla I^t\| \geq k\sigma) \geq p_1 \end{cases} \tag{4.29}$$

avec k une constante choisie modélisant ce que l'on appelle négligeable.

En utilisant l'inégalité définie en (4.28) nous remarquons que

$$\mathbb{P}(\|\nabla I^t\| \geq k\sigma) \geq \mathbb{P}(\|\nabla I^o\| - \|\nabla \boldsymbol{\epsilon}\| \geq k\sigma) = \mathbb{P}(\|\nabla I^o\| - k\sigma \geq \|\nabla \boldsymbol{\epsilon}\|) \tag{4.30}$$

Au lieu de résoudre le problème (4.29) nous pouvons, à l'aide de cette inéquation, trouver des conditions suffisantes sur $\|\nabla I^o\|$ en résolvant le problème suivant :

$$\begin{cases} \text{Quelles sont les conditions sur } \|\nabla I^o\| \text{ nous assurant d'avoir} \\ \mathbb{P}(\underbrace{\|\nabla I^o\| - k\sigma}_{a} \geq \|\nabla \boldsymbol{\epsilon}\|) \geq p_1 \end{cases} \tag{4.31}$$

Pour cela remarquons que ϵ_x et ϵ_y sont indépendants et suivent chacun une loi gaussienne $\mathcal{N}(0, \frac{\sigma^2}{2})$. De ce fait, $\|\nabla \boldsymbol{\epsilon}\|$ suit une loi de Rayleigh (voir l'aide 6) de paramètre $\frac{\sigma}{\sqrt{2}}$ dont la fonction de répartition est :

$$\mathbb{P}(X \leq a) = 1 - \exp\left(\frac{-a^2}{\sigma^2}\right) \tag{4.32}$$

En posant $a = \|\nabla I^o\| - k\sigma$ le problème (4.31) se résume donc à trouver a tel que :

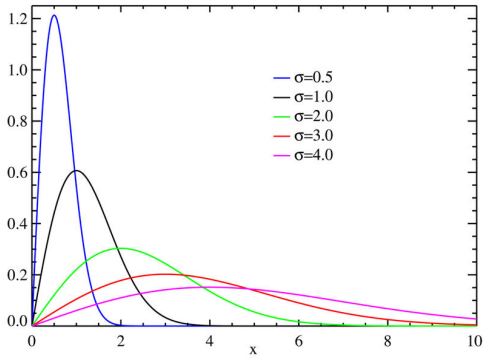
$$1 - \exp\left(\frac{-a^2}{\sigma^2}\right) \geq p_1 \tag{4.33}$$

$$\Leftrightarrow \quad \ln(1-p_1) \geq -\frac{a^2}{\sigma^2} \tag{4.34}$$

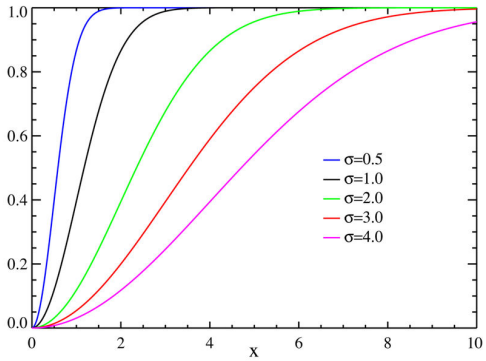
$$\Leftrightarrow \quad \sigma\sqrt{-\ln(1-p_1)} \leq a \text{ ou } -\sigma\sqrt{-\ln(1-p_1)} \geq a \tag{4.35}$$

Mémo 6 – Loi de Rayleigh

Densité de probabilité



Fonction de répartition



$$f(x;\sigma) = \frac{x}{\sigma^2}\exp\left(\frac{-x^2}{2\sigma^2}\right) \qquad P(X \leq a) = 1 - \exp\left(\frac{-a^2}{2\sigma^2}\right)$$

Par définition de la fonction de répartition, $a > 0$. Les solutions au problème (4.31) doivent donc vérifier :

$$\begin{aligned} & \|\nabla I^o\| - k\sigma \geq \sigma\sqrt{-\ln(1-p_1)} \\ \Leftrightarrow \quad & \|\nabla I^o\| \geq \sigma(k + \sqrt{-\ln(1-p_1)}). \end{aligned} \tag{4.36}$$

La figure 4.4 présente en trait plein (resp. en pointillé) les valeurs minimales de $\|\nabla I^o\|$ satisfaisant, en fonction de σ, l'inégalité (4.36) pour $k = 10$ (resp. $k = 5$) pour différentes valeurs de p_1.

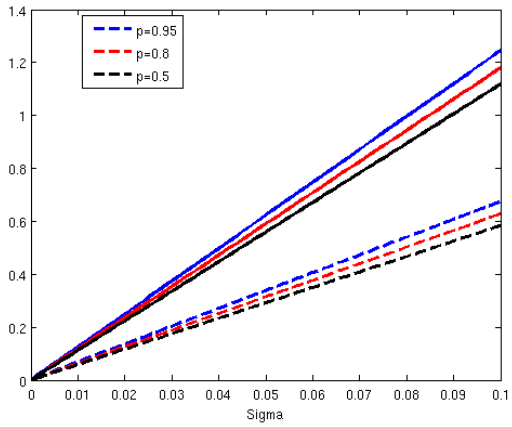


FIGURE 4.4 – Valeurs minimales de $\|\nabla I^o\|$ satisfaisant l'inégalité (4.36) en fonction de σ pour $k = 5$ (en pointillé) et $k = 10$ (en trait plein) pour 3 valeurs différentes de p_1.

On observe notamment que les valeurs limites de $\|\nabla I^t\|$ satisfaisant le problème (4.31) dépendent plus de la notion de *négligeable devant l'ampleur du gradient* que de la probabilité p_1 de respect de cette notion.

Bilan 4.4

Dans cet partie, nous avons donc trouvé des valeurs de $\|\nabla I^o\|$ nous permettant de s'assurer que $\|\nabla I^o\|^2 \gg \sigma^2$ avec une grande probabilité. Ces valeurs doivent être solutions de l'inéquation (4.36) dépendant de ce que l'on définit comme une ***forte probabilité*** *que le bruit soit* ***négligeable devant*** *le signal.*

Dans l'autre sens, afin de s'assurer de ne pas sous-estimer la variance du bruit, on considère $\|\nabla I^t\| \ll \sigma$ lorsque $\|\nabla I^o\| < 2\sigma$. Ces valeurs vont nous permettre de définir la variance dans l'espace des gradients normalisées.

Dans la suite, on choisit le couple $(k = 5, p_1 = 0.95)$. La variance d'un élément dans l'espace des gradients normalisés est définie comme suit :

$$\mathrm{var}\left(\frac{I_x^o}{\|\nabla I^o\|}\right) = \begin{cases} \dfrac{\sigma^2}{2\|\nabla I^o\|^2} & \text{lorsque } \|\nabla I^o\| \geq 6.73\sigma \\ \dfrac{1}{2} & \text{lorsque } \|\nabla I^o\| \leq 2\sigma \\ \dfrac{1}{2} + \dfrac{\|\nabla I^o\| - 2\sigma}{4.73\sigma}\left(\dfrac{\sigma^2}{2\|\nabla I^o\|^2} - \dfrac{1}{2}\right) & \text{sinon} \end{cases} \tag{4.37}$$

où la variance lorsque $2\sigma < \|\nabla I^o\| < 6.73\sigma$ est obtenue par interpolation linéaire entre les deux autres cas.

La figure 4.5 présente pour $\sigma = 0.05$ la variance ainsi définie dans l'espace des gradients normalisés.

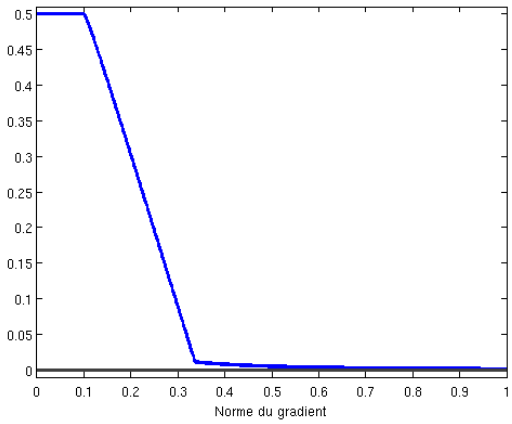


FIGURE 4.5 – Valeur de la variance des gradients normalisés en fonction de la valeur $\|\nabla I^o\|$ pour $\sigma = 0.05$.

Remarque 4.7

Au vu des difficultés rencontrées pour définir la variance dans cet espace là (et des approximations réalisées lors de cette définition) il va sans dire que nous ne cherchons pas à connaître (ni même à estimer) les corrélations entre les erreurs d'observations des différents éléments dans l'espace des gradients normalisés.

4.3 Expériences effectuées

Dans cette section, l'impact d'un bruit d'observation non corrélé en espace sur le processus d'assimilation est étudié. Des exemples d'images bruitées sont présentés en figure 4.6. Ces images correspondent aux différents niveaux de bruit utilisés. Dû à l'échelle de couleur (adaptée à la dernière image), le bruit est peu visible sur les deux premières images. Il est tout de même présent et peut notamment être vu au niveau du vortex sur la seconde image.

Rapport signal/bruit Dans le but de quantifier la fiabilité de l'information disponible dans ces séquences d'images, nous nous réferons au Rapport Signal/Bruit (SNR). Ce dernier est

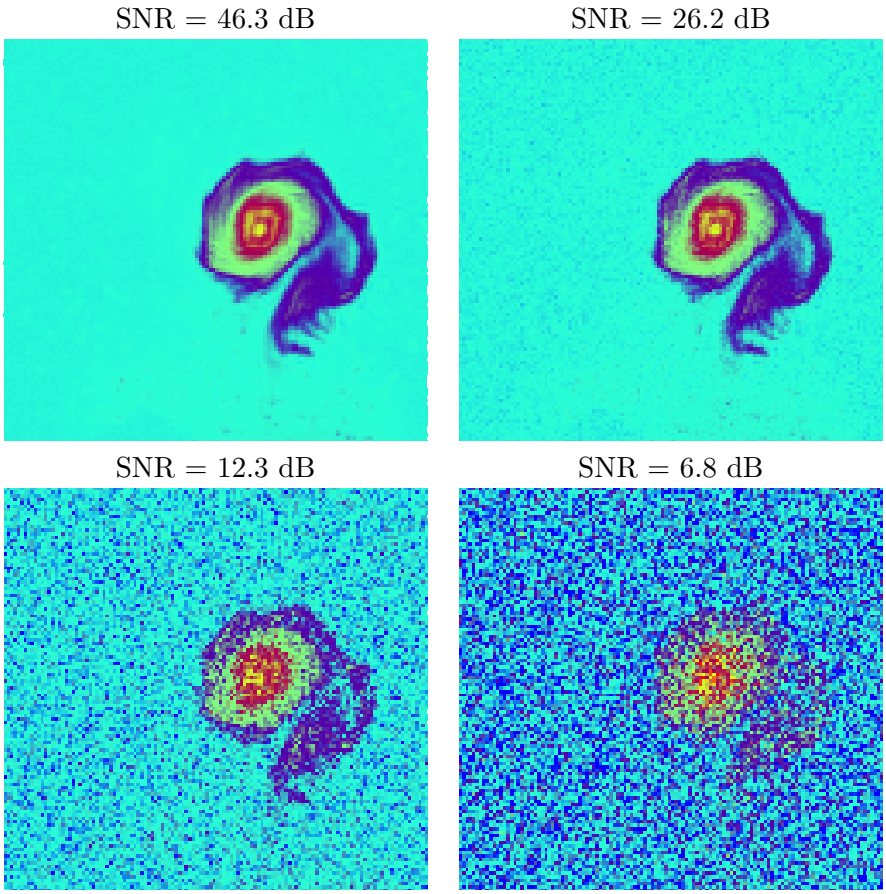


FIGURE 4.6 – Première image de la séquence bruitée pour 4 niveaux de bruits différents. La concentration de traceur passif va de 0 (en bleu) à 1 (en rouge). Pour rappel, la concentration de la séquence non bruitée est comprise entre 0.13 et 0.86.

calculé de la manière suivante :

$$SNR = 10\log_{10}\left(\frac{\sum_{\boldsymbol{x}\in\Omega}(I^t(\boldsymbol{x})^2)}{\sum_{\boldsymbol{x}\in\Omega}(I^t(\boldsymbol{x})-I^o(\boldsymbol{x}))^2}\right). \tag{4.38}$$

Il représente le logarithme du ratio entre l'énergie de l'image et l'energie du bruit. Comme nous pouvons l'observer en figure 4.6, une large valeur de SNR signifie que l'image est de très bonne qualité. A contrario, une faible (ou même négative) valeur signifie que l'image utilisée est fortement corrompue par le bruit.

4.3.1 Impact du niveau de bruit

Dans un premier temps, nous allons nous pencher sur la robustesse à un bruit décorrélé en espace des différents termes d'écarts aux observations introduits précédemment.

Expériences réalisées

But : Etudier l'impact d'un bruit décorrélé sur le processus de minimisation.

Observations utilisées : Chaque expérience d'assimilation est basée sur l'assimilation d'une séquence de 24 images. Le bruit est non biaisé, décorrélé en temps et en espace. La variance du bruit est homogène au sein de chaque séquence.

A savoir : Pour chaque niveau de bruit, 10 scénarios différents sont générés. Les résultats présentés dans les tableaux 4.3 sont les résultats moyens sur 10 expériences.

Autre paramètres : La longueur de corrélation de l'erreur d'ébauche est de 7.65 (pixels). La fenêtre d'assimilation est de 2 h 24. Le traceur passif est donc observé toutes les 6 minutes.

Minimisation : Le nombre d'itérations du processus de minimisation est fixé à 50, quel que soit le terme d'écart aux observations utilisé.

Le tableau 4.3 présente les valeurs finales du ratio $\frac{RMSE(v^a)}{RMSE(v^b)}$ en fonction du rapport signal bruit. Ces pourcentages permettent de quantifier l'amélioration imputable à la phase d'assimilation de la séquence de traceur passif (qui contient donc de l'information de manière indirecte sur la dynamique du système).

	Bruit fort			Bruit faible
SNR	6.8 dB	12.3 dB	26.2 dB	46.3 dB
$\mathcal{T}$Curv	54.8%	26.8%	6.7%	5.0%
$\mathcal{T}$Wav	48.1%	24.6%	6.7%	5.5%
Pixels-scal	48.0%	23.6%	6.1%	4.1%
∇-bloc	48.5%	21.4%	6.1%	4.1%
∇-scal	Divergence	36.6%	7.8%	3.8%
Normale-diag	74.6%	54.8%	29.4%	4.4%

Table 4.3 – Pourcentage d'erreur résiduelle dans la composante méridionale du champ de vitesses analysé en fonction du rapport signal bruit.

Résistance au bruit

On observe, comme attendu, que le résultat de la minimisation est influencé par le niveau du bruit. Aucun des termes d'écart aux observations introduits n'est épargné par ce phénomène bien que certains soient moins résistants que d'autres. Les termes les plus sensibles au bruit sont **Normale-diag** et **∇-scal**. Lorsque la comparaison entre l'image observée et l'image synthétique est effectuée entre gradients normalisés (**Normale-diag**), les résultats sont de mauvaise qualité et ce, même si l'amplitude du bruit est faible.

Capacité de compression des espaces multi-échelles

Les résultats découlant de l'utilisation d'une transformée multi-échelles en association avec un seuillage des coefficients (**$\mathcal{T}$Curv** et**$\mathcal{T}$Wav**) sont de qualité similaire à ceux obtenus dans l'espace des pixels. La compression du signal sur un faible jeu de coefficients (10% des coefficients sont conservés) permet de préserver l'information essentielle tout en compressant la taille du vecteur d'observation. Le taux de compression et les résultats obtenus étant de meilleure qualité dans une base d'ondelettes, l'utilisation de cet espace semble être une meilleure option si l'on désire compresser l'information.

Remarque 4.8

Le taux de compression réel dans l'espace des curvelettes est assez faible à cause de la redondance de l'information dans cet espace. Le vecteur d'observation étant 7 fois plus volumineux dans une trame de curvelettes, conserver 10% *de ces coefficients ne permet pas de compresser beaucoup l'information.*

Impact d'une meilleure modélisation des corrélations

Lorsque la comparaison entre gradients est employée, l'amélioration due à l'utilisation d'une matrice non diagonale est importante. Ces résultats montrent l'intérêt qu'on peut avoir à chercher une meilleure approximation de la matrice de covariance d'erreur d'observation. Cependant,

au vu de l'information perdue lors du passage dans cet espace, les résultats obtenus avec cette pseudo-métrique sont du même ordre de grandeur que ceux obtenus en travaillant dans l'espace des pixels.

Répartition spatiale des erreurs
La figure 4.7 présente la répartition spatiale des erreurs lors de l'assimilation d'une séquence d'images dont le SNR est de 12.3 dB.

Remarque 4.9

L'échelle de couleur est différente pour les résultats découlant de la comparaison entre gradients normalisés.

L'analyse utilisant la comparaison entre gradients normalisés commet une erreur importante au niveau du vortex. Cette erreur provient d'une sous-estimation de l'intensité de ce dernier : au centre du vortex, la vitesse estimée est presque deux fois plus faible que la véritable vitesse. En dehors de cette zone, de nombreuses erreurs sont aussi effectuées. Ces erreurs proviennent du manque d'information fiable dans les régions "lisses" de l'image lorsque ce terme d'écart aux observations est utilisé.

Les autres champs d'erreurs ont de nombreuses caractéristiques communes :
- La structure de ces dernières est plus large en dehors de la zone du vortex,
- L'amplitude des erreurs est moins importante au centre du vortex,
- La structure et la position des erreurs d'analyse commises se ressemblent. C'est particulièrement le cas pour la distance entre pixels et la pseudo-métrique entre gradient (associée à la matrice par blocs diagonaux) qui semblent être des copies (presque) conformes.

On s'aperçoit donc en visualisant ces cartes d'erreurs, qu'elles diffèrent surtout par l'amplitude des différentes structures présentes en leur sein. Le terme d'écart aux observations a assez peu d'importance quant à la structure de ces erreurs (si on excepte la comparaison entre gradients normalisés).

Bilan 4.5

Finalement, nous observons donc qu'aucun terme d'écart aux observations n'est résistant à un bruit décorrélé en espace. Les meilleurs résultats sont obtenus en travaillant dans l'espace des pixels. Lorsque la compression n'est pas une finalité, travailler dans un autre espace ne nous apporte donc pas grand chose dans ce cas là.

4.3.2 Intérêt de la perte d'information en présence de biais

En plus du bruit à proprement parler, les observations utilisées peuvent comporter un biais. En pratique, ce biais est estimé au préalable, de manière à corriger les observations avant leur intégration dans le système d'assimilation. Cependant, une partie de ce biais peut toujours être présente après ce pré–traitement. Les expériences suivantes ont pour but d'illustrer le comportement des différentes distances en présence d'un biais résiduel au sein des images observées.

Description des expériences

But : Evaluer l'impact d'un biais présent au sein des images observées.

Observations utilisées : Un bruit de faible amplitude a été ajouté à la séquence d'images parfaites. Le niveau de bruit est identique à celui utilisé pour générer les séquences d'images ayant un SNR de 26.2 dB. Un biais est ajouté à chaque observation de la séquence d'images. Ce biais a une valeur de 0.02. Pour comparaison, la concentration du traceur est comprise entre 0.13 et 0.86.

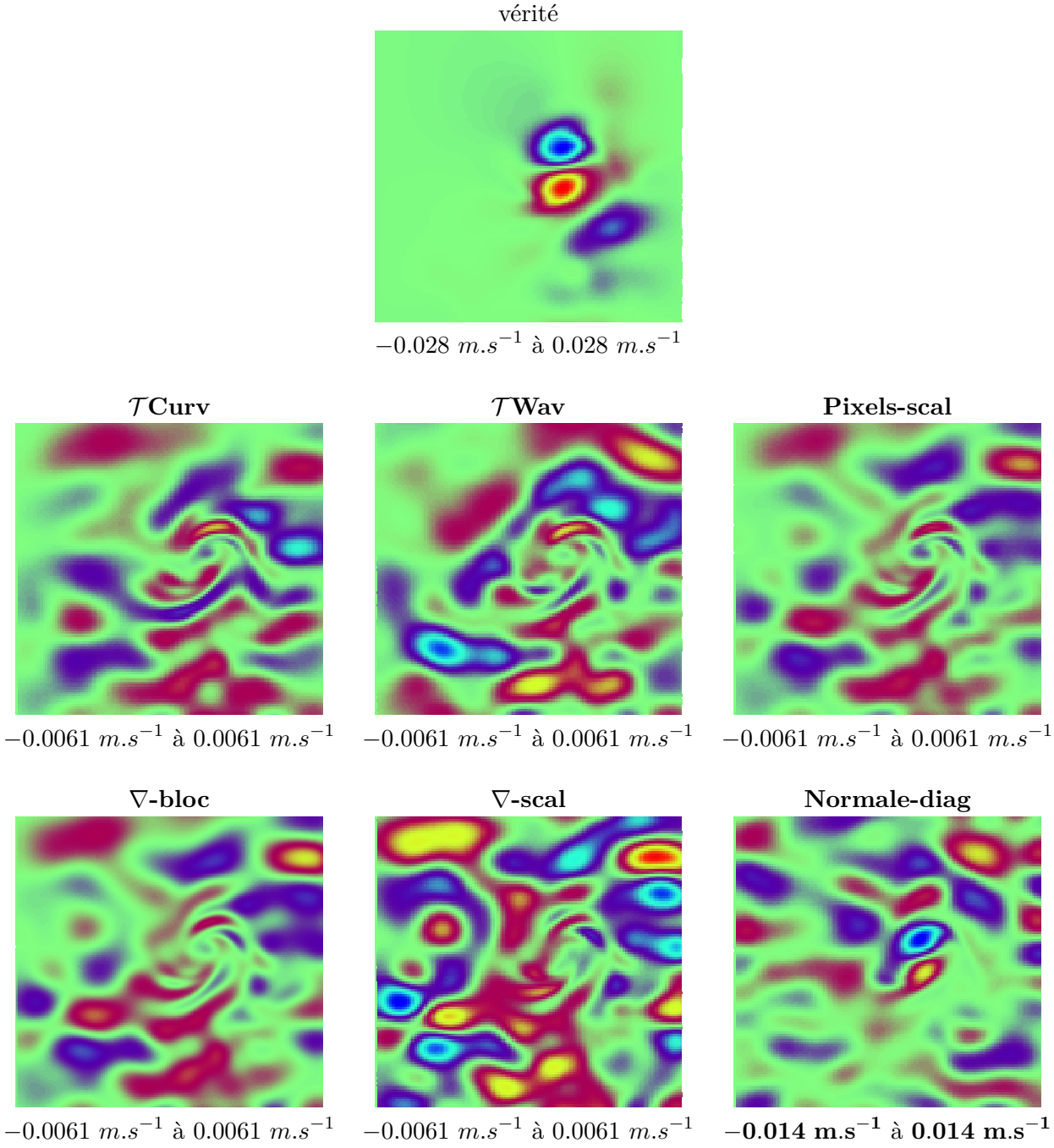


FIGURE 4.7 – Erreur sur la composante zonale du champ de vitesses analysé. Le rapport signal/-bruit de la séquence d'images assimilée est de 12.3 dB.

Autres paramètres : La fenêtre d'assimilation est de 2 h 24. Une image est observée toutes les 6 minutes.

Le pourcentage d'erreur résiduelle en fin d'analyse est présenté dans le tableau 4.4 pour les différentes composantes étudiées.

Remarque 4.10

Lors du seuillage dans la base d'ondelettes, nous avons tenu compte du fait que l'information était biaisée (sans connaitre la valeur du biais). Ceci peut être fait aisément dans ces espaces. En effet, dans une base d'ondelettes (ou de curvelettes) les coefficients de détails ne sont pas affectés par un biais. Seuls les coefficients d'approximation le sont. Par conséquent, ces derniers n'ont pas été utilisés lors de la comparaison entre image observée et image synthétique. De cette manière, seule l'information non biaisée est utilisée (sans avoir à estimer ce biais).

En comparant ces résultats à ceux obtenus précédemment, nous pouvons vérifier que tous les termes d'écarts aux observations sont (ou peuvent être rendus) résistants à un biais d'observation. La seule exception est la comparaison pixel à pixel utilisant l'ensemble de l'information présente dans les observations. En "appauvrissant" l'information contenue dans la séquence d'images, les autres termes d'écart aux observations peuvent ne pas être sensibles à ce biais.

	u	v	ζ
Pixels-scal	87.9%	70.6%	52.4%
∇-scal	10.0%	7.5%	15.6%
∇-bloc	8.0%	6.0%	14.2%
Normale-diag	37.6%	30.2%	54.4%
Ondelettes	9.5%	7.8%	15.6%

Table 4.4 – Effet d'un biais (en plus du bruit) sur l'analyse. Le pourcentage présenté est la valeur moyennée sur 10 expériences issues de 10 tirages du bruit différent.

Remarque 4.11

Les résultats induits par la comparaison dans un espace d'ondelettes sans tenir compte des coefficients d'approximation sont comparables à ceux obtenus dans l'espace des gradients (avec la matrice scalaire). Si les coeffcients d'approximations (4 dans cet exemple) avaient été conservés, les résultats basés sur cette pseudo-métrique auraient été comparables à ceux obtenus dans l'espace des pixels.

4.3.3 Impact de la fréquence d'observation

Depuis le début de l'observation de la Terre par satellite, la fréquence d'acquisition des données n'a de cesse d'augmenter. Dans cette section, des expériences (irréalistes) ont été effectuées afin d'évaluer l'impact du nombre d'observations assimilées (sur une courte) période sur l'analyse effectuée.

Description des expériences

Observations utilisées : Une séquence d'images bruitées. Le rapport signal bruit de la séquence est de 6.8 dB. Cela correspond donc au scénario le plus bruité. Le bruit est décorrélé en temps et en espace et non biaisé.

Fréquence d'observations : Le paramètre variant d'une expérience à l'autre est la fréquence *d'acquisition* des images. Cette dernière varie entre une image toutes les 14.4 s (ce qui correspond à un pas temps) et une image toutes les heures.

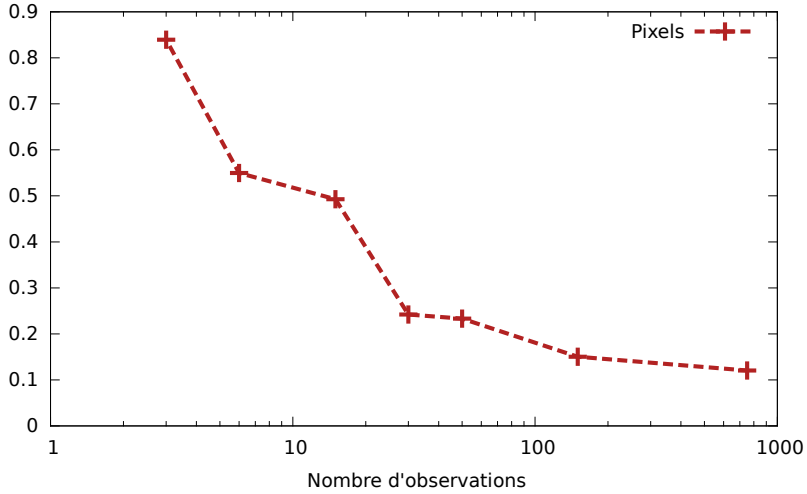


FIGURE 4.8 – Evolution du pourcentage d'erreur résiduelle (sur la composante méridionale de la vitesse) en fonction du nombre d'observations utilisées lors de la phase d'assimilation.

Fenêtre d'assimilation : La fenêtre d'assimilation est de 3 h. Sur cette période, entre 3 et 750 observations sont donc utilisées selon les expériences.

Autres paramètres : La longueur de corrélation des erreurs d'ébauche est de 7.65 (pixels). Le nombre d'itérations de minimisation est de 50.

La figure 4.8 présente l'évolution de l'erreur restant en fin d'assimilation en fonction du nombre d'observations utilisées. On observe ici que, comme attendu, plus nous possédons d'information (dont les erreurs sont décorrélées), meilleure est l'analyse. La décroissance de l'erreur d'analyse est rapide : l'erreur est environ de 55% lorsque 6 observations sont utilisées et de 25% lorsque 30 observations sont utilisées. Après cette limite, l'apport d'information dû à chaque image est beaucoup plus faible. Ces expériences permettent d'illustrer le fait que nous pouvons *en partie compenser* la mauvaise qualité des observations utilisées par leur nombre.

Remarque 4.12

Ces résultats sont cependant à relativiser : ils sont en partie imputables à l'incohérence temporelle des erreurs d'observations. C'est cette trop grande incohérence (temporelle et spatiale) des erreurs qui permet au processus de minimisation de ne pas trop être "attiré" par ces dernières.

5

Impact d'un bruit d'observation corrélé en espace

Sommaire

5.1 Introduction **93**
5.2 Pixels, gradients et gradient normalisés **97**
5.2.1 Approximation dans une base de pixels 97
5.2.2 Approximation dans l'espace des gradients 97
5.2.3 Approximation pour les gradients normalisés 99
5.3 Approximation scalaire dans une base d'ondelettes **99**
5.3.1 Construction 99
5.3.2 Pourquoi est-ce une mauvaise approximation ? 100
5.4 Construction formelle de la matrice de covariance dans une base d'ondelettes **103**
5.4.1 Obtention de la matrice de covariance en mode boîte noire 103
5.4.2 Utilisation de la connaissance de la transformée en ondelette 103
5.4.3 Mais au final, à quoi ça ressemble ? 110
5.4.4 Propriété facilitant la construction lorsque le bruit est identiquement distribué 111
5.5 Approximation diagonale dans une base d'ondelettes **113**
5.5.1 Impact de la diagonalisation 113
5.5.2 Renormalisation 116
5.5.3 Propriété intéressante pour la renormalisation (et autres) 117
5.6 Estimation de la matrice dans une trame de curvelettes **119**
5.6.1 Utilisation de matrice de poids symétrique définie positive 119
5.6.2 Restriction à des matrices diagonales 122
5.7 Estimation de la matrice de covariance dans un espace de Fourier . **128**
5.8 Expériences **129**
5.8.1 Influence du niveau de bruit 131
5.8.2 Impact de R sur le choix de B 139
5.8.3 Quelle quantité d'information est contenue dans les statistiques d'erreurs d'observation ? 142
5.8.4 Attention danger : statistiques à manipuler avec précaution ! 143
5.8.5 Anisotropie et choix de la représentation 147

5.1 Introduction

Dès lors que les observations sont corrompues par un bruit corrélé en espace, le bruit lui-même a une certaine structure. La connaissance de cette structure apporte une information

importante que l'on doit utiliser lors de la phase d'analyse pour tirer parti au maximum du véritable signal présent au sein des images.

Pourquoi est-ce important d'utiliser l'information sur les corrélations d'erreur d'observation ?

Mettons nous un instant à la place d'un algorithme de minimisation dont le but est de régler certains paramètres pour s'approcher au plus près d'un jeu d'observations.

Plaçons nous dans le cas où nous disposons de deux observations dans lesquelles nous avons autant confiance et supposons que nous soyons assez proche de la solution. La figure 5.1 présente, à partir de l'état actuel, 4 états possibles après des changements de paramètre de même norme. Sans plus d'information sur les statistiques d'erreurs d'observations, n'importe lequel de ces états nous apparait comme étant mieux que l'état initial. On souhaite uniquement se rapprocher au plus près des observations tout en changeant assez peu les paramètres initiaux.

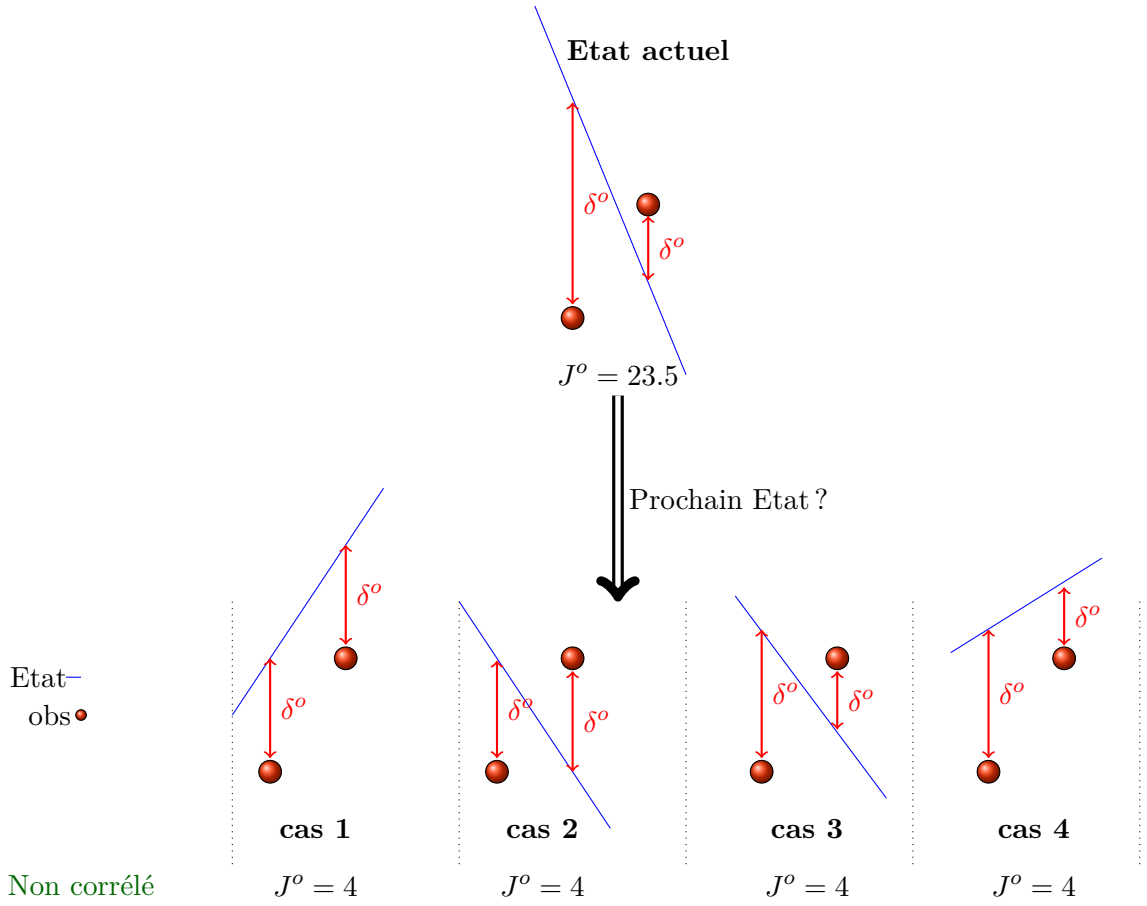


FIGURE 5.1 – Etre un minimiseur n'est pas chose aisée ...

Supposons maintenant savoir que les erreurs d'observations en ces deux points sont fortement corrélées (de manière positive). Les possibilités 2 et 3 deviennent, si nous savons être proches du véritable signal, des choix aberrant. En effet, il y a de grande chance pour que l'écart entre le véritable état et les observations soit de même signe en ces deux points. En utilisant cette information supplémentaire, on s'oblige donc à s'approcher des observations de manière cohérente avec les statistiques d'erreurs.

Lorsque de nombreuses informations sont disponibles, ne pas tenir compte des corrélations existantes peut déteriorer l'analyse. Un groupe d'observations fortement corrélées (de manière

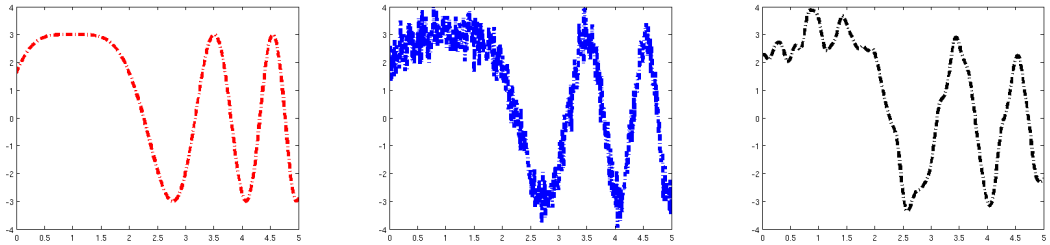


FIGURE 5.2 – A gauche le vrai signal est présenté. Le signal est perturbé par un bruit additif gaussien non corrélé (au centre) et corrélé (à droite). L'énergie du bruit est la même dans les deux cas.

positive) peut, si les corrélations sont ignorées, donner l'illusion d'être en accord et de fait tirer l'analyse vers lui. L'exemple 5.1 illustre le problème de distinction entre signal et bruit lorsque ce dernier est fortement corrélé en espace.

Exemple 5.1

> *La figure 5.2 présente un signal et ce même signal perturbé par un bruit additif gaussien non corrélé (au centre) et corrélé (à droite). Sans connaissance à priori sur les statistiques du bruit, à l'oeil nu, le signal perturbé par un bruit corrélé en espace peut apparaître comme le vrai signal alors qu'il est assez clair que celui perturbé par un bruit décorrélé est bruité.*
>
> *Cette distinction que nous ne sommes pas capables de faire à l'oeil nu, le processus de minimisation ne sera pas capable de la faire si nous lui fournissons une fausse information sur la distribution du bruit.*
> *Si, par exemple, nous indiquons que le bruit n'est pas corrélé en espace, il y a fort à parier que l'état analysé tende à se rapprocher des deux bosses de grande amplitude qui apparaissent comme cohérentes (vis–à–vis des statistiques d'erreurs d'observations spécifiées). Lui spécifier les échelles de corrélation du bruit l'empêchera (en partie) d'effectuer ceci.*
> *Le cas contraire est aussi problématique : fournir l'information que le bruit est lisse, conduit à considérer que les oscillations dues au bruit décorrélé en espace font partie du signal.*

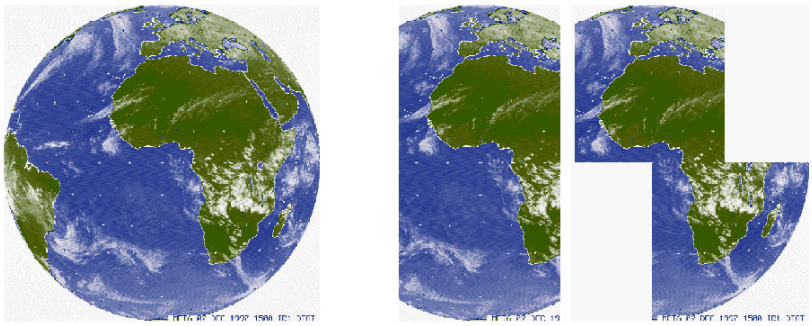
Afin d'aider notre ami l'algorithme de minimisation, il est donc primordial d'incorporer la connaissance sur la structure du bruit que l'on possède. Cette connaissance permet une minimisation cohérente utilisant l'ensemble de l'information disponible au sein des observations. Une conséquence fâcheuse si on ne le fait pas réside dans le fait qu'utiliser plus d'observations ne mène pas nécessairement à une meilleure analyse [Dando et al., 2007].

En intégrant les corrélations d'erreurs d'observation dans la phase d'analyse, on spécifie au minimiseur qu'il existe des manières privilégiées de réduire l'écart entre l'état analysé et les observations (figure 5.1).

Pourquoi est-ce compliqué à utiliser ?

Bien que cette information sur les corrélations soit extrèmenent intéressante, elle est en pratique compliquée à utiliser. Il y a différentes raisons à cela :

1. Tout d'abord, il est nécessaire d'estimer les corrélations existantes au sein du champ d'erreur du signal observé. Différents travaux s'intéressent à estimer les statistiques d'erreur d'observation, par ex. au sein des AMV pour [Bormann et al., 2003, Bormann et al., 2014] ou des températures de brillance pour [Stewart, 2010, Weston, 2011, Stewart et al., 2013a]. Cependant, estimer ces corrélations est une tâche complexe, et ce, en partie à cause des différentes sources d'erreurs d'observations. En effet, comme recensé par exemple dans [Stewart, 2010, Weston et al., 2014], les corrélations peuvent provenir entre autre des erreurs de mesure (la prise de mesure d'un instrument), des erreurs de pré–traitement



All Channels except HRV HRV (nominal and alternative coverage)

FIGURE 5.3 – Couverture des images satellites (nommés Level 1.5 images data) qui constitue un des produits majeurs fournis par le système MSG (Meteosat Second Generation). A gauche la couverture pour tous les canaux à l'exception des canaux haute résolution (la résolution est de 3 km sous le satellite). A droite, la couverture des canaux haute résolution (la résolution est de 1 km sous le satellite).

(englobant l'ensemble des traitements effectués entre l'information receuillie et celle utilisé au cours de l'assimilation) ou des erreur de représentativités (provenant par exemple des échelles non résolues par le modèle).

2. Lorsqu'on accède à la connaissance des corrélations, il n'est pas évident de l'utiliser : la fonction coût définie en équation (4.3) utilise l'inverse de la matrice de covariance d'erreur d'observation.

   Par exemple, pour tenir compte des corrélations présentes dans une unique image du satellite METEOSAT 9 [12], il est nécessaire d'inverser une matrice de taille $\sim 10^6 \times 10^6$. La taille de la matrice à inverser augmente encore sensiblement lorsqu'on souhaite prendre en compte la dimension temporelle des corrélations ou les corrélations entre les erreurs effectuées sur les différents canaux.

3. Supposons ce problème d'inversion résolu (ainsi que le problème de stockage de la matrice inverse). Le problème suivant réside dans le coût induit par la multiplication entre une matrice pleine et un vecteur (celui-ci est de l'ordre de $O(n^2)$ où n est le nombre d'éléments du vecteur). Au vu du faible temps disponible pour effectuer la minimisation, il est crucial de réduire ce coût au minimum en utilisant par exemple des matrices inverses creuses [Stewart et al., 2013b] .

4. Lorsque les trois premiers points sont réglés, le dernier problème est dû au fait que les observations disponibles ne sont pas toujours les mêmes. Le travail fait précédemment doit pouvoir être adapté à chaque nouveau "jeu" d'observations. Par exemple, concernant les observations satellites de l'océan, il n'est pas rare que des nuages viennent cacher certaines zones des images observées. Par conséquent, une approximation de la matrice de covariance construite dans le cadre où toutes les observations sont disponibles peut ne pas pouvoir s'appliquer lorsque seule une partie du signal est disponible. Il est donc nécessaire de pouvoir tenir compte de la spécificité de chaque image et de pouvoir définir l'inverse de la matrice de covariance lorsque certaines observations viennent à manquer.

Les 3 premiers points sont communs à la formation d'une matrice de corrélation d'erreur d'ébauche et à la matrice de covariance d'erreur d'observation. Le dernier point est quant à lui spécifique aux matrices d'erreur d'observation.

12. Les chiffres précédent proviennent d'un document technique disponible sur http://www.eumetsat.int/website/home/Data/TechnicalDocuments/index.html disponible dans la section MSG sous section 0 ° SEVIRI Level 1.5 Image Data.

Axe de travail

Dans la suite de ce manuscrit, on suppose le problème d'estimation de la matrice de covariance résolu et donc d'avoir à disposition une matrice de covariance d'erreur d'observation (ou tout au moins avoir accès aux opérateurs pouvant modéliser cette matrice de covariance).

Le problème lié au coût de calcul induit par la multiplication matrice vecteur dans l'équation (4.3) est ici résolu en imposant l'utilisation de matrice de covariance diagonale (quelque soit l'espace de travail).

Dans ce chapitre, nous abordons l'intérêt des espaces multi-échelles afin de modéliser par des matrices diagonales des matrices de corrélations complexes. Il est en effet montré en section 5.5 et 5.6 que de telles matrices peuvent reproduire de manière honorable des matrices de corrélations complexes dans l'espace des pixels.

La gestion de la variation de l'information disponible d'une image à l'autre est abordée au chapitre 6. Nous verrons un moyen peu coûteux d'adapter une approximation diagonale dans une base d'ondelettes lorsque des nuages viennent masquer une partie de l'image.

5.2 Pixels, gradients et gradient normalisés

5.2.1 Approximation dans une base de pixels

Mémo 7 – Notations

Pour rappel les notations employées pour les matrices de covariances sont expliquées au tableau 4.2. La matrice $\tilde{\mathbf{R}}^{\text{cor}}_{pix,\text{scal}}$ doit être comprise comme étant une approximation scalaire (scal) de la véritable matrice de covariance dans l'espace des pixels $\boldsymbol{pix}$.
cor indique que la véritable matrice de covariance dans l'espace des pixels n'est pas diagonale.

Lorsque le bruit est réparti de manière homogène en espace (i-e que la variance en chaque pixel est identique), et que l'approximation faite est l'approximation diagonale, la matrice de covariance correspondante dans l'espace des pixels est :

$$\tilde{\mathbf{R}}^{\text{cor}}_{pix,\text{scal}} = \sigma^2 \mathbf{I_n} = \text{diag}(\mathbf{R}_{pix}) \tag{5.1}$$

où σ^2 est la variance en un pixel donné. Cette approximation conduit à considérer que le bruit est indépendant en espace (et en temps), c'est à dire que l'erreur en un point donné ne nous fournit aucune information sur une erreur faite en un autre point. De ce fait, en considérant cette matrice de covariance d'erreur d'observation, nous nous privons d'une large partie de l'information correspondant aux interactions entre les erreurs en différents points.

Remarque 5.1

Pour certains types de bruit, il est possible de construire de meilleurs approximations (non diagonales) dans l'espace des pixels. Par exemple, dans [Stewart et al., 2013b] il est proposé d'utiliser, en 1D, des matrices de Markov. Ces matrices (pleines) sont telles que leur inverse est une matrice tridiagonale. Cependant cette propriété ne semble pas généralisable en plusieurs dimensions.

5.2.2 Approximation dans l'espace des gradients

Nous pouvons, en employant la même méthode qu'en section 4.2.1, calculer la variance de chacune des composantes dans l'espace des gradients.[13]

13. on rappelle que l'implémentation des dérivées partielles est faite à l'aide d'un schéma de différences finies centrées.

La variance de $I_x(i,j)$ s'écrit :

$$\begin{aligned}\mathrm{var}\left(I_x(i,j)\right) &= \mathrm{cov}\left(\frac{I(i+1,j)-I(i-1,j)}{2}, \frac{I(i+1,j)-I(i-1,j)}{2}\right)\\ &= \mathrm{var}\left(\frac{I(i+1,j)}{2}\right) + \mathrm{var}\left(\frac{I(i-1,j)}{2}\right) - 2\frac{\mathrm{cov}\left(I(i+1,j), I(i-1,j)\right)}{4}\end{aligned}$$

En notant $\mathrm{var}\left(I(i+1,j)\right) = \mathrm{var}\left(I(i-1,j)\right) = \sigma^2$ et $\mathrm{cov}\left(I(i+1,j), I(i-1,j)\right) = \sigma^2\alpha$ (où alpha est la corrélation entre les erreurs effectuées en $I(i+1,j)$ et en $I(i-1,j)$) on obtient :

$$\mathrm{var}\left(I_x\right) = \sigma^2\frac{1-\alpha}{2} = \tilde{\sigma}^2 \tag{5.2}$$

Remarque 5.2 (*Comportement de* var (I_x))

Plus α est proche de 1, plus une large part de l'erreur d'observation en $I(i+1,j)$ est identique à l'erreur effectuée en $I(i-1,j)$. De ce fait, plus la variance du bruit dans l'espace des gradients est faible (via la soustraction, on élimine l'erreur "commune" à $I(i+1,j)$ et $I(i-1,j)$).

De plus, si les corrélations sont isotropes (i-e identiques dans chacune des directions), on a $\mathrm{var}\left(I_x\right) = \mathrm{var}\left(I_y\right) = \tilde{\sigma}^2$. L'approximation diagonale dans l'espace des gradients, correspond donc à une matrice de covariance proportionnelle à l'identité :

$$\tilde{\mathbf{R}}^{\mathrm{cor}}_{\nabla,\mathrm{scal}} = \tilde{\sigma}^2\mathbf{I} \tag{5.3}$$

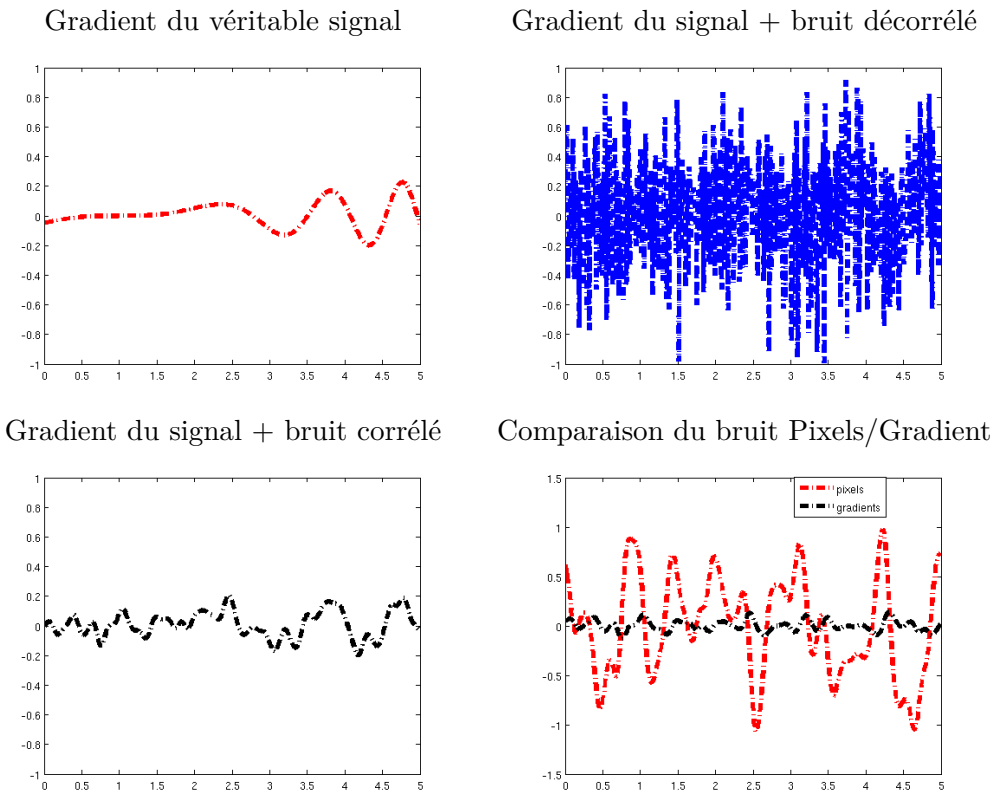


FIGURE 5.4 – Illustration de l'effet d'un bruit corrélé et non corrélé dans l'espace des gradients pour les versions du signal présentées en figure 5.2. La dernière figure présente la comparaison du bruit dans l'espace des pixels et des gradients.

En quoi travailler avec les gradients peut-il être un avantage ?

La figure 5.4 présente le signal dans l'espace des gradients pour les 3 cas abordés en figure 5.2.
Bien que le bruit corrélé et non corrélé aient une variance identique dans l'espace des pixels, on observe que leur variance n'est absolument pas la même dans l'espace des gradients (voir équation 5.2). Cette dernière est fortement réduite lorsque le bruit est corrélé (positivement).
La figure en bas à droite montre l'amplitude du bruit dans l'espace des pixels (rouge) et dans l'espace des gradients (noir). On observe que l'amplitude du bruit est bien moins grande dans l'espace des gradients que dans l'espace des pixels bien que le bruit semble autant corrélé dans les deux espaces. Il y a donc des raisons d'espérer que la solution analysée, basée sur un terme d'écart aux observations dans cet espace, soit moins perturbée par un bruit corrélé que lorsque l'écart aux observations est calculé entre pixels.

Remarque 5.3

L'approximation par bloc diagonaux effectuée dans le cas d'observation dont l'erreur est non corrélée (voir section 4.2.1) n'est dans ce cas plus valide. De nombreuses autres corrélations entre les différentes composantes du gradients apparaissent lorsque le bruit est à l'origine corrélé en espace.
Cette approximation conduisait à considérer que les "oscillations" du bruit autour du véritable état (dans l'espace des gradients) avaient de fortes chances d'être dans des directions opposées tous les deux points (ce qui est une "traduction grossière" de corrélation négative entre l'erreur faite en $I_x(i)$ et $I_x(i+2)$).
Apporter cette information dans le cas où le bruit est corrélé de manière positive (dans l'espace des pixels) induit donc l'utilisation d'une information supplémentaire sur la distribution du bruit erronée comme l'illustre la figure 5.4.

5.2.3 Approximation pour les gradients normalisés

Concernant les gradients normalisés, nous fixons toujours la variance en fonction de la norme du gradient :

$$\mathrm{var}\left(\frac{I_x^o}{\|\nabla I^o\|}\right) = \begin{cases} \dfrac{\sigma^2}{2\|\nabla I^o\|^2} & \text{lorsque } \|\nabla I^o\| \geq 6.73\sigma \\ \dfrac{1}{2} & \text{lorsque } \|\nabla I^o\| \leq 2\sigma \\ \dfrac{1}{2} + \dfrac{\|\nabla I^o\| - 2\sigma}{4.73\sigma}\left(\dfrac{\sigma^2}{2\|\nabla I^o\|^2} - \dfrac{1}{2}\right) & \text{sinon} \end{cases} \tag{5.4}$$

Le bruit étant moins répercuté dans l'espace des gradients lorsque le bruit est corrélé, ces variances sont sur–estimées. Cela devrait permettre de s'appuyer sur des informations plus fiables que dans le cas où le bruit est décorrélé.

5.3 Approximation scalaire dans une base d'ondelettes

5.3.1 Construction

Dans un espace de type ondelette, l'approximation la plus simple consiste à prendre l'équivalent de l'approximation dans l'espace des pixels :

$$\begin{aligned} \tilde{\mathbf{R}}^{\mathrm{cor}}_{\mathbf{W},\mathrm{scal}} &= \mathbf{W}\mathrm{diag}(\mathbf{R}_{pix})\mathbf{W}^T && (5.5) \\ &= \mathbf{W}\tilde{\mathbf{R}}^{\mathrm{cor}}_{pix,\mathrm{scal}}\mathbf{W}^T \\ &= \sigma^2\mathbf{I_n} && (5.6) \end{aligned}$$

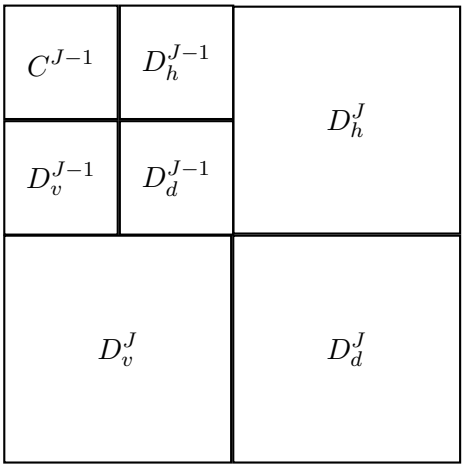


FIGURE 5.5 – Pour rappel, la disposition des coefficients lors d'une décomposition en ondelettes 2D. Le bloc D_d^j (resp D_v^j et D_h^j) contient les coefficients de détails diagonaux (resp. verticaux et horizontaux) à l'échelle j. Le bloc C^{j-1} contient les coefficients d'approximations à l'échelle $j-1$. Dans ce cas, seuls deux niveaux de décompositions ont été effectuées.

Ceci revient à considérer que le bruit agit de la même manière sur chacun des coefficients dans l'espace d'ondelettes.

Remarque 5.4

Lorsque ce choix de modélisation est effectué pour les matrices (5.6) et (5.1), la fonction d'écart aux observations dans l'espace des pixels est identique à celle dans l'espace des ondelettes (voir remarque 3.13). Par conséquent, cette matrice est utilisée uniquement en combinaison avec une opération de seuillage (voir équation (3.31)). De cette manière, nous pouvons évaluer l'impact de la compression de l'information sur la minimisation.

Cependant, comme l'illustre l'équation (5.5), cette approximation effectue l'hypothèse diagonale avant d'avoir "*transporté*" la matrice de covariance dans l'espace des ondelettes. Ce défaut dans l'approche est important car il revient à se priver de l'information supplémentaire que peut nous fournir un espace structuré. Effectuer l'hypothèse diagonale *avant* d'avoir procédé au changement d'espace revient à supposer que le bruit est décorrélé en espace (dans l'espace des pixels).

La figure 5.6 permet d'illustrer la perte d'information due à cette approche. Les images du haut présentent un tirage de bruit blanc gaussien corrélé (à gauche) et non corrélé (à droite) dans l'espace des pixels. Les images du bas présentent la décomposition de ces tirages dans une base d'ondelettes (plus précisément le logarithme de la valeur absolue des coefficients). On observe, sur une unique réalisation du bruit, que la décomposition de ces deux types de bruit diffère fortement. Lorsque le bruit est corrélé en espace, la valeur absolue des coefficients dépend fortement de l'échelle et de l'orientation de la fonction d'ondelettes (ce qui n'est pas le cas pour un bruit décorrélé en espace).

En supposant que les statistiques du bruit dans un espace d'ondelettes sont les mêmes dans ces deux cas, on perd donc une information importante : les différents coefficients d'ondelettes ne sont pas affectés de la même manière lorsque le bruit est corrélé en espace.

5.3.2 Pourquoi est-ce une mauvaise approximation ?

Afin de confirmer le fait que la variance du bruit dépend bien de l'échelle et de l'orientation regardée, nous pouvons estimer la variance des coefficients d'ondelettes. Dans le but d'avoir une estimation "*correcte*" de la variance pour un type de corrélation donnée, le tirage d'un nombre

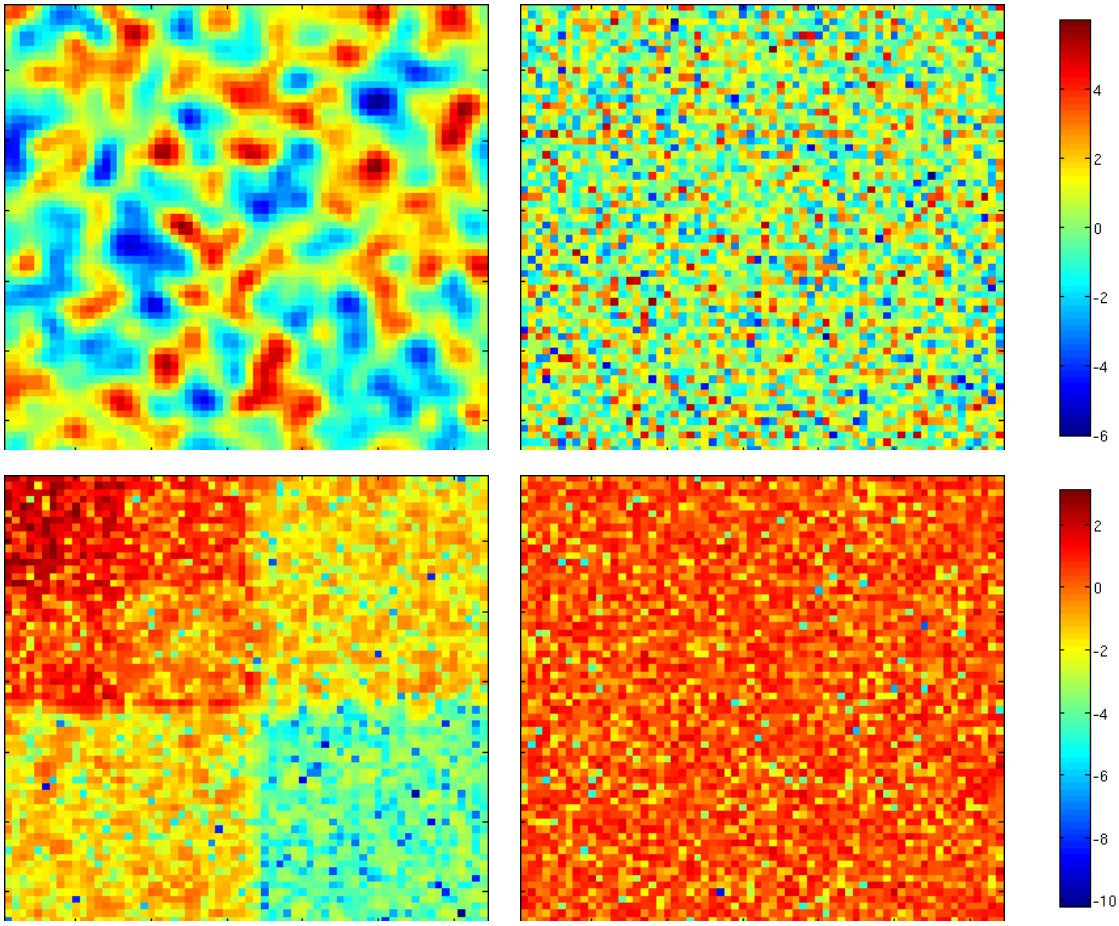


FIGURE 5.6 – L'image en haut à gauche présente un exemple de bruit corrélé en espace tandis qu'en haut à droite un bruit décorrélé en espace est présenté. En bas à gauche est présentée le logarithme (en base 10) de la valeur absolue des coefficients dans une base d'ondelettes ($\log_{10}|c_{j,k,l}|$) correspondant à la décomposition du bruit corrélé en espace. Le stockage des coefficients dans une base d'ondelettes est organisé comme décrit en figure 5.5. En bas à droite, la même chose est présentée pour le bruit indépendant en espace.

important de réalisation du bruit est nécessaire. Dans la suite, ce nombre de tirage est fixé à 1000.

Principe

Le principe consiste pour chacun des tirages à générer un bruit blanc gaussien $\boldsymbol{\epsilon} \sim \mathcal{N}(0, Id_n)$*, à le corréler en espace en appliquant* $\mathbf{R}_{pix}^{1/2}$ *de manière à obtenir* $\boldsymbol{\epsilon}_{cor} = \mathbf{R}_{pix}^{1/2}\boldsymbol{\epsilon}$ *puis à appliquer la transformée en ondelettes pour avoir accès à* $\boldsymbol{\epsilon}^{\mathbf{W}}(i) = \mathbf{W}\boldsymbol{\epsilon}_{cor}$*.*

Une fois ces opérations effectuées, il ne reste plus qu'à calculer une estimation de la variance de chacun des coefficients dans la base d'ondelettes en appliquant la formule suivante :

$$\mathrm{var}\,(c_{j,k,l}) = \frac{1}{m-1}\sum_{i=1}^{m}(\eta^{\mathbf{W}}_{j,\boldsymbol{k},l}(i))^2 - \frac{m}{m-1}\left(\sum_{i=1}^{m}\frac{\eta^{\mathbf{W}}_{j,\boldsymbol{k},l}(i)}{m}\right)^2. \tag{5.7}$$

où $m = 1000$ *est le nombre de tirage indépendant du bruit corrélé en espace.*

La figure 5.7 présente une estimation de la variance du bruit pour les ondelettes de Daubechies à 4 moments nuls pour les deux bruits présentés en figure 5.6. Afin de mieux distinguer la variance des coefficients à petite échelle, le *logarithme de la variance* est présenté.

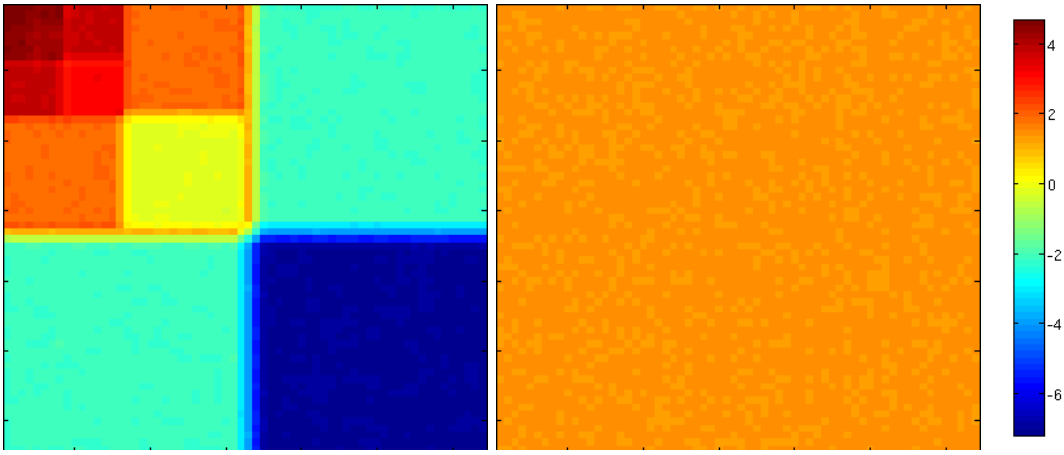


FIGURE 5.7 – Ces images présentent le logarithme de la variance estimés dans une base d'ondelettes. A gauche, la variance est estimeee pour un bruit corrélé en espace. Un tirage de ce bruit est présenté en figure 5.6. A droite, la variance est estimée pour un bruit décorrélé en espace.

Nous pouvons tout d'abord observer, en accord avec la définition de la matrice de covariance dans un espace d'ondelettes (4.18), que la variance dans une base d'ondelettes ne dépend ni de l'échelle, ni de l'orientation lorsque le bruit est décorrélé.

Au contraire, lorsque celui-ci est corrélé en espace, la variance (estimée) du bruit est fortement dépendante de l'échelle et de l'orientation. Il y a, pour cet exemple, un ratio d'environ **150 000** entre la plus grande variance (108.65) et la plus petite variance ($7.2\ 10^{-4}$) estimée pour le bruit corrélé en espace [14]. Cela signifie que certains coefficients, dans la base d'ondelettes, sont très peu perturbés par l'ajout d'un bruit corrélé (et lisse).

Remarque 5.5 (*Ondelettes périodique*)

Lors de l'estimation de la variance d'un bruit corrélé dans un espace d'ondelettes, on remarque que les coefficients de détails situés proche des bords ont une variance supérieure aux autres coefficients. Ceci est dû au fait que le bruit n'est pas périodique. Il y a ainsi une discontinuité au niveau des bords de l'image. Le bruit étant moins lisse dans ces zones, une partie plus importante est portée par les coefficients de détails.

Remarque 5.6 (*Autre interprétation*)

Un bruit corrélé en espace (lorsque le bruit est assez lisse) peut s'écrire de manière parcimonieuse dans des bases d'ondelettes. Le bruit considéré étant lisse, les coefficients de petite échelle portent peu d'informations sur le bruit. Par conséquent, lorsque les images bruitées sont utilisées, les coefficients de petite échelle, peu pertubés par le bruit, sont plus fiables (et apportent donc une information intéressante sur la position des discontinuités).

Ainsi, en "*approchant*", pour un bruit corrélé en espace, la matrice de covariance dans un espace d'ondelettes par la matrice définie en équation (5.6), on fait l'hypothèse que les deux champs de variances présentés en figure 5.7 sont identiques. De ce fait, on perd l'intégralité de l'information que l'on peut espérer glaner (et incorporer à moindre coût dans un processus d'assimilation) en travaillant dans un espace structuré.

Dans la section suivante, nous allons voir différentes manières d'obtenir de manière formelle la matrice de covariance dans une base d'ondelettes. La construction de cette matrice est ici basée sur la connaissance de la matrice de covariance dans l'espace des pixels.

14. Ce même ratio est de 1.36 pour le bruit décorrélé.

5.4 Construction formelle de la matrice de covariance dans une base d'ondelettes

Dans cette section, deux méthodes pour obtenir la matrice de covariance sont présentées. La première peut s'appliquer quelque soit le changement d'espace effectué. La seconde est quand à elle spécifique à la transformée en ondelettes. Pour cette seconde méthode de construction, nous aurons besoin d'utiliser la connaissance du principe de la décomposition et recomposition d'un signal dans une base d'ondelettes (introduit au chapitre 2).

5.4.1 Obtention de la matrice de covariance en mode boîte noire

Il est possible, sans aucune connaissance portant sur la transformée en ondelette, de construire la matrice de covariance dans une base d'ondelettes.
Pour cela, il suffit d'avoir accès aux opérateurs $\mathbf{W}$,$\mathbf{W}^T$ et de se rappeler que :

$$\mathbf{R_W} = \mathbf{W}\mathbf{R}_{Pix}\mathbf{W}^T. \tag{5.8}$$

La colonne i de la matrice de covariance peut-être obtenue en effectuant l'opération suivante :

$$\mathbf{R_W}(:, i) = \mathbf{R_W}\delta_i \tag{5.9}$$

où le vecteur δ_i est composé des éléments suivants :

$$\delta_i(x) = \begin{cases} 0 \text{ si } x \neq i \\ 1 \text{ sinon} \end{cases} \tag{5.10}$$

Cet algorithme est aisément implémentable, et très facilement adaptable à une autre transformée (sous réserve de connaître la transformée adjointe).
Il requiert en tout et pour tout (si n est le nombre d'élément dans l'espace des pixels) :

1. n transformée directe (d'un coût unitaire en $O(n)$),
2. n transformée adjointe (d'un coût unitaire en $O(n)$),
3. n multiplication matrice – vecteur (d'un coût unitaire en $O(n^2)$).

Remarque 5.7

Ce calcul est donc très gourmand en temps mais possède l'avantage d'être facilement parallélisable. De plus, si certaines propriétés de la matrice $\mathbf{R_W}$ *sont connues (voir section 5.4.4), le nombre d'opérations peut drastiquement diminuer (d'un facteur* n *environ).*

5.4.2 Utilisation de la connaissance de la transformée en ondelette

Une manière, certes plus complexe (mais plus rapide car évitant les multiplication matrice vecteur), pour obtenir la matrice de covariance dans une base d'ondelettes est d'effectuer le produit $\mathbf{W}\mathbf{R}_{pix}\mathbf{W}^T$ en utilisant les opérateurs $\mathbf{W}$ et $\mathbf{W}^T$. Ce travail peut être fait en construisant formellement les matrices $\mathbf{W}$ et $\mathbf{W}^T$ comme proposé dans [Azimifar et al., 2012] pour des signaux en 1–D. Ce produit peut aussi être effectué en appliquant successivement des opérateurs (basés sur une étape de transformée directe) à la matrice $\mathbf{R}_{pix}$. C'est cette façon de construire la matrice de covariance qui est présentée ici. Elle a été proposée dans [Vannucci and Corradi, 1999] pour des signaux 1–D et est étendue ici aux signaux en deux dimensions.

Mémo 8 – Notations dans des bases d'ondelettes

Dans cette section (et les suivantes), il est nécessaire de rentrer dans les détails de la décomposition en ondelettes.

— Le coefficient d'approximation à l'échelle j positionné en $\mathbf{k} = (k_1, k_2)$ est noté

$$c^j[k_1, k_2].$$

— L'ensemble des coefficents d'approximations $c^j[k_1, k_2]$ à une échelle j fixée forme la matrice

$$C^j.$$

— Le coefficient de détail vertical (resp. horizontal et diagonal) à l'échelle j positionné en (k_1, k_2) est noté

$$d_v^j[k_1, k_2] \quad (\text{resp. } d_h^j[k_1, k_2] \text{ et } d_d^j[k_1, k_2]).$$

— L'ensemble des coefficients de détails verticaux (resp. horizontaux et diagonaux) à une échelle j fixé forment la matrice

$$D_v^j \quad (\text{resp. } D_h^j \text{ et } D_d^j).$$

Le cadre analytique

Nous avons vu dans la section 2.2.2 que les coefficients d'approximations à l'échelle j peuvent être obtenus à partir des coefficients d'approximation à l'échelle $j+1$ via l'utilisation de deux applications linéaires H_c et H_r par :

$$C^j = H_c H_r C^{j+1} \tag{5.11}$$

où $H_r C^{j+1}$ correspond au résultat obtenu après le passage d'un filtre passe-bas appliqué à chacune des lignes de C^{j+1}, et H_c correspond à un filtre passe-bas appliqué à chacune des colonnes (de $H_r C^{j+1}$).

Ces applications font intervenir une décimation (i-e que l'information est stockée sur un nombre de coefficient plus petit) de telle sorte que lorsque C^{j+1} est de taille $2^n \times 2^n$ alors C^j est de taille $2^{n-1} \times 2^{n-1}$.

Nous avons aussi vu que les coefficients de détails (permettant de reconstruire les détails perdus lors de l'approximation) peuvent s'écrire de la manière suivante :

$$D_v^j = G_c H_r C^{j+1}, \tag{5.12}$$
$$D_h^j = H_c G_r C^{j+1}, \tag{5.13}$$
$$D_d^J = G_c G_r C^{j+1}. \tag{5.14}$$

La figure 5.5 rappelle la disposition usuelle de ces coefficients dans une base d'ondelettes. Le mémo 8 spécifie les notations employées pour les coefficients dans des bases d'ondelettes.

Notations employées pour décrire la matrice de covariance

Afin de décrire la formation et l'organisation de la matrice de covariance dans une base d'ondelettes, nous utilisons les notations introduites dans [Vannucci and Corradi, 1999].

La matrice de covariance d'erreur d'observation entre les coefficients d'approximations (à l'échelle $j+1$) est notée

$$CC^{\{j+1,j+1\}}. \tag{5.15}$$

Celle entre les coefficients de détails diagonaux (à l'échelle j) et les coefficients d'approximations (à l'échelle j) est notée :

$$CD_d^{\{j,j\}}. \tag{5.16}$$

Les covariances entre les coefficients d'approximation à l'échelle $j-k$ (avec $0 < k < j$) et les coefficients de détails horizontaux à l'échelle j sont stockées dans la matrice

$$CD_h^{\{j-k,j\}}. \tag{5.17}$$

Remarque 5.8

Alors que les matrices de covariances entre coefficients situés à la même échelle (par ex. $CC^{\{j,j\}}$ ou $D_hD_d^{\{j,j\}}$) sont des matrices carrées, la matrice faisant intervenir des coefficients situés à différentes échelles (par ex. $CD_h^{\{j-k,j\}}$) sont rectangulaires. Par exemple, la matrice $CD_h^{\{j-k,j\}}$ possède 2^k fois plus de colonnes que de lignes.

La matrice de covariance (dans une base d'ondelettes) est organisée sous la forme présentée en figure 5.8.

				$CD_v^{\{J-1,J\}}$	$CD_h^{\{J-1,J\}}$	$CD_d^{\{J-1,J\}}$
				$D_vD_v^{\{J-1,J\}}$	$D_vD_h^{\{J-1,J\}}$	$D_vD_d^{\{J-1,J\}}$
				$D_hD_v^{\{J-1,J\}}$	$D_hD_h^{\{J-1,J\}}$	$D_hD_d^{\{J-1,J\}}$
				$D_dD_v^{\{J-1,J\}}$	$D_dD_h^{\{J-1,J\}}$	$D_dD_d^{\{J-1,J\}}$
				$D_vD_v^{\{J,J\}}$	$D_vD_h^{\{J,J\}}$	$D_vD_d^{\{J,J\}}$
				$D_hD_v^{\{J,J\}}$	$D_hD_h^{\{J,J\}}$	$D_hD_d^{\{J,J\}}$
				$D_dD_v^{\{J,J\}}$	$D_dD_h^{\{J,J\}}$	$D_dD_d^{\{J,J\}}$

FIGURE 5.8 – Organisation de la matrice de covariance dans une base d'ondelettes.

Comment obtenir cette matrice (la théorie)

Intéressons nous maintenant à la manière d'obtenir les différentes parties de la matrice de covariance présentée en figure 5.8.

Nous supposons avoir accès à la connaissance de la matrice de covariance entre les coefficients d'approximation à l'échelle $j+1$ ($CC^{\{j+1,j+1\}}$).

Remarque 5.9

En pratique, cette connaissance signifie que pour débuter l'algorithme il faut connaître la matrice de covariance dans l'espace des pixels.

De manière analytique, les blocs composant $\mathbf{R_W}$ peuvent être obtenus en appliquant les filtres passe-bas (H) et passe-haut (G) employés dans la transformée en ondelettes.
Ainsi, la covariance entre les coefficients d'approximations à l'échelle j s'écrit

$$\operatorname{cov}\left(C^j, C^j\right) = H_c H_r \left[CC^{\{j+1,j+1\}}\right] H_r^T H_c^T \tag{5.18}$$

De manière analogue, la covariance entre les détails horizontaux et les coefficients d'approximation à l'échelle j se lit :

$$\begin{aligned} \operatorname{cov}\left(C^j, D_h^j\right) &= CD_h^{\{j,j\}} \\ &= \operatorname{cov}\left(H_c H_r C^{j+1}, G_c H_r C^{j+1}\right) \\ &= H_c H_r \left[CC^{\{j+1,j+1\}}\right] H_r^T G_c^T. \end{aligned} \tag{5.19}$$

Les covariances entre éléments situés à des échelles différentes (par exemple entre les coefficients d'approximations à l'échelle $j-1$ et les coefficients de détails horizontaux à l'échelle j) s'obtiennent par récurrence :

$$\begin{aligned} \operatorname{cov}\left(C^{j-1}, D_h^j\right) &= CD_h^{\{j-1,j\}} \\ &= \operatorname{cov}\left(H_c H_r C^j, D_h^j\right) \\ &= H_c H_r \left[CD_h^{\{j,j\}}\right] \\ \Big(&= H_c H_r H_c H_r \left[CC^{\{j+1,j+1\}}\right] H_r^T G_c^T \Big) \end{aligned} \tag{5.20}$$

Ainsi, de manière récursive, il est possible d'obtenir la matrice de covariance dans une base d'ondelettes via l'application de filtres à partir de la connaissance de la matrice de covariance dans l'espace des pixels.

L'algorithme 1 résume les différentes grandes étapes nécessaires à l'obtention de la matrice de covariance dans une base d'ondelettes.

Du cadre analytique au cadre algorithmique

Bien que la partie précédente nous fournisse les grandes étapes nécessaires pour obtenir la matrice de covariance dans l'espace des ondelettes, il faut rappeler que les applications H_r, H_c, G_r et G_c ne sont bien connues que sous forme d'opérateurs (et non sous forme matricielle). Ne cherchant pas à faire intervenir de coûteux produits matriciels, il convient dès lors de savoir comment appliquer ces opérateurs à la matrice de covariance dans l'espace des pixels.

Comment allons nous procéder ?

Nous allons tout d'abord rappeler les formules permettant de calculer la covariance entre deux éléments donnés (pour une même échelle).
Nous verrons ensuite comment appliquer nos opérateurs (i–e les étapes de transformées en ondelettes 1–D ou 2–D) à la matrice $CC^{\{j+1,j+1\}}$ pour former les matrices de covariances entre éléments de l'échelle j.

Algorithm 1 Grandes étapes de l'algorithme utilisé pour obtenir la matrice de covariance dans une base d'ondelettes

Initialisation :
$CC^{\{j+1,j+1\}} \longleftarrow \mathbf{R}_{pix}$
A partir de $CC^{\{j+1,j+1\}}$ construire les 10 matrices $CC^{\{j,j\}}$, $D_hD_h^{\{j,j\}}$, $D_hD_v^{\{j,j\}}, \cdots$

for i=j,k (avec k le nombre d'échelles de la transformée en ondelette) **do**

Former à partir de $CC^{\{i,i\}}$ les 10 matrices $CC^{\{i-1,i-1\}}$, $CD_h^{\{i-1,i-1\}}, \cdots$

for l=i,j **do**

Former à partir de la matrice $CD_h^{\{i,l\}}$ les matrices $CD_h^{\{i-1,l\}}$, $D_hD_h^{\{i-1,l\}}$, $D_vD_h^{\{i-1,l\}}$, $D_vD_h^{\{i-1,l\}}$
Faire de même à partir des es matrices $CD_v^{\{i,l\}}$ et $CD_d^{\{i,l\}}$

end for
end for

Commençons par rappeler les formules utilisées pour obtenir les coefficients d'approximation et de détails à l'échelle j par rapport aux coefficients d'approximation à l'échelle $j+1$ dans une base d'ondelettes séparable :

$$c^j[k_1,k_2] = \sum_{n,m\in\mathbb{Z}} h[n-2k_1]h[m-2k_2]c^{j+1}[n,m] \tag{5.21}$$

$$d_v^j[k_1,k_2] = \sum_{n,m\in\mathbb{Z}} h[n-2k_1]g[m-2k_2]c^{j+1}[n,m] \tag{5.22}$$

$$d_h^j[k_1,k_2] = \sum_{n,m\in\mathbb{Z}} g[n-2k_1]h[m-2k_2]c^{j+1}[n,m] \tag{5.23}$$

$$d_d^j[k_1,k_2] = \sum_{n,m\in\mathbb{Z}} g[n-2k_1]g[m-2k_2]c^{j+1}[n,m]. \tag{5.24}$$

La covariance entre deux éléments d'approximation à la même échelle est :

$$cov(c^j[k_1,k_2],c^j[k_3,k_4]) = \sum_{n,m,o,p\in\mathbb{Z}} h[n-2k_1]h[m-2k_2]h[o-2k_3]h[p-2k_4]cov(c^{j+1}[n,m],c^{j+1}[o,p]), \tag{5.25}$$

tandis que celle entre coefficients de détails (horizontaux) et d'approximation s'écrit :

$$cov(c^j[k_1,k_2],d_h^j[k_3,k_4]) = \sum_{n,m,o,p\in\mathbb{Z}} h[n-2k_1]h[m-2k_2]g[o-2k_3]h[p-2k_4]cov(c^{j+1}[n,m],c^{j+1}[o,p]). \tag{5.26}$$

Nous devons donc effectuer une *quadruple convolution* pour obtenir, à partir des covariances à l'échelle j, les covariances à l'échelle $j+1$. Cette quadruple convolution s'apparente à une étape de la transformée en ondelettes séparable en 4–D.

Convolution sur m et p

Avant d'effectuer les premières phases de convolution, nous devons stocker la matrice de covariance sous la forme présentée en figure 5.9. Cette matrice est composée de blocs L_q^i où chaque bloc correspond à la matrice de covariance entre les éléments de la ligne i et les éléments de la ligne q.

L_1^1	L_2^1	$\cdots$	$L_{n_{max}}^1$
L_1^2	L_2^2	$\cdots$	$L_{n_{max}}^2$
$\vdots$	$\vdots$	$\ddots$	$\vdots$
$L_1^{n_{max}}$	$L_2^{n_{max}}$	$\cdots$	$L_{n_{max}}^{n_{max}}$

FIGURE 5.9 – Matrice de covariance organisée de telle sorte que chaque bloc L_j^i contienne les covariances entre les éléments des lignes i et j. Le nombre de ligne de l'image est noté n_{max}.

Afin d'effectuer les phases de convolutions sur m et p, il faut appliquer une étape de la transformée en ondelettes 2–D sur chaque bloc de cette matrice. Ainsi, chaque bloc L_q^i est transformé sous la forme suivante :

$$\begin{pmatrix} H_c H_r L_q^i & G_c H_r L_q^i \\ H_c G_r L_q^i & G_c G_r L_q^i \end{pmatrix} \tag{5.27}$$

Remarque 5.10

Un élément du bloc $H_c H_r L_q^i$ est de la forme

$$r_1[i, k_2, q, k_4] = \sum_{m,p \in \mathbb{Z}} h[m - 2k_2] h[p - 2k_4] cov(c^{j+1}[i, m], c^{j+1}[q, p]) \tag{5.28}$$

où k_2 et k_4 dépendent de sa position dans la matrice.

De même, un élément du bloc $G_c H_r L_q^i$ s'écrit :

$$r_3[i, k_2, q, k_4] = \sum_{m,p \in \mathbb{Z}} h[m - 2k_2] g[p - 2k_4] cov(c^{j+1}[i, m], c^{j+1}[q, p]) \tag{5.29}$$

Une fois ces étapes effectuées, il ne reste plus qu'à effectuer le produit de convolution sur n et o pour obtenir les coefficients de covariances à l'échelle j.

Réorganisation de la matrice Avant d'appliquer la seconde partie des opérateurs il nous faut cependant réorganiser la matrice sous la forme présentée en figure 5.10.
Pour cela, nous devons regrouper les éléments du type r_1 (présentés en équation (5.28)) dans le coin haut à droite.

Remarque 5.11

Plus précisément, nous devons réorganiser la matrice de covariance de la sorte à former les blocs $C1_{k_2}^{k_4}$ comprenant les coefficients du type $r_1[*, k_2, *, k_4]$. Le même processus doit être appliqué aux 3 autres types de coefficients afin d'avoir accès à une matrice de covariances aisément utilisable par la suite.

$H_r \left[CC^{\{j+1,j+1\}}\right] H_r^T$	$G_r \left[CC^{\{j+1,j+1\}}\right] H_r^T$
$H_r \left[CC^{\{j+1,j+1\}}\right] G_r^T$	$G_r \left[CC^{\{j+1,j+1\}}\right] G_r^T$

FIGURE 5.10 – Organisation de la matrice souhaitée

Convolution sur n et o

A partir de cette matrice, on effectue les mêmes opérations que précédemement (à savoir une transformée 2–D sur chacun des blocs). Lorsque cette étape est réalisée, nous obtenons bien la covariance entre les coefficients de l'échelle j (voir équation (5.25) et (5.26)). Nous devons cependant procéder à une réorganisation (au sein de chaque quart de la matrice obtenue) afin d'avoir la matrice disposée sous la forme présentée en figure 5.11.

$CC^{\{J,J\}}$	$CD_v^{\{J,J\}}$	$CD_h^{\{J,J\}}$	$CD_d^{\{J,J\}}$
$D_vC^{\{J,J\}}$	$D_vD_v^{\{J,J\}}$	$D_vD_h^{\{J,J\}}$	$D_vD_d^{\{J,J\}}$
$D_hC^{\{J,J\}}$	$D_hD_v^{\{J,J\}}$	$D_hD_h^{\{J,J\}}$	$D_hD_d^{\{J,J\}}$
$D_dC^{\{J,J\}}$	$D_dD_v^{\{J,J\}}$	$D_dD_h^{\{J,J\}}$	$D_dD_d^{\{J,J\}}$

FIGURE 5.11 – Organisation de la matrice de covariance dans une base d'ondelettes.

Bilan 5.1

On a ainsi pu voir comment, à partir de la connaissance de $CC^{\{j+1,j+1\}}$ via l'application de transformées en ondelette 2D, nous pouvions former les différents blocs présentés en figure 5.11. Hors opération de réorganisation, nous avons effectué $n_{max}^2 + m_{max}^2$ étapes de la transformée en ondelette 2D où n_{max} et m_{max} sont le nombre de lignes et de colonnes de l'image originale. Cependant le travail n'est pas encore terminé pour obtenir la matrice

complète présentée en figure 5.8.

Remarque 5.12

La matrice formée étant symétrique (par définition), un grand nombre de transformées en ondelettes peuvent être évitées (un peu moins de la moitié).

Covariances entre coefficients d'échelles différentes

Pour obtenir les blocs de la matrice de covariance entre coefficients d'échelles différentes, il faut effectuer *pratiquement* les mêmes opérations. Les phases décrites aux paragraphes précédents sont identiques à l'exception de la transformée utilisée : celle-ci doit être unidimensionnelle.

Obtention de la matrice complète

Pour obtenir la matrice après plusieurs niveaux de décomposition, il faut réeffectuer ces opérations plusieurs fois (sur des matrices dont le nombre d'éléments est divisé par 16 (ou 4 pour les blocs inter-échelles)). L'algorithme 1 en résume les principales étapes.

Remarque 5.13

Obtenir la matrice de covariance dans une base d'ondelettes est bien plus compliqué par ce biais que par la méthode présentée en section 5.4.1. Cependant, hors phase de réorganisation de la matrice, le nombre d'opération est drastiquement réduit : il correspond uniquement aux étapes de transformées directes et inverses de l'autre méthode. On évite ainsi la partie la plus coûteuse de l'algorithme (qui consiste en la multiplication matrice vecteur) au prix de réorganisations successives de la matrice. En pratique, bien que ce soient ces phases de réorganisations qui demandent (à l'heure actuelle) le plus de temps de calcul, cette dernière méthode est nettement plus rapide dès lors que la matrice est de grande taille.

Remarque 5.14

Les matrices de covariances utilisées sont souvent creuses dans l'espace des pixels. Deux lignes ou colonnes de pixels éloignés spatialement peuvent donc ne pas comporter de pixels corrélés entre eux (ce qui signifie que $L_j^i = 0$). En utilisant l'algorithme précédent, cela peut permettre d'économiser un grand nombre de transformées en ondelettes réduisant d'autant plus le nombre d'opérations à effectuer (et le temps de réorganisation).

Bilan 5.2

Nous avons donc vu en section 5.4.1 et en section 5.4.2 comment former la matrice de covariance d'erreur d'observation dans une base d'ondelettes à partir de sa connaissance dans l'espace des pixels.

Remarque 5.15

Bien qu'ayant dans certains cas de bonnes propriétés, ces matrices sont de trop grande taille pour pouvoir se permettre leur inversion. En section, 5.5 nous allons cependant voir le gain d'information que peut apporter l'utilisation de matrice diagonale dans un espace d'ondelettes. Mais avant ceci présentons ce que nous avons obtenu en fin de l'algorithme.

5.4.3 Mais au final, à quoi ça ressemble ?

Après avoir lu en détail (voir même en diagonale) la section précédente, cette question est on ne peut plus légitime. La figure 5.12 présente, sur un exemple, la matrice de corrélation obtenue dans une base d'ondelettes (à droite) à partir celle dans l'espace des pixels (à gauche).
Comme on peut le voir sur cet exemple, l'hypothèse d'une matrice de covariance diagonale dans une base d'ondelettes est loin d'être respectée. En réalité cette matrice comporte plus d'éléments que son équivalent dans l'espace des pixels (894 352 contre 222 784).

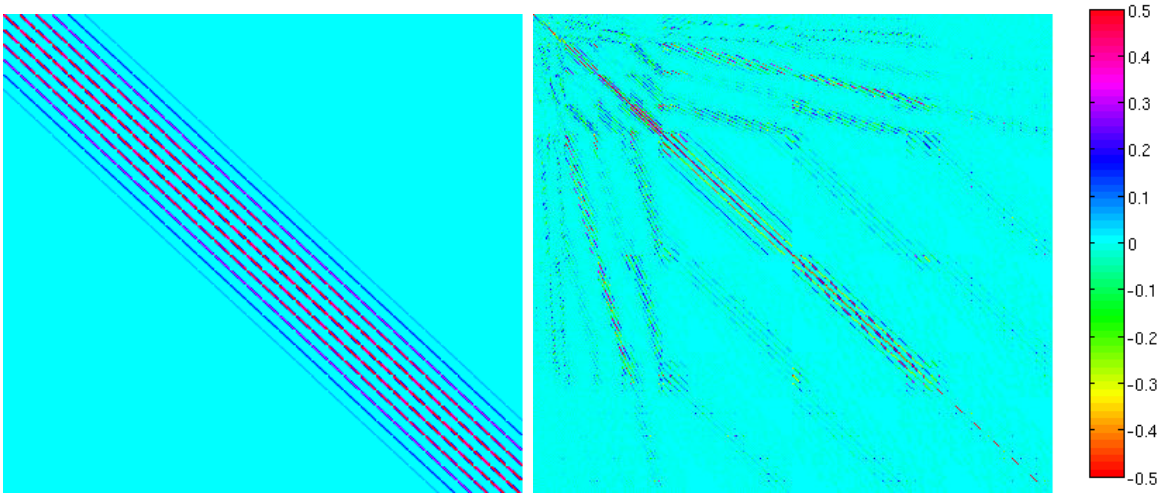


FIGURE 5.12 – Matrice de corrélation dans l'espace des pixels (à gauche) et dans la base des ondelettes de Daubechies à 4 moments nuls à droite. Cette dernière est stockée de la manière présentée en figure 5.8. De manière à mieux visualiser les corrélations existantes, l'échelle de couleur est comprise entre $[-0.5; 0.5]$ dans les deux cas.
La taille de l'image originale est de 32×32 pixels. Par conséquent, ces matrices comportent 1 048 576 éléments. Parmis ces éléments dans l'espace des pixels 222 784 ne sont pas nuls, contre 894 352 pour les ondelettes.

Cependant, il y a peu de coefficients de corrélations importants dans cet espace. En effet, lorsqu'on regarde le nombre de coefficients de corrélation dont la valeur absolue est supérieure à un certain seuil, cette proportion s'inverse rapidement. La figure 5.13 présente le nombre d'éléments de la matrice de corrélation dont la valeur est supérieure à un certain seuil dans l'espace des pixels (en bleu) et dans une base d'ondelettes (en rouge). L'axe des ordonnées est présenté en échelle logarithmique. On observe une décroissance rapide de ce nombre en fonction du seuil choisi dans les deux cas.

En observant cette figure, on s'aperçoit qu'un grand nombre d'éléments extra–diagonaux de la matrice de corrélation dans une base d'ondelettes ne sont pas significatifs et donc que l'information est plus concentrée sur la diagonale. Par exemple, environ 2500 coefficients de corrélations (en comptant les 1024 éléments diagonaux) ont une norme $L1$ supérieur à 0.5 dans une base d'ondelettes contre 20 000 dans l'espace des pixels.

Cette observation se vérifie lors du calcul du pourcentage d'information (en norme L_1) portée par la diagonale de la matrice de corrélation à savoir :

$$\frac{\sum_i |C(i,i)|}{\sum_{i,j} |C(i,j)|}.$$

Pour les ondelettes, ce pourcentage est de 8.7% tandis qu'il est de 4% pour les pixels. Par conséquent, le pourcentage d'information porté par la diagonale de la matrice de corrélation est deux fois plus important dans le cas des ondelettes que dans le cas des pixels.

5.4.4 Propriété facilitant la construction lorsque le bruit est identiquement distribué

Avant de passer aux bonnes propriétés des matrices diagonales dans des bases d'ondelettes, intéressons–nous à une propriété facilitant grandement la construction de matrices de covariances dans ces bases dans certains cas.

Lorsque le bruit est identiquement distribué dans l'espace des pixels, il est en effet *peu coûteux* (par rapport au cas précédent) de construire la matrice de covariance dans une base d'ondelettes.

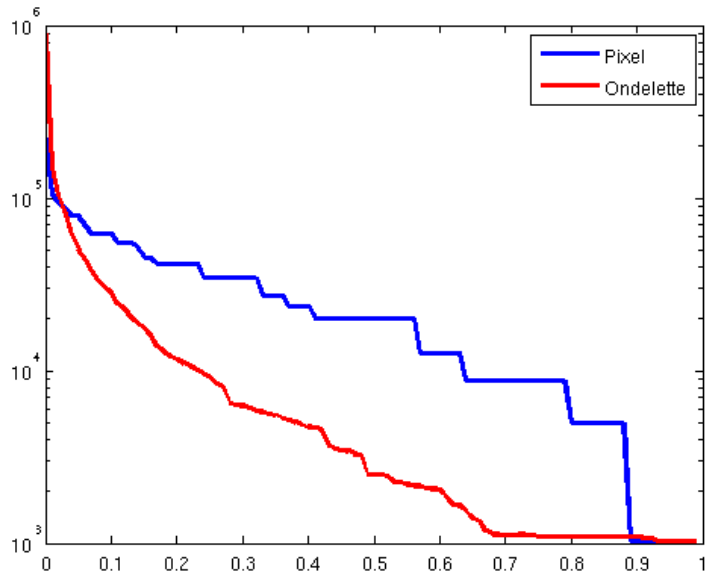


FIGURE 5.13 – Nombre de coefficients de la matrice de corrélation dont la valeur absolue est supérieure à un certain seuil dans l'espace des pixels (en bleu) et dans une base d'ondelettes (en rouge). Cet exemple correspond aux matrices présentées en figure 5.12.

Afin de simplifier les calculs (et les notations) le théorème suivant est proposé en $1-D$ (mais reste vrai en plusieurs dimensions).

Propriété 5.1 (*Conservation de l'invariance par translation*)

Lors de l'utilisation d'ondelettes périodiques, si les termes de covariances entre les coefficents d'approximations à l'échelle $j+1$ sont invariants par translation (et périodiques), i-e que :

$$\operatorname{cov}\left(c^{j+1}[k], c^{j+1}[r]\right) = \operatorname{cov}\left(c^{j+1}[k+p], c^{j+1}[r+p]\right) \quad \forall k\, r \in \mathbb{N}\ \forall p \in \mathbb{Z}. \tag{5.30}$$

alors cette propriété est conservée entre les termes $c^j \in V^j$ et $d^j \in W^j$, autrement dit :

$$\operatorname{cov}\left(c^j[k_1], c^j[k_2]\right) = \operatorname{cov}\left(c^j[k_1+p], c^j[k_2+p]\right) \quad \forall k_1\, k_2 \in \mathbb{N}\ \forall p \in \mathbb{Z}. \tag{5.31}$$

$$\operatorname{cov}\left(d^j[k_1], d^j[k_2]\right) = \operatorname{cov}\left(d^j[k_1+p], d^j[k_2+p]\right) \quad \forall k_1\, k_2 \in \mathbb{N}\ \forall p \in \mathbb{Z}. \tag{5.32}$$

$$\operatorname{cov}\left(c^j[k_1], d^j[k_2]\right) = \operatorname{cov}\left(c^j[k_1+p], d^j[k_2+p]\right) \quad \forall k_1, k_2 \in \mathbb{N}\ \forall p \in \mathbb{Z}. \tag{5.33}$$

Démonstration : Par définition de la transformée en ondelettes, on a :

$$\begin{aligned}\operatorname{cov}\left(c^j[k_1], c^j[k_2]\right) &= \operatorname{cov}\left(\sum_n h[n-2k_1]c^{j+1}[n], \sum_m h[m-2k_2]c^{j+1}[m]\right) \\ &= \sum_{n,m} h[n-2k_1]h[m-2k_2]\operatorname{cov}\left(c^{j+1}[n], c^{j+1}[m]\right)\end{aligned} \tag{5.34}$$

En posant $n' = n+2p$ et $m' = m+2p$ dans l'équation (5.34), on obtient :

$$\operatorname{cov}\left(c^j[k_1], c^j[k_2]\right) = \sum_{n,m} h[n'-2(k_1+p)]h[m'-2(k_2+p)]\operatorname{cov}\left(c^{j+1}[n'-2p], c^{j+1}[m'-2p]\right) \tag{5.35}$$

En utilisant le fait que les corrélations sont invariantes par translations, i-e,

$$\operatorname{cov}\left(c^{j+1}[n], c^{j+1}[n]\right) = \operatorname{cov}\left(c^{j+1}[n-2p], c^{j+1}[n-2p]\right)$$

on obtient que :

$$\begin{aligned}\operatorname{cov}\left(c^{j}[k_1], c^{j}[k_2]\right) &= \sum_{n',m'} h[n'-2(k_1+p)]h[m'-2(k_2+p)] \operatorname{cov}\left(c^{j+1}[n'], c^{j+1}[m']\right) \\ &= \operatorname{cov}\left(c^{j}[k_1+p], c^{j}[k_2+p]\right) \qquad (5.36)\end{aligned}$$

Les mêmes étapes de calcul permettent d'obtenir les résultats des équations (5.32) et (5.33). ■

Remarque 5.16

On remarque que nous ne nous sommes pas servis de la connaissance des filtres h et g dans cette démonstration. Ainsi ce résultat reste valable quelques soient les filtres choisis. Par conséquent, lors d'un autre changement d'espace, si le calcul des coefficients d'approximation (et de détails) est effectué en tout point k de l'espace en employant les même filtres cette propriété est valable. C'est donc aussi le cas (par exemple) lors de l'utilisation de la transformée en curvelettes.

Cette propriété, qui reste valable en deux dimensions, est particulièrement intéressante pour construire la matrice de covariance. Par exemple, lorsqu'on utilise la méthode de type *boîte noire* présentée en section 5.4.1, il suffit de connaître la colonne de la matrice R correspondant à un élément (caractérisé par son échelle et son orientation) pour connaître l'ensemble de la matrice à cette échelle et cette orientation (à une permutation près).

A titre d'exemple, il suffit de 10 applications de (5.9) afin de connaître l'ensemble des coefficients composant la matrice $\mathbf{R_W}$ lorsqu'on effectue une décomposition en ondelettes bidimensionnelles sur 3 niveaux.

Remarque 5.17

Cette propriété est aussi intéressante combinée avec la méthode décrite en section 5.4.1. Dans ce cas là, l'ensemble de l'information à utiliser lors des transformations est contenu dans les blocs $\left\{L_1^j\right\}_{j=1,\frac{n_{max}}{2}}$ (car les covariances sont périodiques). Le nombre de transformées en ondelettes à effectuer est donc drastiquement réduit (ainsi que le nombre de coefficients à stocker).

Au vu de la remarque introductive du chapitre 4 (appuyée par l'équation (4.8)), on peut "profiter" de la propriété 5.1 dès lors que le modèle de corrélation (dans l'espace des pixels) est invariant par translation. Cependant, lorsque le bruit n'est pas périodique, le modèle de corrélation n'est pas invariant par translation. Par conséquent, cette bonne propriété n'est valable que pour les coefficients de détails et d'approximations éloignés des bords. Ce phénomène est par exemple visible en figure 5.7.

Remarque 5.18

Les fonctions d'ondelettes et d'échelle dans la base de Haar n'ayant pas un support périodique vérifient donc cette propriété, et ce, même si le bruit n'est pas périodique. C'est la seule ondelette vérifiant cette propriété lorsque le bruit n'est pas périodique.

5.5 Approximation diagonale dans une base d'ondelettes

5.5.1 Impact de la diagonalisation

Si nous souhaitions (et surtout pouvions) utiliser l'ensemble de l'information contenue dans la matrice de covariance dans une base d'ondelettes, nous aurions exactement la même information

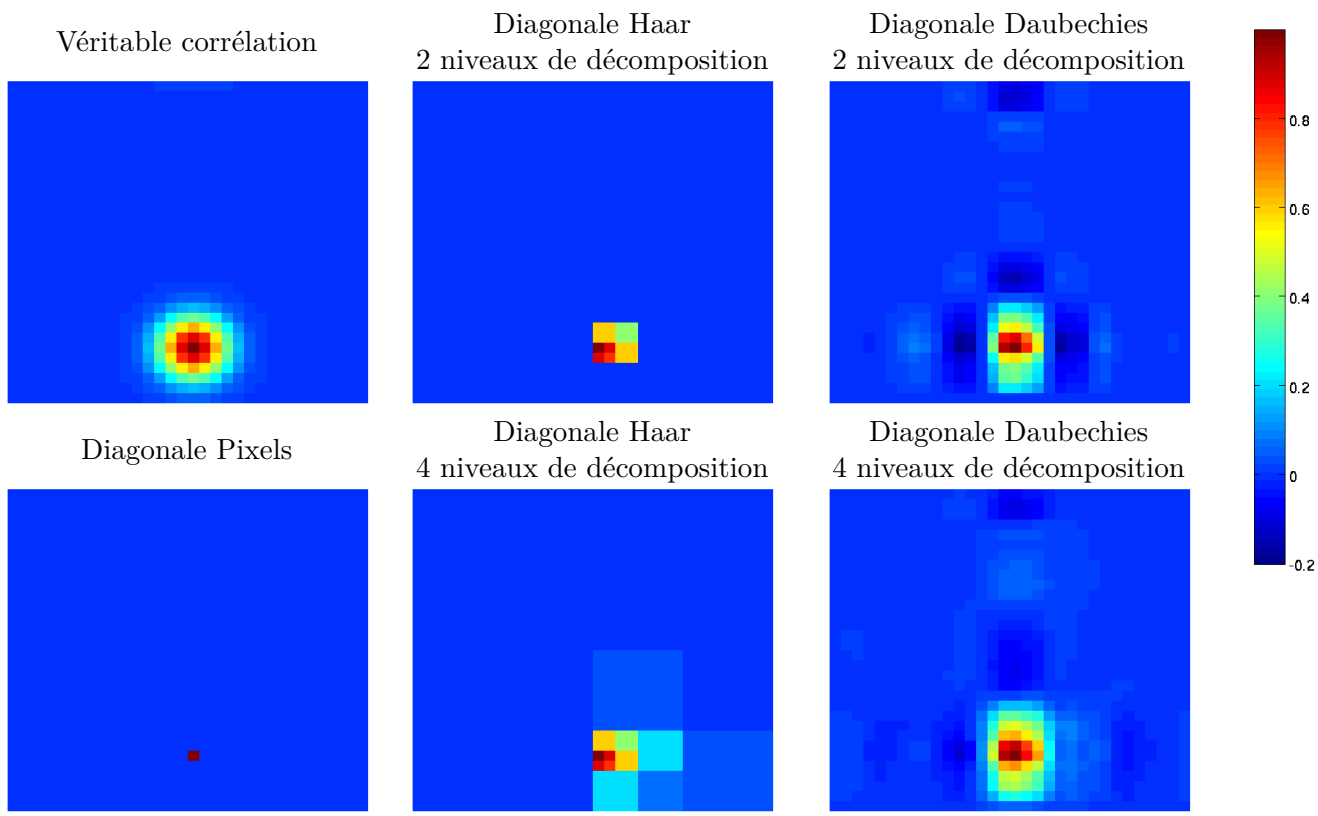


FIGURE 5.14 – Les corrélations entre un pixel et ses voisins sont présentées pour une image de taille 32×32. Le véritable modèle de corrélation est présenté en haut à gauche. Les autres figures présentent l'effet d'une approximation diagonale dans différents espaces.

que dans l'espace des pixels (mais exprimée dans une autre base). Dans la suite seules des approximations diagonales des matrices de covariances sont utilisées. Ces approximations sont notées

$$\tilde{\mathbf{R}}^{\text{cor}}_{\mathbf{W},\text{diag}} = \text{diag}(\mathbf{W}\mathbf{R}_{pix}\mathbf{W}^T). \tag{5.37}$$

Dans cette partie, nous allons illustrer les gains potentiels dus à l'utilisation d'une matrice diagonale dans une base d'ondelettes (par rapport à l'utilisation d'une matrice diagonale dans l'espace des pixels). Pour cela, nous allons nous intéresser à la matrice équivalente, dans l'espace des pixels, à la matrice diagonale dans une base d'ondelettes définie par :

$$\mathbf{R}_{pixEq} = \mathbf{W}^T\tilde{\mathbf{R}}^{\text{cor}}_{\mathbf{W},\text{diag}}\mathbf{W}. \tag{5.38}$$

Remarque 5.19

Quand aucune approximation n'est effectuée dans une base d'ondelettes, la matrice $\mathbf{R}_{pixEq} = \mathbf{R}_{pix}$. *Lorsqu'une approximation diagonale est réalisée, en comparant ces matrices, nous pouvons donc avoir une idée des covariances qu'il est possible de représenter.*

La figure 5.14 présente les corrélations d'un pixel avec ses voisins pour une approximation diagonale dans différents espaces. Contrairement à ce qu'il se passe dans l'espace des pixels, lorsqu'on travaille dans un espace d'ondelettes, certaines corrélations sont prises en compte. Ces corrélations dépendent du choix de l'ondelette et du niveau de décomposition utilisé. Cependant aucune des deux ondelettes choisies ne permet de représenter de manière correcte l'isotropie du bruit. Ce phénomène est d'autant plus marqué avec les ondelettes de Haar.

L'utilisation d'une matrice de covariance diagonale dans une base d'ondelettes produit aussi des corrélations entre des pixels non corrélés à l'origine, corrélations qui peuvent être négatives.

Bilan 5.3

Malgré ces défauts, l'approximation obtenue est tout de même de qualité correcte : il y a une forte ressemblance entre les vraies corrélations et celles induites par l'utilisation d'une matrice diagonale (avec les ondelettes de Daubechies).

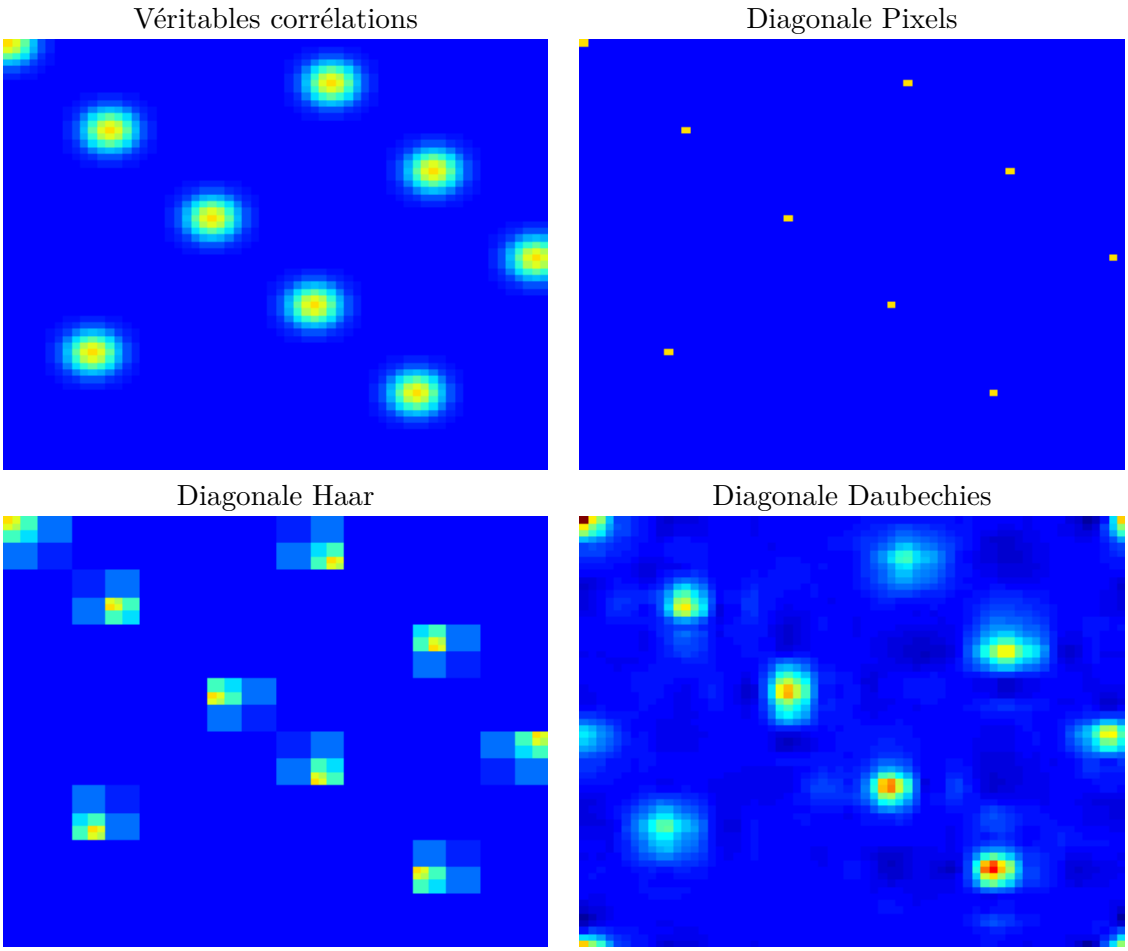


FIGURE 5.15 – En haut à gauche on présente la vraie covariance entre l'erreur en un pixel et l'erreur chez ses voisins pour 10 pixels différents répartis au sein de notre image. Les autres figures présentent l'effet d'une approximation diagonale dans différents espaces.

La figure 5.15 présente les covariances obtenues en 9 pixels situés à différents endroits de l'image (pour une image en comportant 64×64). On s'aperçoit que :

— l'approximation diagonale dans une base d'ondelettes fait fortement varier le modèle de corrélation d'un pixel à un autre.

— la variance obtenue dans l'espace des pixels, pour les ondelettes de Daubechies, dépend de la localisation.

— Le pixel en haut à gauche de l'image est corrélé avec les pixels situés dans les autres coins de l'image.

Le dernier point provient du fait que l'ondelette choisie est périodique. Lorsque l'image n'est pas périodique il faut donc faire attention à l'influence potentielle que peuvent avoir ces corrélations parasites.

Remarque 5.20

Les ondelettes de Haar sont les seules ondelettes orthonormales non périodiques possédant

en tout point de l'espace la même fonction de base et ne sont donc pas affectées par ce problème.

Remarque 5.21

Ce problème peut être évité par l'emploi d'ondelettes de bords à la place des ondelettes périodiques. Pour ce type d'ondelettes, les filtres associés aux actions de décomposition et de recomposition varient en fonction de la localisation sur l'image de telle sorte à ne jamais utiliser des pixels situés de l'autre côté. Voir par exemple [Mallat, 2008] pour plus de précisions sur ce type d'ondelettes. Cependant, en utilisant ce type d'ondelettes, la propriété 5.1 permettant de construire à moindre coût la matrice de covariance lorsque le bruit est invariant par translation n'est plus respectée : le filtre change près des bords de l'image.

5.5.2 Renormalisation

En figure 5.15, nous avons observé que la variance induite par l'utilisation d'une matrice de covariance diagonale dans l'espace des pixels varie de point en point. Afin d'illustrer le fait que la variance change d'un point à l'autre, nous allons nous intéresser à la variance obtenue dans l'espace des pixels lorsque la matrice de covariance est considérée diagonale dans les bases d'ondelettes de Haar, de Daubechies à 2 moments nuls et de Daubechies à 4 moments nuls. La figure 5.16 présente les résultats obtenus (pour un même modèle de corrélation) dans ces différentes bases lorsque la décomposition en ondelettes s'effectue sur un niveau (première ligne) et sur trois niveaux (seconde ligne).

On observe tout d'abord que ce problème n'apparait pas lors de l'utilisation des ondelettes de Haar. Ces dernières permettent d'avoir une variance invariante par translation dans l'espace des pixels lorsqu'un seul niveau de décomposition est appliqué. Ce n'est pas le cas pour les ondelettes de Daubechies à 2 et 4 moments nuls.

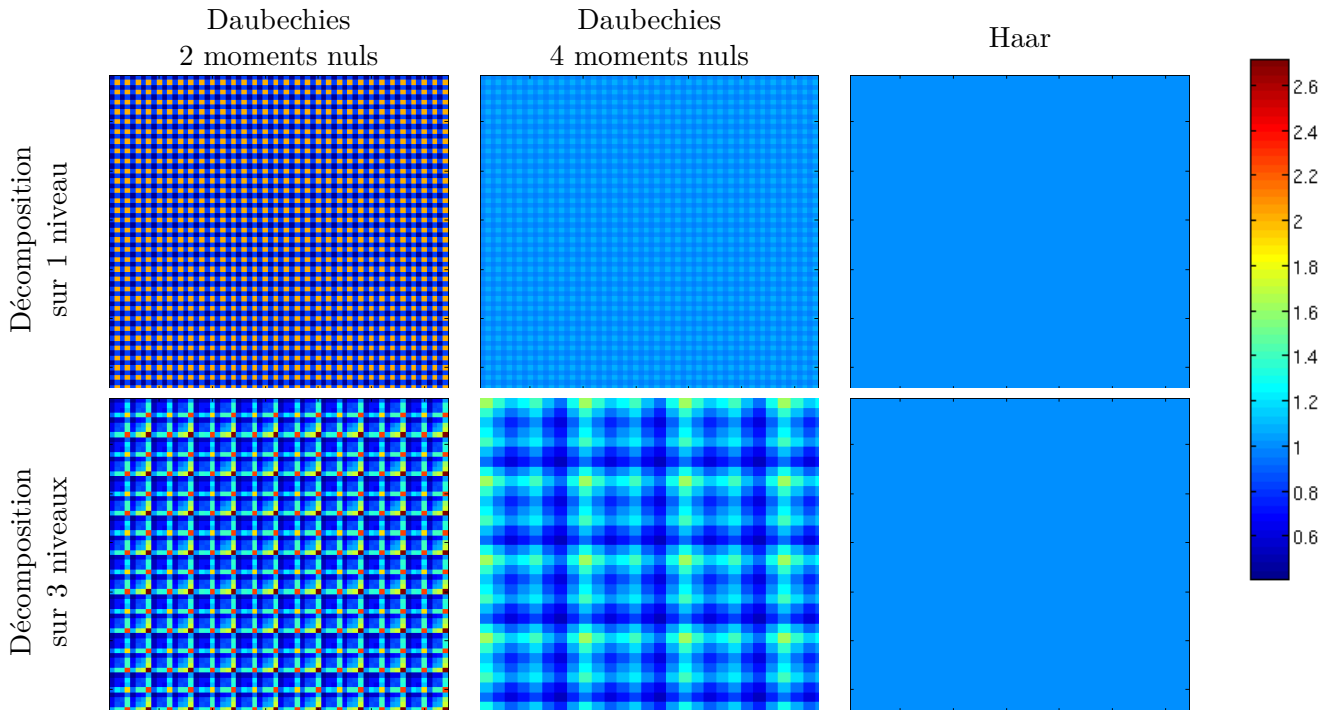


FIGURE 5.16 – Les différentes figures montrent la variance obtenue dans l'espace des pixels lorsque seule la diagonale de la matrice dans l'espace des ondelettes est considérée.

Pour caractériser la dispersion de la variance, nous allons calculer le ratio entre la plus grande variance induite par l'utilisation d'une matrice diagonale et la plus petite (dans une base donnée). Etant donné que seules des matrices de corrélations sont "approchées", ce ratio correspond à la différence de confiance accordée à des pixels (dans lesquels on devrait avoir la même confiance).

Pour les ondelettes de Daubechies à 4 moments nuls, ce ratio est relativement peu élevé (1.1 pour 1 niveau et 2.6 pour 3 niveaux de décomposition). Cependant pour les ondelettes de daubechies à 2 moments nuls, ce ratio commence à devenir important : dans l'exemple présenté en figure 5.16 il est de 5.7 pour 1 niveau de décomposition et de 6.6 pour 3 niveaux. Ainsi, en utilisant cette ondelette et la diagonale de la matrice de covariance, on aurait (artificiellement) presque 7 fois plus confiance dans certains pixels que dans d'autres. Ce ratio a tendance à augmenter avec le nombre de niveau de décomposition utilisé dans la base d'ondelettes.

Une méthode pour résoudre ce problème, si on considére qu'il n'y a pas de raisons valables[15] à ce changement de confiance, consiste à effectuer une renormalisation de telle sorte que la matrice

$$\mathbf{R}_{pixEq} = \Lambda \mathbf{W}^T \mathrm{diag}(\mathbf{W}\mathbf{R}_{pix}\mathbf{W}^T)\mathbf{W}\Lambda \tag{5.39}$$

soit une matrice de corrélation (contrairement à la définition introduite en équation (5.38)). Lors du calcul de la fonction coût, on effectue donc une "renormalisation" de l'influence de chaque pixels avant d'effectuer la transformée en ondelettes.

5.5.3 Propriété intéressante pour la renormalisation (et autres)

Dans l'exemple présenté en figure 5.16, on observe un motif qui se répète dans chacun des cas. Ce motif est de taille 2×2 pixels lorsque la décomposition est effectuée sur 1 niveau et de taille 8×8 pixels lorsque la décomposition est effectuée sur 3 niveaux.
Cette répétition de motif est intéressante quand on souhaite renormaliser : en calculant les coefficients de renormalisation en quelques pixels (via par exemple la méthode présentée en section 5.4.1) on peut obtenir l'ensemble des coefficients de renormalisation.

Comme dans la partie 5.4.4, la propriété suivante est présentée en 1–D (afin notamment d'utiliser des notations plus "légères") mais reste valable en plusieurs dimensions.

Propriété 5.2

Si les covariances entre coefficients d'approximations et de détails à une échelle j donnée sont (en quelque sorte) invariant par translation de 2^{l-1} pixels, i-e :

$$\begin{aligned}
\mathrm{cov}\left(c^j[p], c^j[k]\right) &= \mathrm{cov}\left(c^j[p + m\ 2^{l-1}], c^j[k + n\ 2^{l-1}]\right) \quad \forall p, k, m, n \in \mathbb{Z}. && (5.40)\\
\mathrm{cov}\left(d^j[p], d^j[k]\right) &= \mathrm{cov}\left(d^j[p + m\ 2^{l-1}], d^j[k + n\ 2^{l-1}]\right) \quad \forall p, k, m, n \in \mathbb{Z}. && (5.41)\\
\mathrm{cov}\left(c^j[p], d^j[k]\right) &= \mathrm{cov}\left(c^j[p + m\ 2^{l-1}], d^j[k + n\ 2^{l-1}]\right) \quad \forall p, k, m, n \in \mathbb{Z}. && (5.42)
\end{aligned}$$

alors les covariances entre coefficients d'approximation à l'échelle $j+1$ (obtenus après reconstruction) sont invariantes par translation de 2^l pixels :

$$\mathrm{cov}\left(c^{j+1}[p], c^{j+1}[k]\right) = \mathrm{cov}\left(c^{j+1}[p + m\ 2^l], c^{j+1}[k + n\ 2^l]\right) \quad \forall p, k, m, n \in \mathbb{Z}. \tag{5.43}$$

Démonstration : En utilisant la formule de recomposition présentée dans le mémo 9, la covariance

15. comme par exemple une meilleure approximation des covariances

Mémo 9 – Formule de recomposition

La formule de recomposition permettant d'obtenir les coefficients d'approximations à l'échelle $j+1$ connaissant les coefficients d'approximation et les coefficients de détails à l'échelle j est la suivante :

$$c^{j+1}[p] = \sum_{n_1} h[p-2n_1]c^j[n_1] + \sum_{n_1} g[p-2n_1]d^j[n_1]. \tag{5.44}$$

entre éléments d'approximation à l'échelle $j+1$ s'écrit :

$$\begin{aligned}
\operatorname{cov}\left(c^{j+1}[p], c^{j+1}[k]\right) &= \sum_{n_1,n_2} h[p-2n_1]h[k-2n_2]\operatorname{cov}\left(c^j[n_1], c^j[n_2]\right) \\
&+ \sum_{n_1,n_2} g[p-2n_1]h[k-2n_2]\operatorname{cov}\left(d^j[n_1], c^j[n_2]\right) \\
&+ \sum_{n_1,n_2} h[p-2n_1]g[k-2n_2]\operatorname{cov}\left(c^j[n_1], d^j[n_2]\right) \\
&+ \sum_{n_1,n_2} g[p-2n_1]g[k-2n_2]\operatorname{cov}\left(d^j[n_1], d^j[n_2]\right).
\end{aligned}$$

En effectuant les changements de variables suivants

$$n_3 = n_1 - 2^{l-1} \quad n_4 = n_2 - 2^{l-1}$$

cette équation se réécrit sous la forme :

$$\begin{aligned}
\operatorname{cov}\left(c^{j+1}[p], c^{j+1}[k]\right) &= \sum_{n_3,n_4} h[p+2^l-2n_3]h[k+2^l-2n_4]\operatorname{cov}\left(c^j[n_3-2^{l-1}], c^j[n_4-2^{l-1}]\right) \\
&+ \sum_{n_3,n_4} g[p+2^l-2n_3]h[k+2^l-2n_4]\operatorname{cov}\left(d^j[n_3-2^{l-1}], c^j[n_4-2^{l-1}]\right) \\
&+ \sum_{n_3,n_4} h[p+2^l-2n_3]g[k+2^l-2n_4]\operatorname{cov}\left(c^j[n_3-2^{l-1}], d^j[n_4-2^{l-1}]\right) \\
&+ \sum_{n_3,n_4} g[p+2^l-2n_3]g[k+2^l-2n_4]\operatorname{cov}\left(d^j[n_3-2^{l-1}], d^j[n_4-2^{l-1}]\right).
\end{aligned}$$

En utilisant le fait que les covariances sont invariantes par translation de 2^{l-1} pixels, cette équation devient :

$$\begin{aligned}
\operatorname{cov}\left(c^{j+1}[p], c^{j+1}[k]\right) &= \sum_{n_3,n_4} h[p+2^l-2n_3]h[k+2^l-2n_4]\operatorname{cov}\left(c^j[n_3], c^j[n_4]\right) \\
&+ \sum_{n_3,n_4} g[p+2^l-2n_3]h[k+2^l-2n_4]\operatorname{cov}\left(d^j[n_3], c^j[n_4]\right) \\
&+ \sum_{n_3,n_4} h[p+2^l-2n_3]g[k+2^l-2n_4]\operatorname{cov}\left(c^j[n_3], d^j[n_4]\right) \\
&+ \sum_{n_3,n_4} g[p+2^l-2n_3]g[k+2^l-2n_4]\operatorname{cov}\left(d^j[n_3], d^j[n_4]\right) \\
&= \operatorname{cov}\left(c^{j+1}[p+2^l], c^{j+1}[k+2^l]\right).
\end{aligned}$$

■

Pour rappel, la propriété 5.4.4 nous assure, lorsque les covariances sont invariantes par translation dans l'espace des pixels, que les variances sont invariantes par translation entre les coefficients de l'échelle j. La combinaison de cette propriété avec la propriété précédente nous assure donc que le "modèle" de corrélation pour un pixel donné induit par l'utilisation d'une matrice diagonale dans une base d'ondelettes se répète tous les 2^l pixels où l caractérise le nombre de niveaux de décomposition.
Pour effectuer la phase de renormalisation présentée en équation (5.39) il nous faut, au plus, calculer $2^l \times 2^l$ coefficients de renormalisation (en 2 dimensions) si les covariances sont invariantes par translation au sein de chaque bloc de la matrice présentée en figure 5.8.

Remarque 5.22

La propriété précédente provient du fait que la transformée en ondelettes sur $\mathbf{l}$ *échelles est invariante (en un certain sens) aux translations de* $2^{\mathbf{l}}$ *pixels.*

5.6 Estimation de la matrice dans une trame de curvelettes

Contrairement à ce qui a été effectué avec des bases d'ondelettes, aucune méthode analytique pour obtenir la "vraie" matrice de covariance dans une trame de curvelette n'est utilisée. En effet, créée par la méthode présentée en section 5.4.1, la matrice $\mathbf{R_C} = \mathbf{C}\mathbf{R}_{Pix}\mathbf{C}^T$ ne nous serait d'aucune utilité. Cette matrice est en effet de rang n avec $n \ll m$ (où n et m sont le nombre d'éléments dans la base des pixels et la trame de curvelettes) est donc non inversible.

Mémo 10 – Trame de Parseval

Pour rappel, la famille de curvelettes utilisée est une trame de Parseval. Ceci rend le travail dans cette trame plutôt aisé car :

$$\mathbf{C}^T\mathbf{C} = \mathbf{I_n}. \tag{5.45}$$

Dans la suite nous nous servirons plusieurs fois de ce fait. Cependant, il ne faut pas oublier que $\mathbf{C}^T \in \mathcal{M}_{n,m}$ est uniquement une *pseudo-inverse* de $\mathbf{C} \in \mathcal{M}_{m,n}$:

$$\mathbf{C}\mathbf{C}^T = A \in \mathcal{M}_{m,m} \tag{5.46}$$

avec A une matrice non inversible (de rang n).

En section 4.2.3 nous avions défini la matrice de pondération dans la trame des curvelettes, notée $\mathbf{K_C}$, pour un bruit décorrélé de telle sorte que :

$$\mathbf{C}^T\mathbf{K_C}\mathbf{C} = \mathbf{R}_{pix}^{-1}. \tag{5.47}$$

Lorsque $\mathbf{R}_{pix} = \sigma^2\mathbf{I_n}$, cela nous a conduit au chapitre 4 à considérer la matrice

$$\mathbf{K_C} = \sigma^2\mathbf{I_m} \tag{5.48}$$

car $\mathbf{C}^T\mathbf{C} = \mathbf{I_n}$.

Remarque 5.23

La matrice définie en (5.47) est telle que :

$$\begin{aligned} J^o_{pix} &= (I^o - I^{synth})^T\mathbf{R}_{pix}^{-1}(I^o - I^{synth}) \\ &= (I^o - I^{synth})^T\mathbf{C}^T\mathbf{K_C}\mathbf{C}(I^o - I^{synth}) = J^o_{\mathbf{C}} \end{aligned}$$

Lorsque $\mathbf{R}_{pix} = \mathbf{I_n}$ *la matrice (5.48) n'est pas la seule matrice satisfaisant cette condition. C'est par exemple aussi le cas de la matrice* $\mathbf{C}\mathbf{C}^T$.

5.6.1 Utilisation de matrice de poids symétrique définie positive

Pour chaque matrice de covariance $\mathbf{R}_{pix}$ inversible, il existe en fait différentes matrice de poids $\mathbf{K_C}$ vérifiant cette relation. C'est par exemple toujours le cas de la matrice $\mathbf{K_C} = \mathbf{C}\mathbf{R}_{pix}^{-1}\mathbf{C}^T$.

Bien que cette matrice puisse toujours être utilisée, elle comporte deux défauts majeurs :

— elle est bâtie en utilisant l'inversion de la matrice $\mathbf{R}_{pix}$ (inversion qu'on ne souhaite pas effectuer) et son coût de construction (connaissant $\mathbf{R}_{pix}^{-1}$) est probablement trop cher.

— elle ne peut pas être diagonale (et inversible) car cette matrice est seulement semi-définie positive (car $\mathbf{C}$, $\mathbf{C}^T$ et $\mathbf{R}_{pix}^{-1}$ sont des opérateurs de rang n). Par conséquent, en utilisant cette définition on se priverait délibéremment dans certains cas de l'utilisation de matrices plus *"sympathiques"*.

Peut-on utiliser des matrices de pondération diagonales ?

Mémo 11 – Notations

Dans cette partie nous utilisons uniquement des matrices de pondération dans l'espace des curvelettes. Ces matrices sont notées $\mathbf{K_C}$ lorsqu'il s'agit des matrices exactes au sens de l'équation (5.47). Lorsqu'on utilise des matrices de pondérations approchées la notation $\tilde{\mathbf{D}}_{\mathbf{C},approx}$ est employée.

Comme pour la transformée en ondelettes, dans une trame de curvelettes nous nous intéressons exclusivement dans ces travaux à la construction de matrices de pondération diagonales.

La propriété suivante nous assure que lorsque $\mathbf{K_C}$ est une matrice diagonale positive alors la matrice $\mathbf{C}^T\mathbf{K_C}\mathbf{C}$ est bien une matrice de covariance inversible. Elle nous assure donc qu'utiliser une matrice de covariance diagonale dans une trame de curvelettes a un certain sens (i-e qu'on est bien en train de minimiser une certaine distance de Mahalanobis dans l'espace des pixels).

Propriété 5.3

Soit $\mathbf{K_C}$ une matrice symétrique définie positive de $\mathcal{M}_{m,m}$. La matrice $\mathbf{C}^T\mathbf{K_C}\mathbf{C}$ est une matrice définie positive (dans $\mathcal{M}_{n,n}$).
De plus, toute matrice de covariance inversible $\mathbf{R}_{pix}$ peut être obtenue par l'équation (5.47) en choisissant une matrice $\mathbf{K_C}$ symétrique définie positive.

Démonstration : Première partie On sait que $\mathbf{K_C}$ est une matrice symétrique définie positive dans $\mathcal{M}_{m,m}$. Par conséquent, pour tout vecteur $y \in \mathbb{R}^m \setminus \{0_m\}$ on a :

$$y^T\mathbf{K_C}y > 0 \tag{5.49}$$

Ceci est donc vrai pour tout vecteur $y_c = \mathbf{C}x$ non nul. Or y_c est non nul $\forall x \in \mathbb{R}^n \setminus \{0_n\}$ (car C est de rang n). Par conséquent, nous avons $\forall x \in \mathbb{R}^n \setminus \{0_n\}$:

$$x^T\mathbf{C}^T\mathbf{K_C}\mathbf{C}x > 0. \tag{5.50}$$

La matrice $\mathbf{C}^T\mathbf{K_C}\mathbf{C}$ est donc bien une matrice de covariance inversible dans $\mathcal{M}_{n,n}$.

Seconde partie Cherchons maintenant à voir quelles sont les conditions sur la matrice $\mathbf{K_C}$ pour que $\mathbf{R}_{pix}^{-1}$ s'écrive sous la forme $\mathbf{C}^T\mathbf{K_C}\mathbf{C}$.

Pour cela prenons une matrice $\mathbf{C}^T\mathbf{K_C}\mathbf{C}$ avec $\mathbf{K_C}$ une matrice quelconque. Afin que cette matrice puisse être une matrice de covariance inversible il faut et il suffit qu'elle soit une matrice symétrique définie positive, i-e, que

$$x^T\mathbf{C}^T\mathbf{K_C}\mathbf{C}x > 0 \quad \forall x \in \mathbb{R}^n \setminus \{0_n\}. \tag{5.51}$$

Prenons une décomposition [16] en valeur singulière de $\mathbf{C}$:

$$\mathbf{C} = U\Sigma V^T \tag{5.52}$$

avec U une matrice unitaire de taille $m \times m$, Σ une matrice de taille $m \times n$ dont les coefficients diagonaux sont des réels positifs ou nuls (et tous les autres coefficents sont nuls) et V est une matrice unitaire de taille $n \times n$.

16. La décomposition en valeur singulière d'une matrice dégénérée n'est pas unique.

Le mémo 10 nous rappelle que la matrice $\mathbf{C}^T\mathbf{C} = V\Sigma^T U^T U\Sigma V^T = \mathbf{I_n}$. Les matrices U et V étant unitaires, on a donc :

$$\Sigma^T\Sigma = \mathbf{I_n}. \tag{5.53}$$

Par conséquent, la matrice Σ est la matrice suivante :

$$\Sigma = \begin{pmatrix} \mathbf{I_n} \\ 0 \end{pmatrix}. \tag{5.54}$$

En utilisant cette décomposition en valeur singulière, la condition (5.51) devient :

$$x^T V\Sigma^T U^T \mathbf{K_C} U\Sigma V^T x > 0 \quad \forall x \in \mathbb{R}^n \setminus \{0_n\}. \tag{5.55}$$

Dénommons la matrice $R_u = U^T\mathbf{K_C}U$. Cette matrice s'écrit, sans perte de généralité, sous la forme :

$$R_u = \begin{pmatrix} R_{u1} & R_{u2} \\ R_{u3} & R_{u4} \end{pmatrix} \tag{5.56}$$

avec R_{u1} une matrice de taille $n \times n$. Introduisons le vecteur

$$y = V^T x. \tag{5.57}$$

Ce vecteur est non nul lorsque x est non nul.
En remarquant que $\Sigma^T R_u \Sigma = R_{u1}$ la condition (5.55) s'écrit donc sous la forme :

$$y^T R_{u1} y > 0 \quad \forall y \in \mathbb{R}^n \setminus \{0_n\}. \tag{5.58}$$

Afin de satifaire cette condition, il suffit donc que R_{u1} soit une matrice symétrique définie positive.

Quelque soit la matrice $\mathbf{R}_{pix}^{-1}$, on peut donc choisir une matrice $\mathbf{K_C}$ de rang plein de telle sorte que les matrices $R_{u2} = R_{u3} = 0_{m-n}$ et par exemple $R_{u4} = \mathbf{I_{m-n}}$. La matrice R_u ainsi définie est alors symétrique définie positive (et par conséquent $\mathbf{K_C} = UR_uU^T$ l'est aussi). ■

Il suffit donc que $VR_{u1}V^T = \mathbf{R}_{pix}^{-1}$, i-e $R_{u1} = V^T\mathbf{R}_{pix}^{-1}V$ pour qu'à partir de la matrice $\mathbf{K_C}$ on obtienne la matrice $\mathbf{R}_{pix}^{-1}$ via l'équation (5.47) :

$$\begin{aligned}
\mathbf{C}^T\mathbf{K_C}\mathbf{C} &= \mathbf{C}^T U R_u U^T \mathbf{C} \\
&= \underbrace{V\Sigma^T U^T}_{\mathbf{C}^T} U \begin{pmatrix} V^T\mathbf{R}_{pix}^{-1}V & R_{u2} \\ R_{u3} & R_{u4} \end{pmatrix} U^T \underbrace{U\Sigma V^T}_{\mathbf{C}} \\
&= V\Sigma^T \begin{pmatrix} V^T\mathbf{R}_{pix}^{-1}V & R_{u2} \\ R_{u3} & R_{u4} \end{pmatrix} \Sigma V^T \quad \text{car } U \text{ est unitaire} \\
&= VV^T\mathbf{R}_{pix}^{-1}VV^T \quad \text{en utilisant l'équation (5.54)} \\
&= \mathbf{R}_{pix}^{-1} \quad \text{car } V \text{ est unitaire}
\end{aligned}$$

Le choix de $\mathbf{K_C}$ n'est donc absolument pas unique. Il existe ainsi une infinité de choix possible pour définir $\mathbf{R}_{pix}^{-1}$.

Remarque 5.24

Bien que cette propriété ait été énoncée uniquement pour les curvelettes, elle s'applique dès lors que l'opérateur $\mathbf{C}$ *est de rang* n*. Pour voir cela, il suffit de décomposer l'opérateur* $\mathbf{C}$ *sous la forme* $U\Sigma\Lambda V^T$ *avec* Λ *une matrice carré diagonale telle que* $\Sigma^T\Sigma = \mathbf{I_n}$ *et poser* $y = \Lambda V^T$ *au lieu du changement proposé en équation (5.57). Par conséquent, quelque soit le type de trame considérée, on est certain qu'utiliser une matrice de pondération diagonale dont les éléments sont positifs induit bien l'utilisation d'une matrice de covariance inversible dans l'espace des pixels.*

Petite disgression sur la trace de la matrice $\mathbf{K_C}$

La propriété suivante nous renseigne sur le lien existant entre la trace de la matrice $\mathbf{K_C}$ et celle de $\mathbf{R}_{pix}^{-1}$ définie par l'équation (5.47). Cette propriété nous renseigne sur le fait que la trace de la matrice $\mathbf{K_C}$ doit nécessairement être plus grande que celle de $\mathbf{R}_{pix}^{-1}$ (si on souhaite que $\mathbf{C}^T\mathbf{K_C}\mathbf{C}$ ait la même trace que $\mathbf{R}_{pix}^{-1}$).

Propriété 5.4

Lorsque la matrice $\mathbf{K_C}$ est choisie symétrique définie positive de rang plein, on a :

$$Tr(\mathbf{K_C}) > Tr(\mathbf{C}^T\mathbf{K_C}\mathbf{C}) \tag{5.59}$$

Démonstration :

$$Tr(\mathbf{C}^T\mathbf{K_C}\mathbf{C}) = Tr(\Sigma^T U^T \mathbf{K_C} U \Sigma) \tag{5.60}$$

car V est une matrice unitaire et l'opération trace est invariante par permutation cyclique. De plus, comme précédemment $\Sigma^T U^T \mathbf{K_C} U \Sigma = R_{u1}$. Par conséquent $Tr(\mathbf{C}^T\mathbf{K_C}\mathbf{C}) = Tr(R_{u1})$.

$$Tr(\mathbf{K_C}) = Tr(U^T\mathbf{K_C}U) = Tr(R_{u1}) + Tr(R_{u2}) \tag{5.61}$$

car U est une matrice unitaire et la trace est invariante par permutation cyclique.

Etant donné que $\mathbf{K_C}$ est de rang m et symétrique définie positive, la matrice $U^T\mathbf{K_C}U$ l'est aussi. Donc $Tr(R_{u2}) = Tr(\mathbf{K_C}) - Tr(\mathbf{C}^T\mathbf{K_C}\mathbf{C}) > 0$. ■

Cette propriété est intéressante dans le sens où elle nous indique que la somme des poids dans une trame de curvelettes est nécessaiement plus grande que son équivalent dans l'espace des pixels lorsque la matrice choisie est diagonale (et inversible). On a donc, en quelque sorte, une perte de somme de variance lors du passage de $\mathbf{K_C}$ à la matrice $\mathbf{R}_{pix}^{-1}$ définie en équation (5.47).

Remarque 5.25

La propriété précédente reste valable quelque soit la trame de Parseval choisie. L'unique condition utilisée dans cette propriété étant que $\mathbf{C}^T\mathbf{C} = \mathbf{I_n}$.

5.6.2 Restriction à des matrices diagonales

Le problème dans une trame de curvelettes (et dans d'autres trames) est que, connaissant $\mathbf{R}_{pix}$, il est complexe de construire une matrice $\mathbf{K_C}$ symétrique définie positive telle qu'étant l'équivalente de $\mathbf{R}_{pix}^{-1}$ au sens de l'équation (5.47).

Qu'allons nous faire ?

Dans cette section, deux approximations diagonales différentes, notées $\tilde{\mathbf{D}}_{\mathbf{C},approx}$ sont étudiées. La première peut être mise en place de manière simple, mais possède quelques problèmes intrinsèques. La seconde est plus compliquée à mettre en place. Aucune expérience d'assimilation n'a été effectuée en utilisant cette seconde technique (si on excepte les expériences faites dans le chapitre précédent avec la matrice $\mathbf{K_C} = \mathbf{I_m}$) qui doit encore être un peu développée. Cependant, cette seconde technique est présentée afin de mieux cerner les problèmes inhérents à la première méthode. Elle illustre le potentiel des curvelettes (et d'autres trames) pour représenter des matrices de covariances complexes via l'utilisation de matrices diagonales.

Comme dans la section précédente, nous allons régulièrement visualiser les matrices $\mathbf{R}_{pixEq}$ induites par l'utilisation de matrice de pondération. Cette dernière est calculée de la manière suivante :

$$\mathbf{R}_{pixEq} = \left(\mathbf{C}^T\tilde{\mathbf{D}}_{\mathbf{C},approx}\mathbf{C}\right)^{-1}. \tag{5.62}$$

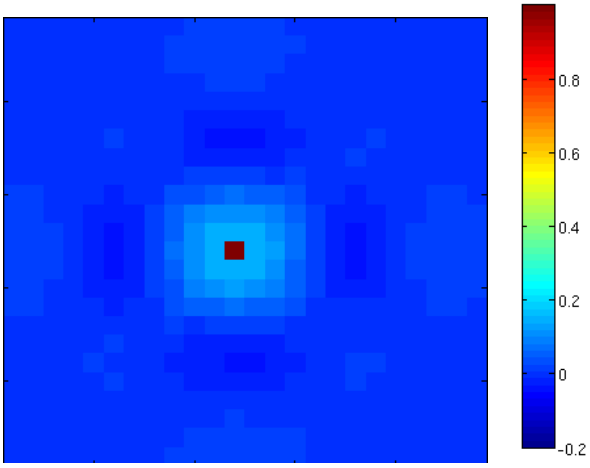


FIGURE 5.17 – Corrélation entre un pixel et ses voisins induite par l'utilisation de la matrice $\tilde{\mathbf{D}}_{\mathbf{C},var}$ dans le cas où $\mathbf{R}_{pix} = \mathbf{I_n}$.

Comparer qualitativement les matrices $\mathbf{R}_{pixEq}$ et $\mathbf{R}_{pix}$ permet alors de visualiser les approximations réalisées lors de l'utilisation de la matrice de pondération $\tilde{\mathbf{D}}_{\mathbf{C},approx}$.

Construction via estimation de la variance

La première manière envisagée pour construire une matrice de pondération diagonale consiste à estimer la variance dans l'espace des curvelettes. L'approximation diagonale $\tilde{\mathbf{D}}_{\mathbf{C},var}$ est alors définie comme l'inverse de la diagonale de la matrice de covariance :

$$\tilde{\mathbf{D}}_{\mathbf{C},var} = \left(\mathrm{diag}(\mathbf{C}\mathbf{R}_{pix}\mathbf{C}^T)\right)^{-1}. \tag{5.63}$$

Afin de s'épargner des heures de calculs pour obtenir une approximation de qualité discutable, seule une estimation de $\mathbf{C}\mathbf{R}_{pix}\mathbf{C}^T$ est effectuée. Cette dernière est réalisée par l'algorithme, décrit en section 5.3.2 pour les ondelettes, permettant d'estimer la variance de chaque coefficient de curvelettes (en supposant la connaissance de $\mathbf{R}_{pix}^{1/2}$).

Choix non optimal ...
Il est bon de noter que ce choix de matrice de poids ne correspond pas nécessairement (si elle existe) à la matrice de pondération diagonale permettant de représenter sans approximation la matrice $\mathbf{R}_{pix}^{-1}$.

Afin d'illustrer ceci, la méthode précédente a été employée pour déterminer la matrice diagonale correspondant à un bruit indépendant en espace et identiquement distribué.

La figure 5.17 présente (via l'utilisation de $\mathbf{R}_{pixEq}$) les corrélations obtenues par cette estimation. La covariance d'un pixel avec ses voisins induite par l'utilisation de la matrice $\tilde{\mathbf{D}}_{\mathbf{C},var}$ (équation (5.63)) est présentée.

On observe, que contrairement à la matrice $\mathbf{K_C} = \mathbf{I_m}$ cette matrice ne permet pas de représenter correctement la matrice $\mathbf{R}_{pix}$. Chaque pixel n'est ainsi pas décorrélé de ses voisins (même si les corrélations équivalentes sont assez peu importantes). Phénomène plus inquiétant (non illustré sur la figure 5.17) la variance de chaque pixel est fortement sous évaluée (d'un facteur 7.2 correspondant à la redondance de la transformée en curvelettes) lorsque cette technique est employée. Il est donc nécessaire d'effectuer une renormalisation afin de ne pas sous-estimer la variance. Dans ce cas, l'écart aux observations est défini comme :

$$(\mathbf{C}\Lambda I^o - \mathbf{C}\Lambda I^t)^T \tilde{\mathbf{D}}_{\mathbf{C},var}(\mathbf{C}\Lambda I^o - \mathbf{C}\Lambda I^t) \tag{5.64}$$

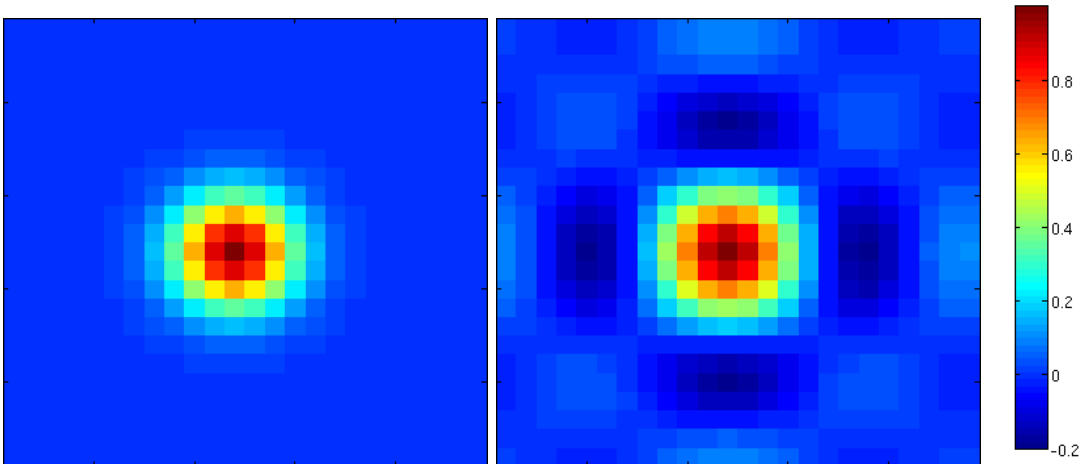


FIGURE 5.18 – Corrélation entre un pixel et ses voisins. A gauche les véritables corrélations sont rappelées. A droite les corrélations induites par l'utilisation de la matrice $\tilde{\mathbf{R}}_{\mathbf{C},var}$ est présentée.

où Λ est une matrice de renormalisation.

Cette renormalisation peut être faite de manière simple lorsqu'à l'origine le même modèle de corrélation est valable en tout pixel (voir remarque 5.22). Sinon calculer les coefficients de renormalisation peut s'avérer très onéreux.

Ce phénomène de sous-estimation de la variance n'est pas un cas isolé : dans tous les exemples traités il y a une sous-estimation importante de cette dernière. Malgré tout, les corrélations sont assez bien représentées et nous verrons en section 5.8 que ce choix de matrice est en pratique redoutablement efficace dans les cas considérés (au moins en comparaison avec les méthodes non basées sur une distance multi-échelles).

... mais respectant relativement bien les corrélations ...
La figure 5.18 présente les corrélations entre un pixel et ses voisins induites par l'utilisation de $\tilde{\mathbf{D}}_{\mathbf{C},var}$. On observe que le modèle de corrélation est ici bien approché. L'isotropie du bruit est dans ce cadre véritablement prise en compte et constitue donc une amélioration vis–à–vis des approximations obtenues par l'utilisation de matrices diagonales dans des bases d'ondelettes présentées en figure 5.14. De plus, contrairement à ces dernières (voir figure 5.15), le même modèle de corrélation est répété en chaque pixel. Cependant, les corrélations négatives sont plus importantes que dans le cas des ondelettes de Daubechies et la portée des corrélations parasites est bien plus importante (voir figure 5.21).

... dans certains cas Un défaut propre à cette méthode pour estimer la matrice $\tilde{\mathbf{D}}_{\mathbf{C},var}$ est que pour des modèles de corrélations plus complexes, les approximations faites dans l'espace des curvelettes sont assez mauvaises. La figure 5.19 présente, pour un exemple de corrélation anisotrope (en haut à gauche), les corrélations induites par l'utilisation de la matrice $\tilde{\mathbf{D}}_{\mathbf{C},var}$ (en bas à gauche) et la matrice diagonale dans la base d'ondelettes de Daubechies à 4 moments nuls $\tilde{\mathbf{R}}^{\text{cor}}_{\mathbf{W},\text{diag}}$ (en haut à droite). On observe que ces deux matrices n'arrivent pas à modéliser correctement la direction privilégiée dans les corrélations initiales bien que l'approximation dans la trame de curvelette soit de meilleure qualité.

Dans la partie qui suit, une autre méthode pour estimer la matrice de pondération $\tilde{\mathbf{D}}_{\mathbf{C},est}$ est introduite. Nous verrons que cette méthode permet de pallier (en partie) les différents défauts visualisés jusqu'à présent. Cependant cette dernière n'a pas encore été intégrée au code d'assimilation (écrit en FORTRAN) mais simplement testée sous Matlab dans le but de montrer que des solutions aux problèmes décrits précédemment dans cette section existent.

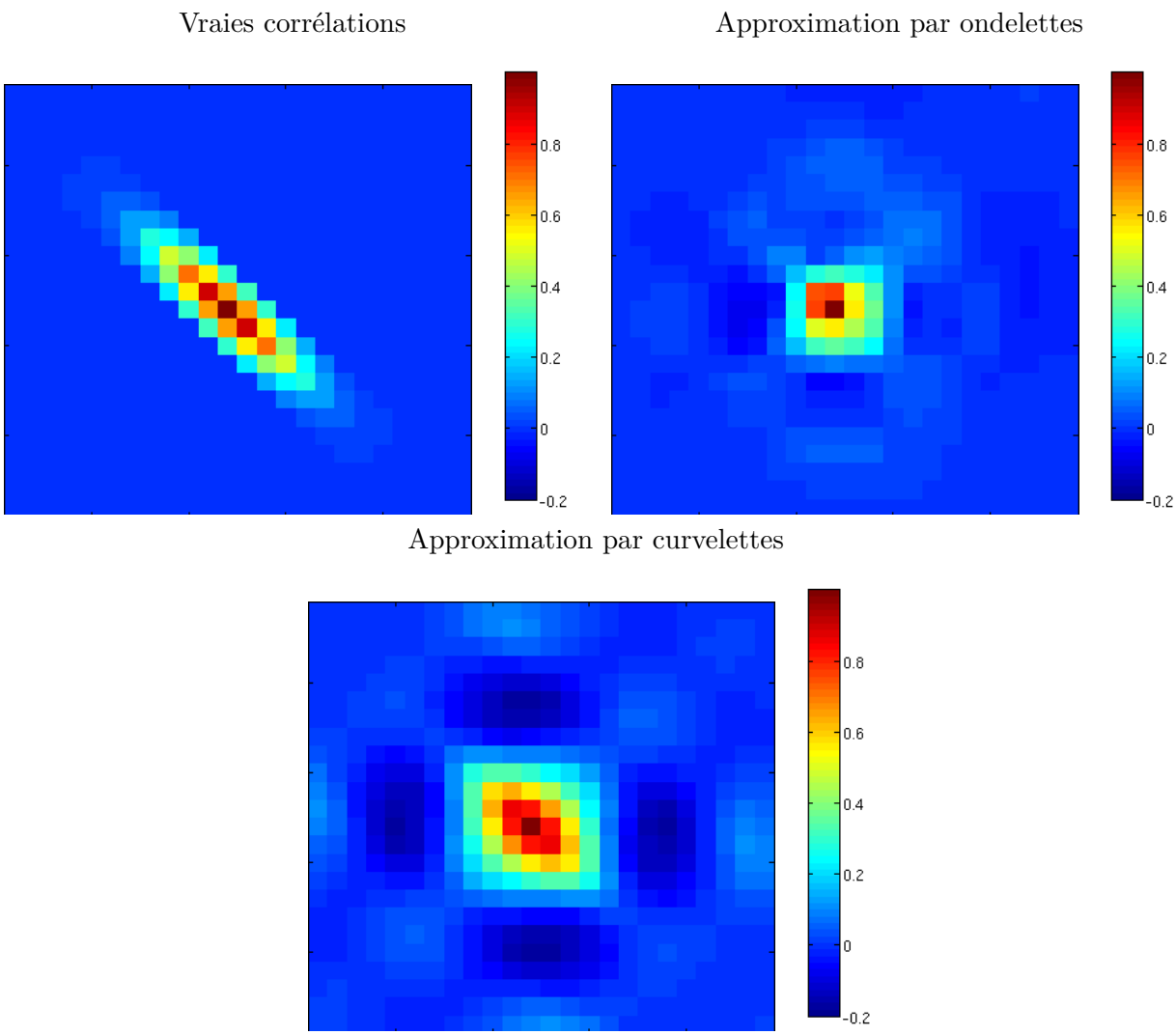


FIGURE 5.19 – Modélisation de covariances anisotropiques. En haut à gauche les vraies covariances entre un pixels et ses voisins. Les images suivantes présentent les covariances induites par l'utilisation d'une matrice diagonale dans une base d'ondelette (en haut à droite), provenant de la matrice de covariance dans une trame de curvelettes (en bas à gauche).

Estimation de la matrice de pondération par minimisation

Au vu de la manière de choisir la matrice $\tilde{\mathbf{D}}_{\mathbf{C},var}$, il est légitime de se demander s'il existe d'autres manières de construire une matrice de pondération diagonale $\tilde{\mathbf{D}}_{\mathbf{C},}$ pouvant conduire à une meilleur prise en compte des corrélations au sein des erreurs.

La méthode présentée ici consiste à définir une **matrice diagonale** positive $\tilde{\mathbf{D}}_{\mathbf{C},est}$ comme un minimum du problème suivant :

$$\begin{cases} \text{Trouver } \tilde{\mathbf{D}}_{\mathbf{C},est} \text{ tel que} \\ \tilde{\mathbf{D}}_{\mathbf{C},est} = \underset{\tilde{\mathbf{D}}_{\mathbf{C},} \in \Omega}{\operatorname{argmin}} \|\mathbf{C}^T \tilde{\mathbf{D}}_{\mathbf{C},} \mathbf{C} \mathbf{R}_{pix} y - y\|^2 \\ \text{où } \Omega \text{ est l'ensemble des matrices diagonales} \\ \text{(dont les éléments sont tous positifs)} \end{cases} \tag{5.65}$$

où y est un vecteur de $\mathbb{R}^n$.

Remarque 5.26

On ne cherche donc pas à inverser la matrice $\mathbf{R}_{pix}$*, mais on cherche la diagonale dans l'espace*

des curvelettes qui permet d'être le plus proche de cette inverse (au sens des moindres carrés) pour un vecteur y. Afin d'être moins sensible au choix du vecteur y, cette minimisation est effectuée (en une seule fois) avec 10 vecteurs y différents.

Afin de simplifier grandement ce problème d'estimation (qui n'est dans un premier temps pas la chose importante), nous supposons que le poids des coefficients de curvelettes ne dépend pas de la position mais uniquement de l'échelle et de la direction. Ceci signifie qu'on souhaite estimer un modèle de corrélation invariant par translation dans l'espace des pixels.

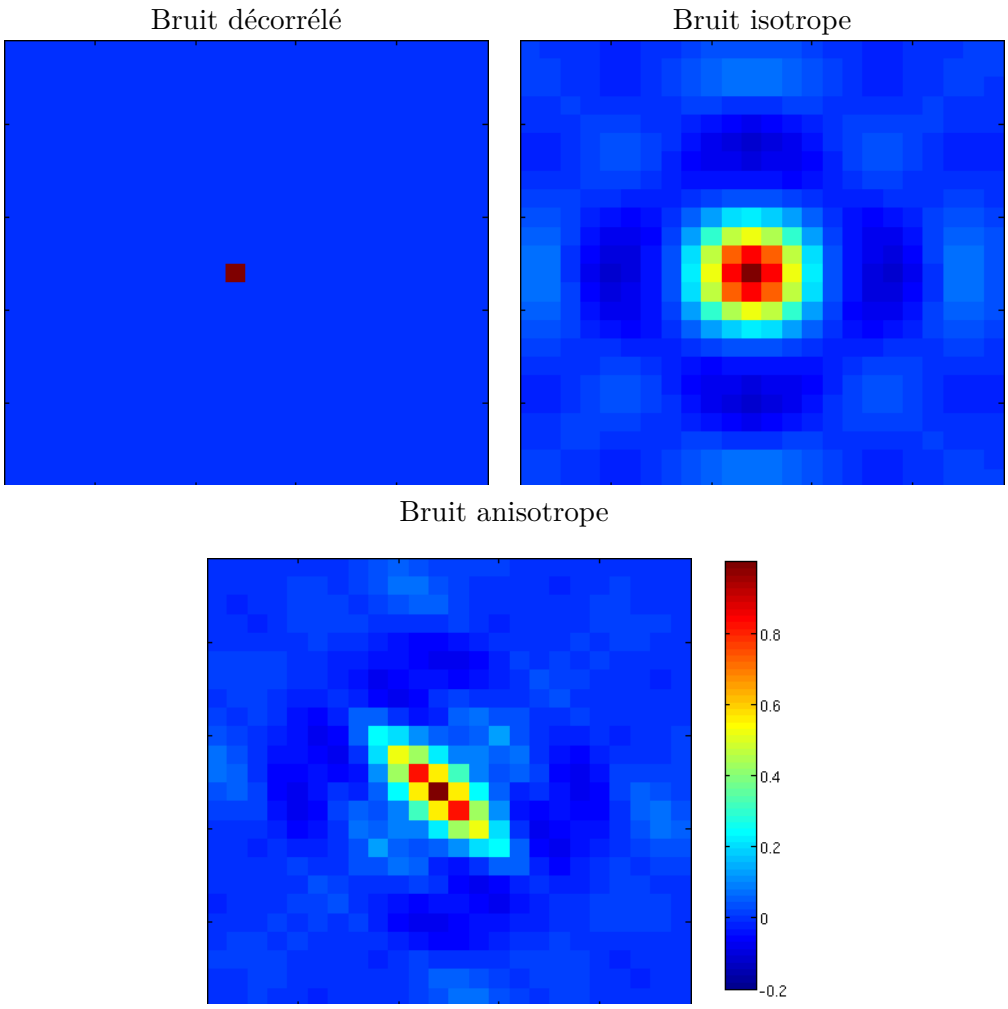


FIGURE 5.20 – Résultat induit par la minimisation de la fonctionnelle (5.65) pour modéliser les corrélations présentées en figure 5.17, 5.18 et 5.19.

Remarque 5.27

Le coût de la minimisation de la fonctionnelle (5.65) est raisonnable. Pour évaluer cette fonction coût il est nécessaire d'effectuer

1. *Le calcul de $x = \mathbf{CR}_{pix}y$ une unique fois en début de minimisation. Le cout de ce calcul est de $O(N^2)$ (pour la multiplication matrice vecteur) + $O(N\log(N))$ (pour l'application $\mathbf{C}$ à un vecteur),*
2. *Le calcul de $z = \tilde{\mathbf{D}}_{\mathbf{C},est}x$ à chaque itération. Ce dernier à un coût de M multiplications (où $M \simeq 7N$ est le nombre d'éléments dans la trame de curvelettes).*
3. *Le calcul de $\mathbf{C}^T z$ dont le coût est en $O(N\log(N))$ (où N est le nombre de pixels de l'image).*

Dès lors l'évaluation du coût à une itération se fait en $O(N\log(N)) + O(M)$ (sauf pour

la première itération). Pour le reste des opérations nécessaires à la minimisation (par ex. calcul du gradient, calcul du hessien) cela dépend de la méthode utilisée.

Des exemples de résultat induit par cette minimisation, réalisée sous matlab via la fonction *lsqcurvfit* utilisant l'algorithme de Levenberg-Marquardt [17] (n'utilisant ni la connaissance du gradient ni celle du Hessien), sont présentés en figure 5.20.

— Les corrélations modélisées par cette méthode lorsque le bruit est décorrélé sont présentées en haut à gauche. On observe que dans ce cas elle concordent aux véritables corrélations : il n'y a pas (contrairement aux résultats présentés en figure 5.17) apparition de corrélations parasites. De plus, contrairement à l'utilisation de $\tilde{\mathbf{D}}_{\mathbf{C},var}$ les variances ne sont pas sous-estimées mais en parfaite adéquation avec les véritables variances.

— L'approximation effectuée lorsque le bruit est isotrope est présentée en haut à droite. On observe que cette approximation est bien meilleure qu'en utilisant des ondelettes et qu'il y a un gain par rapport à l'utilisation de la matrice $\tilde{\mathbf{D}}_{\mathbf{C},var}$: cette nouvelle modélisation respecte un peu mieux la structure aux endroits où des corrélations existent tout en atténuant fortement les corrélations parasites résultant de l'approximation $\tilde{\mathbf{D}}_{\mathbf{C},var}$.

— En bas au centre est présentée l'approximation obtenue lorsqu'on cherche à modéliser les corrélations anisotropes présentées en figure 5.19. On observe qu'on est capable, à partir d'une matrice de pondération diagonale, de bien prendre en compte l'anisotropie des corrélations (ce qui n'était le cas ni avec les ondelettes, ni avec les curvelettes en utilisant la matrice $\tilde{\mathbf{D}}_{\mathbf{C},var}$). L'approximation n'est certes pas parfaite mais la direction privilégiée est bien respectée et *peu* de corrélations parasites font leur apparition.

Remarque 5.28

Comme illustré pour le bruit décorrélé en espace, dans un cas où il existe une matrice de pondération diagonale, cette méthode peut permettre (si le problème de minimisation est bien résolu) de la retrouver. Lorsque la matrice est construite par estimation de la variance, il n'y a aucune raison que ce soit le cas.

Remarque 5.29 (*Variance dans l'espace des pixels*)

Il faut noter qu'avec cette méthode, les variances dans l'espace des pixels ne sont que peu sous (ou sur) estimées. Par exemple, dans le cas des corrélations anisotropes, les variances correspondant à l'utilisation de la matrice $\tilde{\mathbf{D}}_{\mathbf{C},est}$ sont de 1.3 quand celles d'origines étaient de 1.

Bilan 5.4

Bien que la matrice de pondération $\tilde{\mathbf{D}}_{\mathbf{C},var}$ puisse approcher de manière acceptable les corrélations présentes dans le signal dans certains cas, il est possible d'effectuer (toujours avec une matrice diagonale) de bien meilleures approximation de la matrice $\mathbf{R}_{pix}$ au prix d'une minimisation ($\tilde{\mathbf{D}}_{\mathbf{C},est}$).

Corrélations parasites à longue distance

Les corrélations présentés dans les figures 5.18, 5.19 et 5.20 sont en fait un zoom sur les corrélations à courte distance induites par une matrice de corrélation diagonale. Cependant, lors de l'utilisation d'une matrice de pondération dans l'espace des curvelettes, des corrélations longues distances font leur apparition (ce qui était moins le cas avec des bases d'ondelettes). La figure 5.21 présente l'ensemble des corrélations entre un point et les autres points de l'image lors de l'utilisation de différentes distances.

— Le modèle de corrélation recherchée est présenté en haut à gauche. Nous nous cherchons donc à représenter des corrélations anisotropes.

17. Pour plus d'information sur la méthode de minimisation utilisée, se réferer à la documentation de matlab http://www.mathworks.fr/fr/help/optim/ug/lsqcurvefit.html.

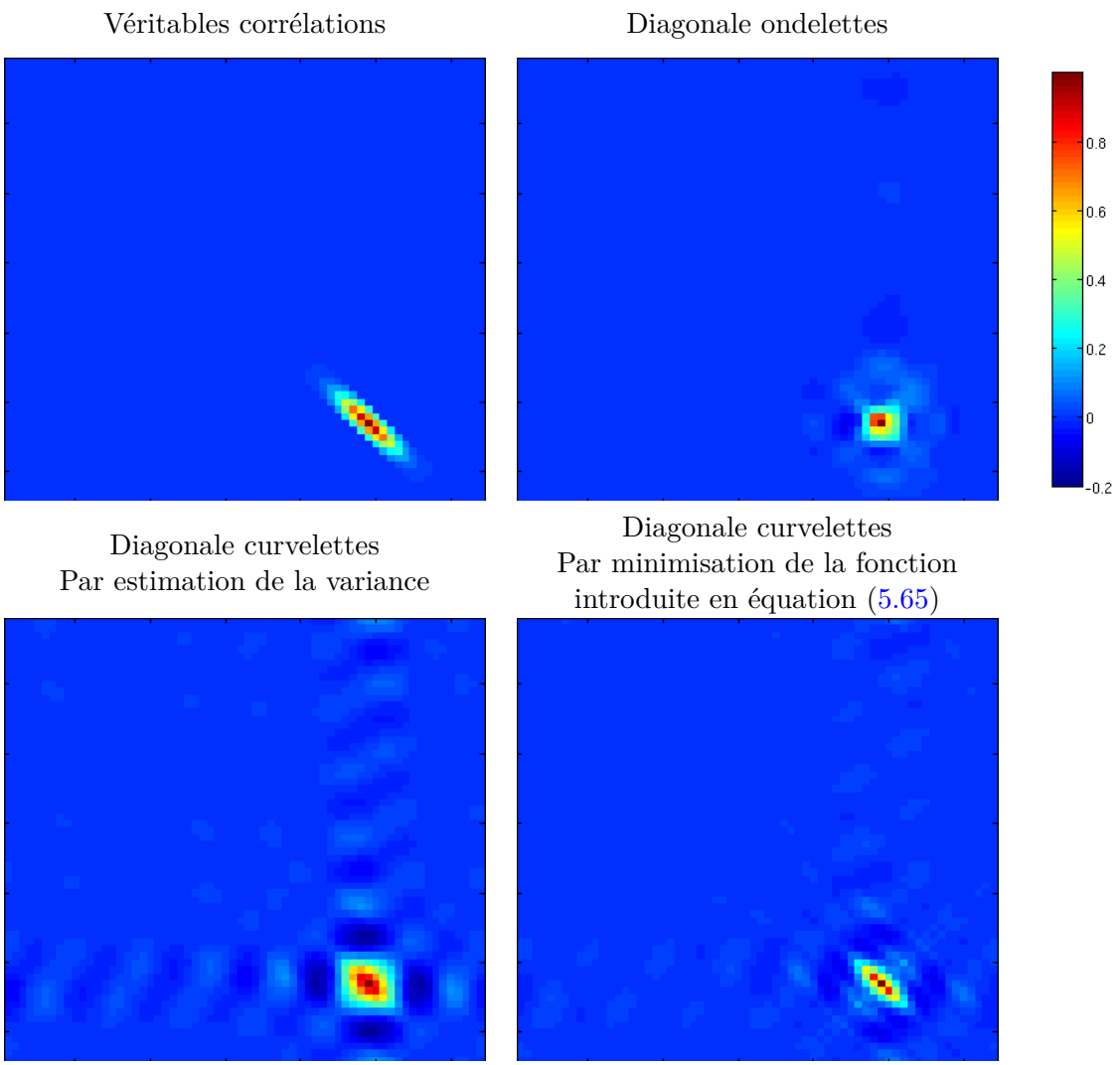


FIGURE 5.21 – Corrélation d'un pixels avec ces voisins. Présentation des effets de différentes matrices de pondération diagonales.

— En haut à droite, les corrélations obtenues avec des ondelettes de Daubechies à 4 moments nuls sont présentés. On observe que bien que ces corrélations ne soient pas les bonnes, elles ne présentent pas de corrélation longue distance parasites (ou du moins, moins que les autres modélisations).
— En bas à gauche sont présentées les corrélations obtenues lors de l'utilisation de la matrice $\tilde{\mathbf{D}}_{\mathbf{C},var}$. Cette matrice, en plus de mal prendre en compte l'anisotropie du bruit comme présenté en figure 5.19 induit des corrélations longues distances d'amplitude non négligeable.
— Ces problèmes sont en partie résolus lors de l'utilisation de la matrice de pondération estimées par minimisation de la fonctionnelle (5.65). Utiliser la matrice $\tilde{\mathbf{D}}_{\mathbf{C},est}$ permet en effet de représenter de manière correcte l'anisotropie du bruit (comme présenté en figure 5.19) et réduit fortement les problèmes liées aux corrélations parasites.

5.7 Estimation de la matrice de covariance dans un espace de Fourier

L'idée consistant à considérer une matrice de covariance d'erreur d'ébauche diagonale après un changement d'espace est une idée assez ancienne. Cette idée a par exemple été mise en œuvre

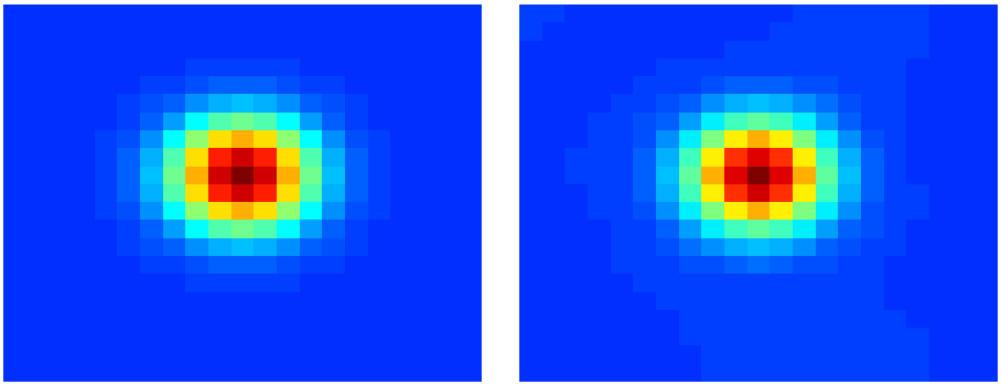
dès [Courtier et al., 1998] en utilisant une transformée de Fourier.

L'espace spectral est connu pour permettre de représenter uniquement des modèles de corrélations homogènes en espace à l'aide d'une matrice diagonale.

Dans les expériences de la partie 5.8, la détermination de la diagonale de la matrice de corrélation est effectuée à l'aide de l'algorithme décrit en section 5.3.2 pour les ondelettes. Cet algorithme permet d'estimer la variance de chaque coefficient dans l'espace spectral lorsque la matrice $\mathbf{R}_{pix}^{1/2}$ est connue.

D'un point de vue numérique, la transformée de Fourier utilisée est empruntée à la librairie MCALab[Fadili et al., 2010a]. Elle provient des travaux de [Frigo and Johnson, 2005] [18].

Exemple de modélisation :
La figure 5.22 présente un exemple de corrélations (entre un pixels et ses voisins) induites par l'utilisation d'une telle transformée (à droite) ainsi que les corrélations souhaitées (à gauche). On observe que l'approximation effectuée est d'une meilleure qualité que les précédentes (dans des bases d'ondelettes (figure 5.14) ou de curvelettes (figures 5.18 et 5.20)).



Vérité

Approximation diagonale dans Fourier

FIGURE 5.22 – Approximation diagonale et modélisation de corrélation homogène isotrope dans Fourier

5.8 Expériences

Au cours des sections précédentes, nous avons introduit, dans chaque espace considéré, une ou plusieurs matrices de covariance. Dans cette section, les (pseudo)métriques combinées à ces différentes matrices sont utilisées au cours d'expériences jumelles pour évaluer leurs intérêts.

Dans un premier temps, nous nous intéressons à l'influence de l'amplitude du bruit sur l'analyse effectuée en utilisant les différentes (pseudo)métriques. Dans un second temps, nous verrons l'impact du choix de la matrice de covariance d'erreur d'observation sur celui de la matrice d'erreur d'ébauche. Dans la section suivante, l'importance d'avoir accès à l'information contenue dans les statistiques d'erreur d'observation est mise en avant. Puis, dans une dernière section, nous présentons des expériences pour lesquelles améliorer la prise en compte des statistiques d'erreur d'observation n'améliore pas nécessairement l'analyse.

18. On peut notamment la télécharger à l'adresse suivante http://www.fftw.org

Génération des corrélations

La distribution du bruit servant d'exemple dans cette section est celle présentée en figure 5.6. Il s'agit du bruit isotrope et homogène en espace ayant servi d'exemple dans les sections précédentes.

Création du bruit

1. Création d'un bruit blanc gaussien indépendant et identiquement distribué. Le vecteur de bruit est noté β. Il suit une loi $\mathcal{N}(0, \sigma^2 \mathbf{I_n})$.

2. Corrélations des erreurs au sein du bruit par l'application d'un opérateur G :

$$\epsilon = G \star \beta. \tag{5.66}$$

$\star$ est le produit de convolution et G un filtre gaussien de taille $(2n+1)(2n+1)$ tel que :

$$G(i,j) = \frac{1}{|G|} \exp\left(-\frac{i^2+j^2}{2\sigma_L^2}\right) \quad \text{for } i,j = -n, ..., n. \tag{5.67}$$

avec $|G| = \sum_{i,j} \exp\left(-\frac{i^2+j^2}{2\sigma_L^2}\right)$. Dans les expériences qui vont suivre, $\sigma_L = 1.5$ et $n = 9$. Le vecteur ϵ suit donc une loi normale $\mathcal{N}(0, \sigma^2 G G^T)$.

Remarque 5.30

Au final les corrélations peuvent être modélisées par des gaussiennes dont la longueur de corrélation est d'environ 3 pixels (dans chacune des directions).

La figure 5.23 présente, pour différents niveaux de bruit, la première image de séquences d'observations utilisées :

— L'image en haut à droite présente le résultat de l'ajout d'un bruit de faible amplitude (SNR = 26.8 dB) ;

— L'image en bas à gauche est telle que le bruit ajouté n'impacte pas la perception visuelle des principales structures présentes au sein de l'image. Le rapport signal bruit est de 20.8 dB pour cette image.

— En revanche, la perception des structures de l'image en bas à droite est altérée par la présence de bruit. Il est, avec une unique image, plus complexe d'avoir une idée claire de la délimitation du vortex et ce, bien qu'on ait une idée de la structure du bruit. Le rapport signal bruit de cette image est de 14.8 dB.

Notations et termes d'écarts aux observations utilisés

Les notations des termes d'écarts aux observations introduits dans les sections précédentes sont les suivantes :

- Les résultats obtenus en utilisant la comparaison :
  - entre pixels sont notés **Pixels-scal**. La matrice de covariance est ici scalaire (équation (5.1)),
  - entre gradients sont notés **∇-scal** lorsque la matrice de covariance est scalaire (équation (5.3)),
  - entre gradients normalisés sont notés **Normale-diag**. La variance de chaque élément est définie en équation (5.4).
- Concernant les transformées en ondelettes, les notations suivantes sont employées :
  - **Haar-diag** lorsque la base de Haar est utilisée avec une matrice de covariance diagonale (équation (5.37)). Dans ce cas, l'ensemble des coefficients est utilisé.

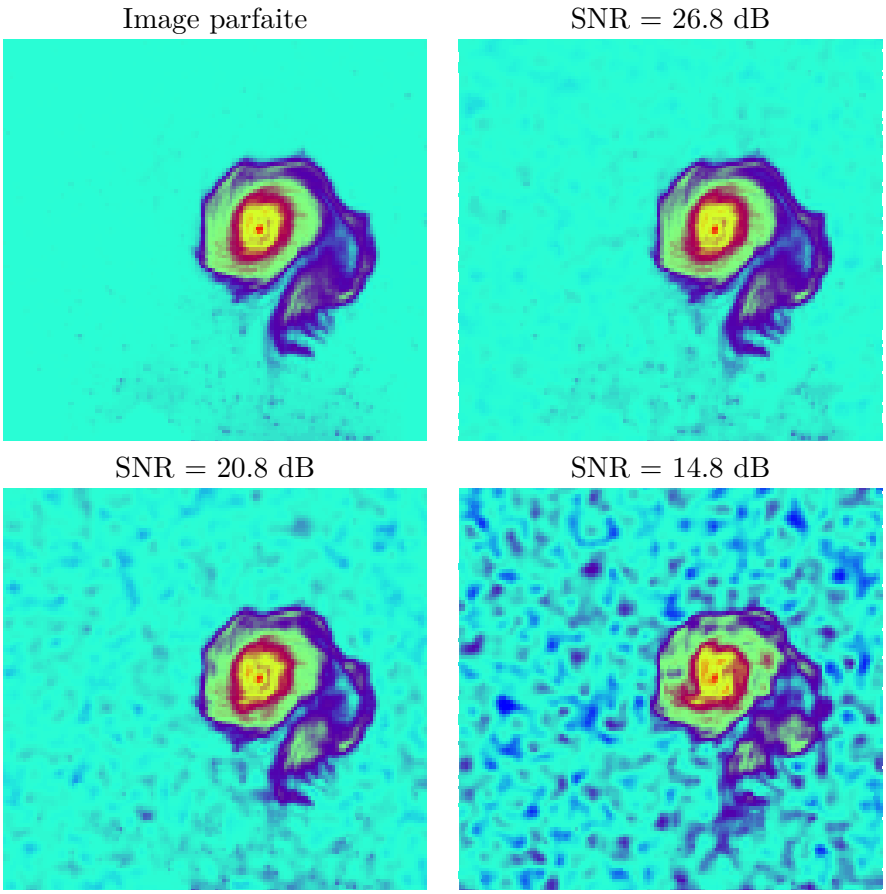


FIGURE 5.23 – Présentation du bruit

  * **Daub-diag** lorsque la base de Daubechies à 4 moments nuls est utilisée avec une matrice de covariance diagonale (équation (5.37)). Dans ce cas, l'ensemble des coefficients est utilisé.
  * $\mathcal{T}$**Wav** lorsque la base de Daubechies à 4 moments nuls est utilisée avec une matrice de covariance scalaire. Dans ce cas, une opération de seuillage (consistant à conserver 10% des coefficients ayant la norme la plus importante) est appliquée.
- Les résultats utilisant la distance entre curvelettes sont notés :
  * **Curv-diag** lorsque la transformée en curvelettes est utilisée en combinaison avec la matrice de pondération $\tilde{\mathbf{D}}_{\mathbf{C},var}$ définie en équation (5.63). Dans ce cas, l'ensemble des coefficients de curvelette est utilisé. Le vecteur d'observation est donc 7 fois plus grand dans cet espace que dans l'espace des pixels.
  * $\mathcal{T}$**Curv** lorsque la transformée en curvelettes est utilisée en combinaison avec une matrice proportionnelle à l'identité. Dans ce cas, une opération de seuillage (3.29) (consistant à conserver 10% des coeffcients ayant la norme la plus importante) est appliquée.
- Les résultats utilisant la distance dans l'espace de Fourier sont notées **Fourier-diag**. Dans ce cadre l'ensemble des coefficients est utilisé conjointement avec la variance estimée de ces derniers.

5.8.1 Influence du niveau de bruit

Expériences effectuées

But : Etudier l'influence du niveau de bruit sur l'analyse.

Observations utilisées : Une séquence de 24 images bruitées. Le niveau de bruit est homogène sur la séquence, corrélé en espace et décorrélé en temps. Trois niveaux de bruits sont considérés. Un exemple d'image bruitée pour chacun des niveaux est présenté en figure 5.23.

Termes d'écarts aux observations : Tous les termes décrits précédemment.

A savoir : Pour chaque niveau de bruit, 10 scénarios de bruits différents sont utilisés. Les résultats présentés dans les tableaux 5.1 à 5.4 sont les résultats moyens sur 10 expériences.

Autres paramètres : La longueur de corrélation d'erreur d'ébauche est de 7.65 pixels. La fenêtre d'assimilation est de 2 h 24 en équivalent atmosphère. Le traceur est observé toutes les 6 minutes.

Utilisation de matrices diagonales dans des espaces multi-échelles

Nous allons tout d'abord nous intéresser au changement de modélisation des erreurs d'observations induit par l'utilisation de matrices diagonales dans des espaces multi-échelles. Pour ces expériences, la totalité du signal présent dans les images est utilisé. La différence entre ces distances réside *uniquement* dans la description des statistiques d'erreur d'observation.

Le tableau 5.1 présente le pourcentage d'erreur restant dans la composante méridionale de la vitesse en fin d'analyse en fonction du niveau de bruit et de la distance considérée.

Table 5.1 – Moyenne sur 10 expériences du pourcentage d'erreur résiduelle dans la composante méridionale du champ de vitesse analysée, v_0^a, par rapport à l'erreur présente dans l'ébauche (v_0^b). Ce pourcentage est calculé pour différents niveaux de bruit.

SNR	Bruit fort 14.8 dB	20.8 dB	Bruit faible 26.8 dB
Curv-diag	7.9%	7.4%	7.9%
Daub-diag	8.3%	7.2%	6.9%
Haar-diag	16.5%	10.1%	7.3%
Pixels-scal	36.8%	21.8%	15.2%
Fourier-diag	7.1%	7%	7%

On observe que, contrairement aux autres distances, la distance entre pixels (associée à une matrice scalaire) n'est pas robuste à un bruit additif corrélé en espace.

Parmi ces dernières, la moins robuste est la distance dans la base d'ondelettes de Haar combinée à une matrice diagonale. Ce résultat n'est pas surprenant au vu des *corrélations induites* par l'utilisation d'une matrice diagonale avec cette représentation du signal (voir figure 5.14). Cependant, on observe qu'approcher (même assez mal) le modèle de corrélations peut avoir un effet bénéfique important : l'influence de l'amplitude du bruit sur l'analyse est fortement diminuée par rapport à la distance entre pixels.

Les distances utilisant, l'espace spectral, la trame de curvelettes ou la base d'ondelettes de Daubechies sont très résistantes à l'augmentation de l'amplitude du bruit. L'analyse dépend très peu du niveau de bruit corrompant les images. Ce dernier est bien mieux pris en compte qu'avec les autres distances et ce, même pour un faible niveau de bruit.

Remarque 5.31

Lorsque l'analyse s'effectue dans la base d'ondelettes de Daubechies (ou une trame de

curvelette, ou dans l'espace spectral), les résultats obtenus avec les séquences d'images les plus bruitées sont meilleurs que ceux utilisant les séquences d'images les moins bruitées associées à la distance pixel à pixel.

Bilan 5.5

Dans ces expériences, la différence d'informations fournie au processus d'assimilation de données réside uniquement dans les statistiques d'erreur d'observations prescrites. Au vu des résultats, on s'aperçoit qu'il est crucial d'utiliser cette information. Même des modélisations assez grossières de ces statistiques (en utilisant les ondelettes de Haar) induisent une différence d'analyse importante.

Gradients, gradients normalisés et pixels sous-échantillonnés

Idée : L'idée de cette section est de modifier l'espace d'observation de telle sorte que ces dernières respectent mieux les hypothèse sur les statistiques d'erreurs d'observation (ou que la violation de ces dernières soit moins grave), sans pour autant avoir à effectuer un travail de modélisation de matrice de covariance important.

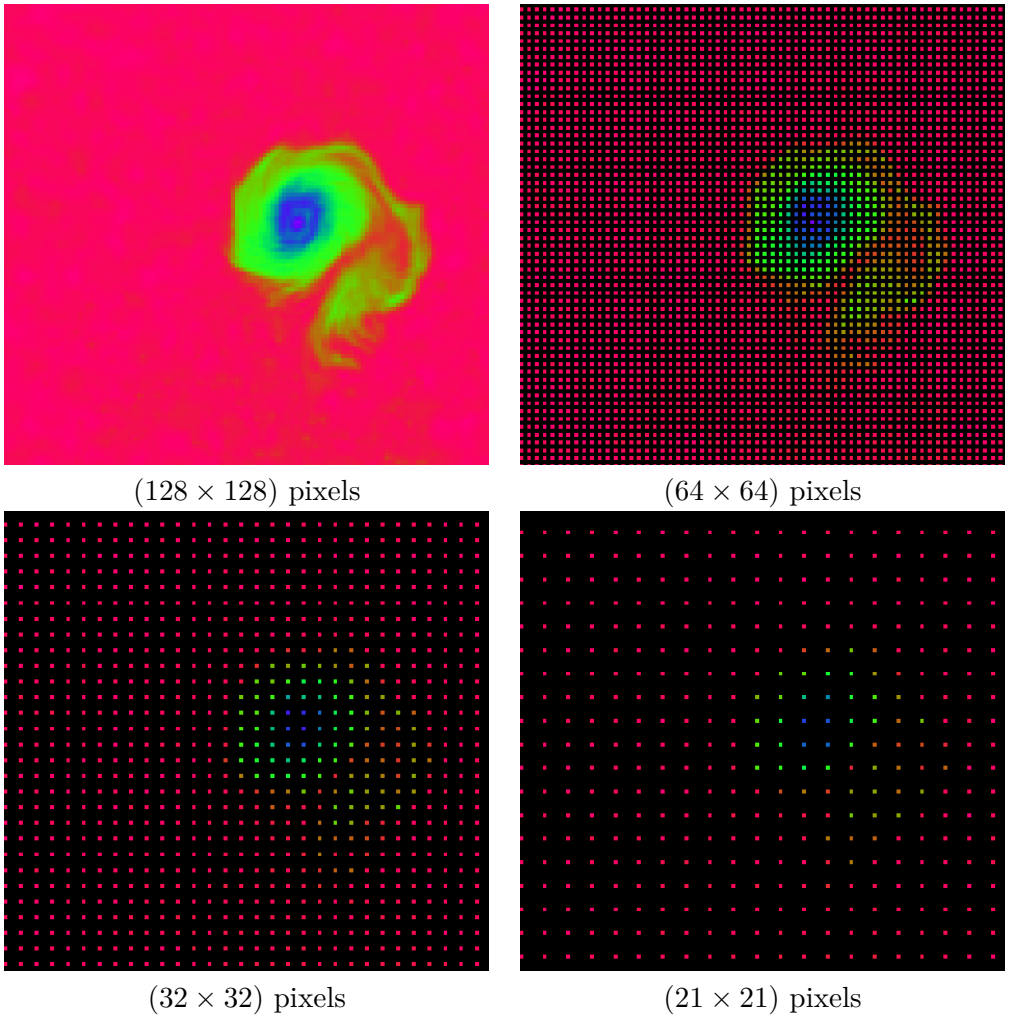


FIGURE 5.24 – Les données complètes sont présentées en haut à gauche. L'information utilisée lorsque 1 pixels sur 2 (resp. 4 et 6) est conservé dans chaque direction est présenté en haut à droite (resp. en bas à gauche et à droite). Les données manquantes sont présentées en noir.

Mise en œuvre : Dans l'espace des pixels, cette idée est souvent utilisée en *sous-échantillonnant*

les données (voir par ex. [Liu and Rabier, 2002]) et/ou en créant des *"super-observations"*. Concernant les premières, l'idée est de conserver uniquement dans le vecteur d'observation des pixels dont les erreurs sont décorrélées (ou presque décorrélées) entre elles. Cela conduit à réduire drastiquement la taille du vecteur d'observation (mais permet en général d'accéder à une meilleur analyse). Dans le tableau 5.2 les résultats sont présentés lorsqu'un pixel sur deux dans chaque direction est conservé (sous le nom de **Pixels-Thin2-scal**) et lorsqu'un pixel sur trois est conservé (**Pixels-Thin3-scal**). Des exemples de sous-échantillonnages sont présentés en figure 5.24. On observe que moins il y a de données, moins la structure du vortex est observée avec précision.

Remarque 5.32

Avoir accès à une séquence trop pauvre en information ne permet que difficilement d'évaluer "visuellement" le champ de vitesses transportant le traceur.

Dans l'espace des gradients, aucun sous-échantillonnage spatial n'est effectué. Le fait que les erreurs soient fortement corrélées en espace implique une réduction de la variance importante dans l'espace des gradients comme présenté en équation (5.2). L'information conservé après la transformation est donc très peu bruité par rapport à l'information initiale contenue dans l'espace des pixels (bien qu'elle soit toujours corrélé en espace).

Résultats : Le tableau 5.2 présente les résultats obtenus en utilisant la comparaison dans l'espace des gradients (avec une matrice scalaire), des gradients normalisés et dans différents sous espaces des pixels.

Table 5.2 – Moyenne sur 10 expériences du pourcentage d'erreur résiduelle dans la composante méridionale du champ de vitesses analysée, v_0^a, par rapport à l'erreur présente dans ce champ à l'ébauche (v_0^b). Ce pourcentage est calculé pour différents niveaux de bruit.

SNR	Bruit fort 14.8 dB	20.8 dB	Bruit faible 26.8 dB
∇-scal	20.2%	13.5%	11.1%
Pixels-scal	36.8%	21.8%	15.2%
Pixels-Thin2-scal	28.8%	19.1%	13.2%
Pixels-Thin3-scal	46.5%	22.9%	15.4%
Normale-diag	53.3%	36.9%	24.4%

- On observe que, comme pour les données non corrélées en espace, l'utilisation des gradients normalisés induit de bien moins bons résultats que les autres (pseudo)métriques. Lorsque les images sont bruitées, utiliser uniquement l'information sur la direction des gradients n'est pas suffisant pour contrôler le champ de vitesse initial. Cette information est en effet trop pauvre et trop bruitée dans les zones lisses de l'image.
- Le sous échantillonnage des données a ici un impact assez faible. Certes les résultats sont meilleurs lorsqu'un pixel sur deux est conservé dans chaque direction (et ce bien que les pixels voisins soient toujours fortement corrélés). Cependant lorsque moins de pixels sont conservés (par ex. un sur trois) les résultats se détériorent. Cette augmentation est encore plus conséquente quand la distance d'échantillonnage augmente (résultats non présentés dans ce document). Ces résultats atypiques du sous-échantillonnage sont certainement en partie imputable à la taille du phénomène observé (le tourbillon a une taille d'environ 20 x 20 pixels) et au fait que les observations utilisées soient des observations indirectes de la variable contrôlée (observation d'un traceur passif pour le contrôle du champ de vitesses).

- Les résultats obtenus en utilisant la comparaison entre gradients sont bien meilleurs. Ceci provient du fait que considérer uniquement l'information sur les gradients de l'image permet de filtrer une partie de ce bruit lorsque ce dernier est lisse (voir remarque 5.2 et figure 5.4). En utilisant seulement une partie du signal, on se retrouve donc à travailler avec une information, certes moins riche, mais surtout beaucoup moins bruitée. Pour le bruit corrompant ces séquences d'images, la perte d'information est largement compensée par la faible amplitude du bruit résiduel dans l'espace des gradients.

Bilan 5.6

La comparaison entre gradients est donc une manière simple (mais efficace) de mieux extraire les informations importantes dans les images lorsque le bruit est lisse. Dans ce cas, l'amélioration ne provient pas d'une meilleure description des statistiques du bruit (comme dans le cas des ondelettes ou des curvelettes) mais de la suppression d'une partie importante du bruit. Cette réduction découle du travail dans un espace moins riche en information. Cependant Travailler dans un espace trop pauvre peut ne pas permettre d'extraire suffisamment d'information des observations pour contrôler le système. C'est ce qui se passe lorsqu'on travail dans l'espace des gradients normalisés lorsque l'image d'origine est lisse. Les résultats obtenus seraient sans doute meilleurs si la séquence d'image ne présentait pas de zone lisse.

Dispersion des expériences

Jusqu'à présent, nous avons simplement parlé de résultat moyen. Ces résultats sont intéressants dans le sens où ils permettent de dire, en moyenne, l'impact d'un terme d'écart aux observations. Cependant, la dispersion des résultats lors des différents scénarios est aussi une information intéressante. En effet, cette dispersion permet d'évaluer la sensibilité de l'analyse à une réalisation particulière du bruit. Bien que nous ayons effectué trop peu de scénarios différents pour calculer des diagnostics de dispersion, nous pouvons tout au moins représenter l'ensemble des résultats obtenus sous forme de boîte à moustache.

Cette représentation est employée en figure 5.25 pour différents termes d'écarts aux observations en fonction du niveau de bruit (jaune pour le niveau le plus élevé et bleu pour le moins élevé).

On observe, comme attendu, que l'analyse effectuée avec des séquences peu bruitées dépend moins du tirage du bruit. Pour les curvelettes, et dans une moindre mesure pour les ondelettes de Daubechies et pour les gradients normalisés, la dispersion des résultats ne varie pas lors de l'augmentation du niveau de bruit.
A la vue du tableau 5.1, ce résultat est assez peu surprenant pour les curvelettes et les ondelettes de Daubechies. En effet en employant une approximation tenant compte des corrélations présentes dans le champs d'erreurs l'impact de ces dernières est fortement réduit. Une conséquence logique en ait une moindre dispersion des résultats.

Par contre, c'est un peu inattendu concernant les gradients normalisés. Dans ce cas, les résultats sont moins bons lorsque le bruit augmente mais toujours aussi peu dispersés.

Les résultats dans l'espace des pixels, dans l'espace des gradients ou avec les bases d'ondelettes de Haar sont plus dépendants du tirage lorsque le niveau de bruit augmente. Ceci résulte d'un plus fort impact du bruit dans la fonction coût lorsque le terme d'écart aux observations est considéré dans ces espaces. Il en résulte que certaines expériences se passent "relativement" bien et d'autre "relativement" mal selon le tirage de ce dernier.

Capacité de compression des transformées multi-échelles

Les résultats obtenus dans l'espace des pixels, dans un sous-espace d'ondelettes et de curvelettes, sont regroupés dans le tableau 5.3. La différence entre les (pseudo)métriques utilisées réside dans

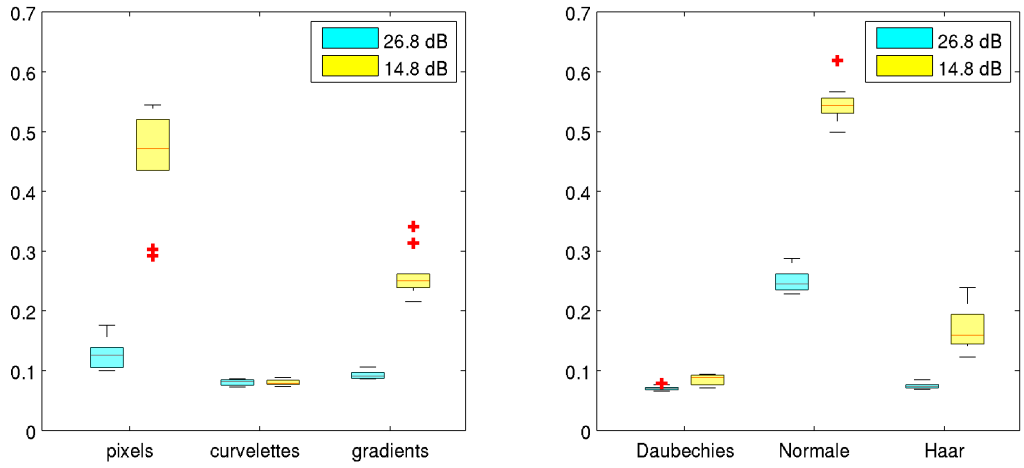


FIGURE 5.25 – Boites à moustaches présentant la dispersion du pourcentage d'erreur sur la composante méridionale du champ de vitesses. Les 10 expériences effectuées avec les séquences peu bruitées ($SNR = 26.8$ dB) sont présentées en bleu, celles avec un bruit important ($SNR = 14.8$ dB) en jaune.

la quantité de signal fourni au processus d'assimilation de données.

Remarque 5.33

Dans ce cas, les statistiques d'erreurs d'observations prescrites ($\tilde{\mathbf{R}}_{\mathbf{W},} = \sigma^2\mathbf{I_n} = \tilde{\mathbf{R}}_{pix,}$; $\tilde{\mathbf{R}}_{\mathbf{C},} = \sigma^2\mathbf{I_m}$) conduisent à minimiser la même fonction coût lorsque l'ensemble de l'information est employée dans chacun des espaces.

On choisit pour ces expériences d'assimiler 10% des coefficients d'ondelettes ou de curvelettes. Les coefficients sélectionnés sont ceux dont la norme est la plus importante. La taille du vecteur d'observations est donc divisée par dix dans la base d'ondelettes et par 1.5 seulement dans la trame de curvelettes (à cause de la redondance d'information).

TABLE 5.3 – Erreur restant sur la composante méridionale du champ de vitesses en fin d'assimilation en fonction du niveau de bruit

SNR	Bruit fort 14.8 dB	20.8 dB	Bruit faible 26.8 dB
$\mathcal{T}$**Curv** (10%)	61.4%	28.4%	16.3%
$\mathcal{T}$**Wav** (10%)	55.4%	29.2%	16.6%
Pixels-scal	36.8%	21.8%	15.2%

On observe que, comme pour la distance entre pixels, les pseudo-métriques dans des bases multi-échelles (combinées avec une matrice de covariance proportionnelle à l'identité) ne sont pas résistantes à un bruit corrélé en espace.

Les résultats obtenus en assimilant seulement une partie de l'information ($\mathcal{T}$**Curv** et $\mathcal{T}$**Wav**) sont moins bons que lorsque l'ensemble de l'information (**Pixels-scal**) est utilisée.

Cependant, si pour des raisons de taille compresser les données est nécessaire, la compression dans une base d'ondelettes semble être une meilleure option. Dans ce cadre, moins de coefficients sont utilisés et les résultats sont "assez" proches des résultats obtenus dans l'espace des pixels.

Remarque 5.34

On observe que plus la séquence d'images est bruitée, plus l'effet de la compression est néfaste. Ceci provient du fait que les coefficients de petite échelle, peu perturbés par un bruit lisse, sont moins conservés par le seuillage lorsque le niveau de bruit augmente. En effet, leur norme dans l'image originale est souvent bien plus faible que celle des coefficients à une échelle plus grossière. Cette différence s'accentue avec la présence d'un bruit corrélé en espace (voir figure 5.6). Le seuillage effecuté se basant uniquement sur la norme du coefficient, les coefficients de détail aux échelles les plus fines sont donc moins conservés lorsque le niveau du bruit augmente.
Lors de l'utilisation d'une matrice diagonale dans la base d'ondelettes de Daubechies (conduisant aux résultats ***Daub-diag*** *du tableau 5.1), les coefficients de détail à l'échelle la plus fine ont une variance* 10 000 *fois moins élevée que les coefficients à l'échelle grossière.*
Par conséquent, dans les expériences compressant l'information, au lieu d'être assimilé avec un poids bien plus fort que les coefficients des échelles plus grossières le poids de ces coefficients de petite échelle est réduit à 0.
En effectuant ce seuillage, on supprime donc les marqueurs les plus fiables des discontinuités plutôt que de les assimiler avec un poids important.

Utilisation de l'information sur les statistiques d'erreurs lors de la compression

La connaissance de la variance de chaque coefficient dans un espace multi-échelles peut être utilisée lors de la phase de compression. Deux manières d'utiliser cette information peuvent être envisagées :

- La méthode de sélection des coefficients n'est pas modifée. Seul leurs poids respectifs dans la fonction coût l'est. Dans la suite, cette compression est nommée **global** ;
- La méthode de sélection des coefficients est modifiée. Par la suite, cette compression est nommée **par échelle** ;

Compression via l'utilisation d'un seuillage global :
La compression peut se faire de manière "habituelle". Afin de représenter au mieux les images observées, les coefficients d'ondelettes ou de curvelettes ayant la norme la plus importante sont conservés. L'utilisation de ces coefficients nous permet de reconstruire les images observées en dégradant le moins possible l'information (aussi bien en terme de signal que de bruit) présente au sein de ces images.

Cet opération de compression est basée sur le seuillage suivant :

$$d^j[k] = \begin{cases} d^j[k] & \text{si } |d^j[k]| > \tau \\ 0 & \text{sinon} \end{cases} \tag{5.68}$$

où τ est fixé de manière à ne conserver qu'un certain pourcentage des coefficients. C'est ce seuillage qui a été appliqué pour les pseudo-métriques du tableau 5.3.

Compression en changeant la méthode de sélection :
L'autre idée consiste, maintenant que nous connaissons la variance, à utiliser l'information sur cette dernière.

Une première idée (qui ne fonctionne pas) consiste à définir le seuillage sur lequel s'appuie la phase de compression par :

$$d^j[k] = \begin{cases} d^j[k] & \text{si } |d^j[k]| > \dfrac{\tau}{\sigma^j[k]} \\ 0 & \text{sinon} \end{cases} \tag{5.69}$$

De fait, cette première idée peut paraître tentante. Cependant, pour le bruit considéré, étant donné la différence de variance entre les coefficients cette méthode conduit à conserver

uniquement les coefficients à petite échelle. Ces coefficients sont certes fiables, mais ils portent très peu d'information (pour s'en convaincre voir la figure 2.11). L'utilisation de ces coefficients de manière seul ne permet pas de fournir suffisament d'information sur le processus physique ayant lieu.

Remarque 5.35

D'ailleurs, comme nous le verrons en section 5.8.4, ces coefficients de petites échelles peuvent aussi être dérangeants lors de la phase de minimisation.

Des coefficients aux échelles plus grossières doivent donc être conservés pour pouvoir avoir une idée du mouvement global ayant lieu. Afin de prendre en compte le fait que les coefficients à petites échelles sont plus fiables que les coefficients aux échelles plus grossière mais comportent bien moins d'information, nous utilisons le seuillage échelle par échelle proposé dans [Souopgui, 2010]. Ce seuillage consiste à conserver la même proportion de coefficients à chaque échelle. On force ainsi lors du processus de compression à conserver des coefficients aux petites échelles (ce que ne fait pas (5.68)) et des coefficients aux échelles plus grossières (ce que ne fait pas (5.69)). Il s'écrit

$$d^j[k] = \begin{cases} d^j[k] & \text{si } |d^j[k]| > \tau_j \\ 0 & \text{sinon} \end{cases} \tag{5.70}$$

Ici le seuil en dessus duquel les coefficients sont conservés dépend de l'échelle considérée. Dans la suite on considère qu'on souhaite conserver la même *proportion* (et non le même nombre) de coefficients à chaque échelle. Par exemple, pour le seuillage dans une base d'ondelettes pour une image de taille 128×128 si on souhaite diviser par 10 la taille du vecteur d'information, on conserve environ 1200 coefficients à la plus petite échelle, 300 coefficients à une échelle plus grossière et ainsi de suite.

Le tableau 5.4 présente les résultats obtenus lorsque la compression est effectuée dans une base d'ondelettes et dans une trame de curvelettes en tenant compte de la variance de chacun des éléments. Le seuillage global correspond à celui utilisé jusqu'à présent (rappelé en équation (5.68)). Les résultats issus de ce seuillage peuvent donc être mis en lien avec ceux présentés au tableau 5.3. On observe que l'utilisation des véritables variances lorsque ce seuillage est effectué permet une amélioration des résultats. Cependant cette amélioration est limitée lorsque le bruit à une amplitude importante. Pour contrebalancer ce fait on peut décider d'utiliser plus de coefficients (pour les ondelettes) où changer la méthode de sélection des coefficients (dans les deux espaces).

Table 5.4 – Erreur restant sur la composante méridionale du champ de vitesses en fin d'assimilation en fonction du niveau de bruit

	Seuillage	Bruit fort 14.8 dB	20.8 dB	Bruit faible 26.8 dB
Curvelette	global (10%)	43.4%	14%	9.5%
	par echelle (10%)	9.8%	8.1%	7.8%
Ondelette	global (20%)	20.3%	11%	9.1%
	par échelle (20%)	13.7%	9.1%	8.4%
Ondelette	global (10%)	46%	17%	13.3%
	par échelle (10%)	16.9%	10%	9%

Le seuillage par échelle (correspondant à celui introduit en équation (5.70)) permet de pallier à ce problème. Le fait de forcer, lors de la phase de compression, à conserver des coefficients de petite échelle permet une amélioration importante de l'analyse. Pour les curvelettes, les résultats finaux en utilisant 10% des coefficients sont à peine moins bons que ceux utilisant l'ensemble des coefficients 5.1. Lorsque l'analyse est effectuée dans une base d'ondelettes, les résutlats de

compression sont eux aussi bons. Cependant, ils sont un peu plus sensibles à l'augmentation de l'amplitude du bruit.

Bilan 5.7

Les performances obtenues lorsque les véritables variables variances sont utilisées permettent d'envisager une compression de l'information sur un faible jeu de coefficients. Cependant cette compression est plus complexe à faire lorsque le bruit est corrélé en espace : dans le cas étudié nous avons en effet le choix entre conserver des coefficients porteur de beaucoup d'information mais fortement bruités et des coefficients fiable mais porteur de peu d'information. Choisir de conserver ces deux types de coefficients permet d'améliorer fortement les performances de l'analyse.

Répartition spatiale des erreurs

5.8.2 Impact de R sur le choix de B

Dans cette section, nous nous intéressons aux interactions entre la matrice de covariance d'erreur d'ébauche et la matrice de covariance d'erreur d'observation.

Mémo 12 – Fonction coût

Pour rappel, la fonction coût utilisée est écrite de la manière suivante :

$$J(\boldsymbol{X}_0) = \alpha_b \underbrace{\|\boldsymbol{X}_0 - \boldsymbol{X}^b\|^2_{\mathbf{B}}}_{J^b} + \underbrace{\sum_{i=0}^{N} \|\mathbf{y_i^o} - \mathbf{H}_i(\boldsymbol{X}_i)\|^2_{\mathbf{R}_k}}_{J^o}$$

où α_b est un paramètre permettant de changer l'influence du terme d'écart à l'ébauche.

Description des expériences

But : Etudier l'importance du choix de la matrice d'erreur d'ébauche (et du poids accordé à cette dernière, i-e α_b) en fonction des statistiques d'erreurs d'observations prescrites.

Observations utilisées : Une séquence de 24 images bruitées. Le niveau de bruit est homogène sur la séquence. Les erreurs sont corrélées en espace et décorrélées en temps. Le rapport signal bruit est de 14.8 dB. Il correspond à un des scénarios utilisant le bruit de forte amplitude (voir figure 5.23).

Termes d'écarts aux observations étudiés : Les termes d'écarts aux observations utilisées sont **Pixels-scal**, **∇-scal**, **Daub-diag** et **Curv-diag**. Les termes **Pixels-scal**, **Daub-diag** et **Curv-diag** assimilent la même information en terme de signal. Seule la représentation des statistiques d'erreur change.

Matrice de covariance d'erreur d'ébauche :

- Pour chaque terme d'écart aux observations, différentes longueur de corrélation d'erreur d'ébauche sont testées.
- Ensuite, pour une longueur de corrélation de 6.25 pixels, différents termes de pondération α_b de l'écart à l'ébauche sont utilisés.

A noter : Contrairement à la partie précédente, une seule expérience est effectuée pour chaque couple (Ecart aux observations, Matrice **B**). Les résultats sont donc plus sensibles à l'influence du bruit que ceux présentés dans les tableaux 5.1, à 5.3 (comme illustré en figure 5.25).

La figure 5.26 présente le pourcentage d'erreur restant dans la vorticité analysée en fonction de la longueur de corrélation (de l'erreur d'ébauche).

Le comportement de l'erreur d'analyse diffère fortement lorque le terme d'écart aux observations prend en compte les corrélations présentes au sein du bruit (**Daub-diag** et **Curv-diag**) ou non (**Pixels-scal** et **∇-scal**).

Lorsque les corrélations au sein du bruit ne sont pas pris en compte (**Pixels-scal** et **∇-scal**), on observe qu'augmenter la longueur de corrélation de l'erreur d'ébauche (figure 5.26) permet d'obtenir de meilleurs résultats. La définition d'une bonne matrice **B** semble donc ici être celle qui contraint l'incrément à être lisse. Afin de compenser la mauvaise spécification de la matrice d'erreur d'observation il est ici nécessaire de mal spécifier la matrice de covariance d'erreur d'ébauche **B**.

A contrario, pour les ondelettes et les curvelettes nous observons le même type de comportement que dans le cadre où les observations étaient parfaites :

- il y a tout d'abord une phase de décroissance rapide des erreurs d'analyse lorsque la longueur de corrélation augmente,
- suivis d'une phase durant laquelle augmenter la diffusion a assez peu d'effet,
- puis, lorsque le nombre d'itérations de diffusion devient trop important, il y a une phase de croissance des erreurs.

Dans ce cas, nous ne sommes pas obligés de contraindre l'incrément à être lisse pour obtenir de bon résultats sur la vorticité analysée. Comme le présente cette figure, trop contraindre l'espace de recherche entraîne de moins bons résultats.

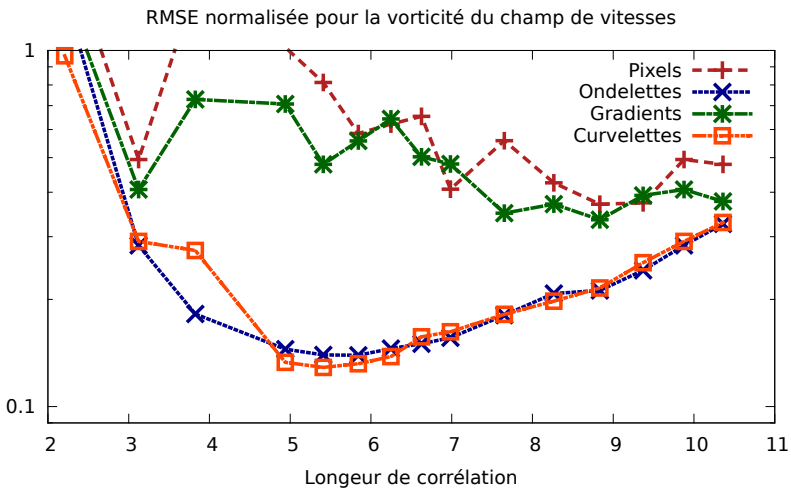


FIGURE 5.26 – Pourcentage d'erreur restant (en terme de RMSE) dans le champ de vorticité en fin d'assimilation en fonction de la longueur de l'erreur d'ébauche.

La figure 5.27 présente le pourcentage d'erreur d'analyse restant en fin d'assimilation en fonction du poids accordé à l'ébauche.

La forme de la courbe change grandement selon la transformation (Gradient, Pixels, Ondelettes, Curvelettes) utilisée.

Lorsque l'hypothèse (de matrice de covariance) diagonale est effectuée dans une base d'ondelettes ou une famille curvelettes, on distingue trois phases principales :

- Lorsque le paramètre α_b est trop important, l'analyse effectuée ne tient que peu compte des observations (à droite sur la courbe). Ceci est attendu et peut être expliqué par le fait que $J^b \gg J^o$ ce qui rend les variations du premier terme bien plus "pénalisant" que les variations du second. A partir d'une certaine valeur de α_b, le processus d'assimilation n'améliore donc que très peu notre connaissance du véritable état analyse.

– Puis, lorsque la valeur du paramètre α_b diminue, on observe une décroissance rapide de l'erreur d'analyse (en fonction de la valeur α_b choisie).
– Enfin, on observe que la valeur de α_b n'a guère d'importance tant que cette dernière n'est pas trop importante : par exemple l'erreur effectuée est identique que α_b vaille 1 ou 10^8.

Lorsque l'hypothèse diagonale est effectuée dans l'espace des pixels ou dans l'espace des gradients, la forme de la courbe de la figure 5.27 varie un peu. Comme précédemment, cette dernière comporte bien les deux premières phases (stagnation si α_b est trop grand puis décroissance rapide lorsque α_b approche de sa valeur optimale). Cependant, une fois cette valeure optimale atteinte, l'erreur d'analyse croit rapidement (quand α_b diminue). Ceci provient du fait qu'une partie de l'erreur d'observation n'est pas considérée comme telle au sein de la fonction coût. Il en résulte que l'analyse incorpore en son sein une partie importante de l'erreur d'observation.

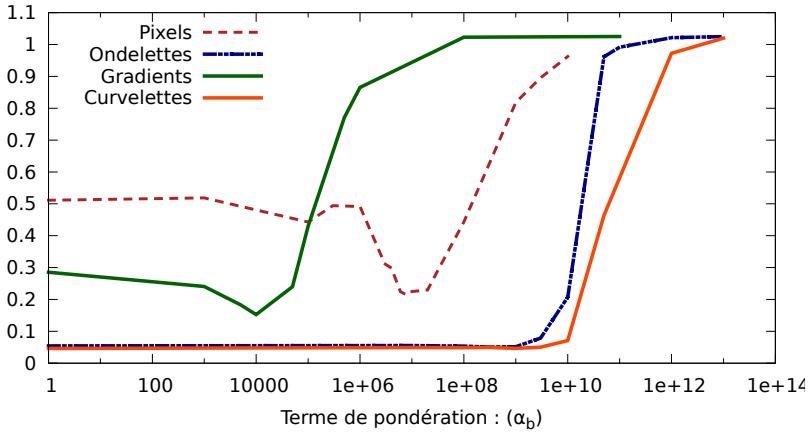


FIGURE 5.27 – Pourcentage d'erreur restant (en terme de RMSE) dans le champ de vitesse zonale en fin d'assimilation en fonction du poids accordé à l'ébauche (α_b) dans la fonction coût (3.42) .

Remarque 5.36

La valeur optimale de α_b varie grandement selon l'espace considéré : cette valeur va de 10^4 dans l'espace des gradients à 10^9 dans la famille de curvelettes. Cette différence peut s'expliquer de deux manières. Dans l'espace des gradients, une partie de l'information (ainsi que du bruit) est filtré (par la transformée). Dès lors, l'écart aux observations est bien moins important dans cet espace que dans les autres.
Concernant la différence pixels, ondelettes et curvelettes, le signal incorporé est identique. Le changement provient des corrélations au sein des erreurs d'observations prises en compte. L'incorporation des corrélations revient à considérer que certains éléments (dans les bases d'ondelettes ou de curvelettes) sont bien moins perturbés (ou presque parfait) que d'autres par un bruit corrélé en espace. Dès lors, lorsqu'un problème de décalage "observations" "équivalent modèle" apparait, le poids du décalage entre $H(\boldsymbol{X}_i)$ et y_i^o au sein de la fonction coût peu exploser dans ces espaces. Cela conduit notamment au fait que $J^o(\boldsymbol{X^b})_{ondelettes} \gg J^o(\boldsymbol{X^b})_{pixels}$.

Bilan 5.8

Le rôle de la matrice de covariance d'erreur d'ébauche est important dans ces expériences. Ce rôle est différent selon les statistiques d'erreurs d'observation utilisées. Lorsque ces dernières sont mal spécifiées, l'augmentation de la longueur de corrélations d'erreur d'ébauche permet de diminuer l'erreur d'analyse en recherchant l'incrément dans un espace plus lisse. Cet

effet n'est plus désiré lorsque les corrélations entre erreurs d'observations sont mieux prises en compte. Dans ce cas, augmenter artificiellement la longueur de corrélation détériore fortement l'analyse.

5.8.3 Quelle quantité d'information est contenue dans les statistiques d'erreurs d'observation ?

Il n'y a bien évidemment pas de réponse générale à cette question. En plus de dépendre des observations utilisées, la réponse à cette question est liée à l'amplitude du bruit et à la forme des corrélations sus-citées.

Cependant, afin d'illustrer à quel point la connaissance sur ces statistiques peut apporter une information importante un nouveau jeu d'expériences jumelles est réalisé.

Expériences effectuées

But des expériences : Quelle est la quantité d'images supplémentaire nécessaire pour remplacer l'information contenue dans les statistiques d'erreurs d'observations ?

Termes d'écarts aux observations : **Pixels-scal**, **∇-scal**, **Daub-diag** et **Curv-diag**. Les termes **Pixels-scal**, **Daub-diag** et **Curv-diag** assimilent la même information en terme de signal. Seule la représentation des statistiques d'erreurs change.

Observations utilisées : Une séquence d'images bruitées. Le bruit est corrélé en espace et décorrélé en temps. Le rapport signal bruit de la séquence d'images est 14.8 Db.

Fréquence d'observation : La fréquence d'observation du vortex varie selon l'expérience.

Autre paramètres : La fenêtre d'assimilation est ici de 3 h. La longueur de corrélation est de 7.65 pixels.

La figure 5.28 présente, en fonction du nombre d'observations utilisées, le pourcentage d'erreur restant dans la composante méridionnale du champ de vitesses.

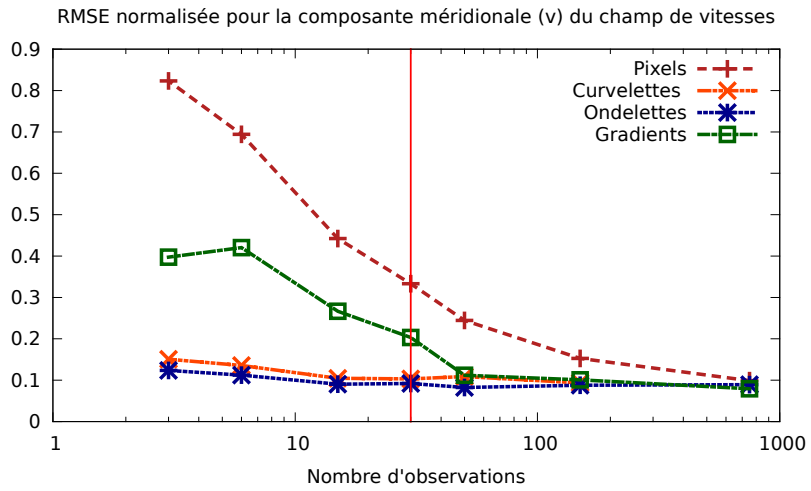


FIGURE 5.28 – Pourcentage d'erreur restant sur la composante méridionale du champ de vitesses en fonction du nombre d'observations utilisées pour des expériences de 3 h. La ligne rouge correspond à la fréquence d'acquisition utilisée dans les expériences des sections précédentes.

On observe un comportement bien différent selon si une partie des connaissances sur les statistiques est intégrée (**Curv-diag** et **Daub-diag**) ou si cette information n'est pas ajoutée au processus de minimisation (**Pixels-scal** et **∇-scal**).

Dans les deux cas, lorsque le système est très bien observé (une observation à chaque pas de temps), les résultats obtenus sont bons pour l'ensemble des termes d'écart aux observations.

Cependant, lorsque le système est mal observé, les distances utilisant l'information sur les statistiques d'erreurs obtiennent des résultats bien meilleurs que les autres. Ainsi, utiliser 3 images avec la distance entre ondelettes permet d'éliminer 88% de l'erreur présente dans l'ébauche. Lorsqu'une image est observée à chaque pas de temps, 91% de l'erreur est supprimée. L'apport supplémentaire dû à chaque image supplémentaire utilisée est donc très faible lors de l'utilisation de cette distance car l'essentiel de l'information sur le mouvement a déjà été extraite avec peu d'images.

Pour obtenir une réduction d'environ 90%, il faut, dans l'espace des pixels, utiliser une observation à chaque pas de temps (soit 750 images) ce qui correspond à quelques centaines de fois plus d'observations que dans une base d'ondelettes.

Remarque 5.37

Ces bons résultats obtenus lorsqu'une image est utilisée à chaque pas de temps sont en partie dus à l'incohérence temporelle du champ d'erreur. Dans des applications réalistes, en plus d'avoir accès à une information disponible de manière moins régulière, le champ d'erreur est très certainement corrélé temporellement.

Pour le terme d'écart entre gradients, le bruit étant fortement diminué grâce à l'appauvrissement de l'information, les résultats sont comparables à ceux obtenus dans une base d'ondelettes ou de curvelettes dès lors qu'au moins 50 observations sont utilisées. La suppression d'une partie de l'information (incluant, entre autre, une large part du bruit) permet donc d'obtenir de bien meilleurs résultats sans tenir compte des statistiques d'erreurs d'observations.

Bilan 5.9

Pour ces séquences d'observations bruitées, utiliser des approximations des statistiques d'erreurs prenant en compte une partie des corrélations spaciale diminue fortement le nombre d'observations nécessaires au contrôle du système. L'information présente dans les observations est dans ce cas bien mieux exploitée.

5.8.4 Attention danger : statistiques à manipuler avec précaution !

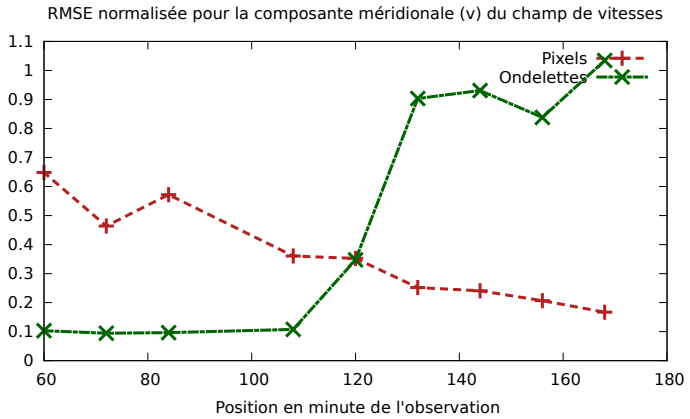


FIGURE 5.29 – Pourcentage d'erreur restant sur la composante méridionale du champ de vitesse en fonction de la différence de temps entre l'ébauche et la première observation.

Dans les paragraphes précédents, l'importance de bien connaître les statistiques d'erreurs a été illustrée sur plusieurs exemples. Cependant, connaître et utiliser ces statistiques ne constitue pas toujours un atout pour produire une bonne analyse.

Description des expériences

But : Déterminer si l'incohérence temporelle du bruit joue un rôle important dans la réussite des expériences précédentes.

Observations utilisées : Une seule image bruitée. Le bruit est corrélé en espace. L'amplitude du bruit est faible (SNR = 26.8 dB).

Temps d'observation : Le temps écoulé entre t_0 et le temps d'observation t_i varie d'une expérience à l'autre.

Terme d'écart aux observations : Pixels-scal et **Daub-diag**. Un troisième terme d'écart aux observations (non décrit précédemment) fait son apparition au cours du paragraphe.

Autres paramètres : La longueur de corrélation d'erreur d'ébauche est de 7.65 pixels.

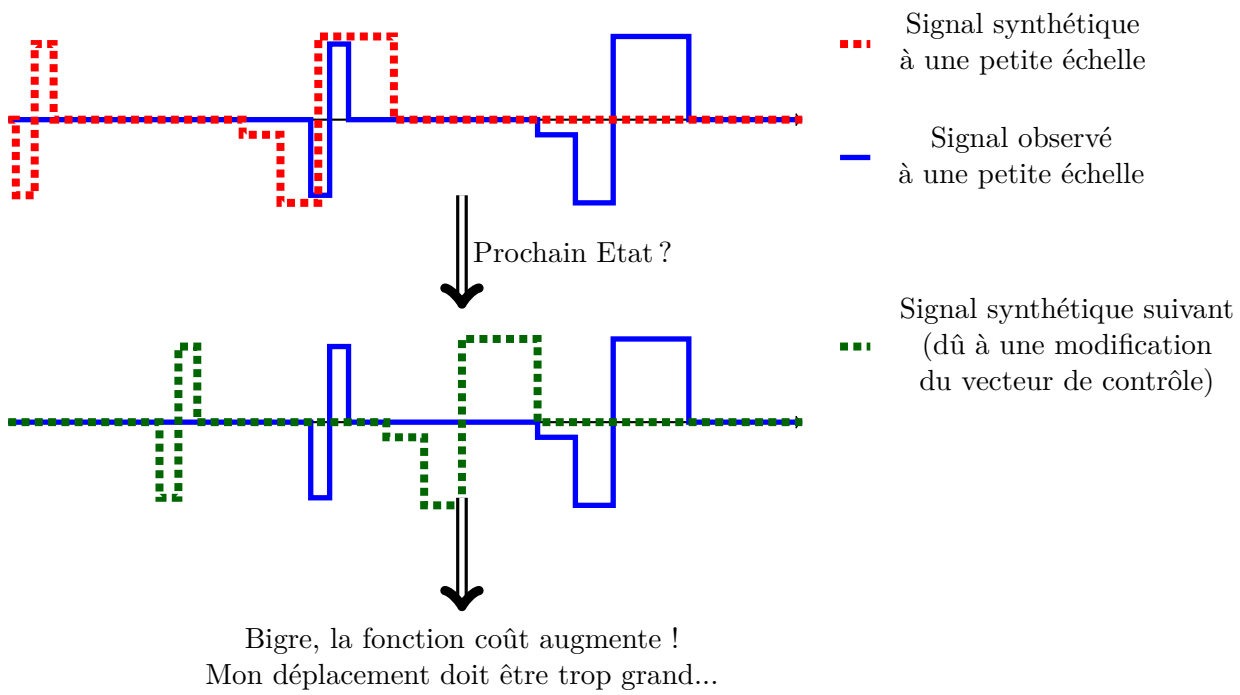


FIGURE 5.30 – Ilustration du problème rencontré pour le signal à petite échelle.

La figure 5.29 présente le pourcentage d'erreur résiduelle sur la composante méridionale du champ de vitesses en fonction du temps d'observation.

On observe que, dans l'espace des pixels, plus l'observation est éloignée du temps initial, meilleure est l'analyse. Par conséquent, accéder à une information éloignée en temps (et assez peu perturbée par le bruit) peut conduire à obtenir de meilleurs résultats lorsque les non-linéarités du modèle s'expriment mais restent assez faible.
La courbe obtenue en utilisant la distance dans la base d'ondelettes de Daubechies est assez surprenante. Lorsque l'observation est assez proche du temps initial, tout se passe bien : les résultats analysés sont de l'ordre de ceux attendus (il reste environ 10% d'erreur sur la composante méridionale du champ de vitesse). Ces premiers résultats sont donc en accord avec les expériences précédentes : en utilisant la même information (en terme de signal) que dans l'espace des pixels, l'utilisation des statistiques d'erreurs sur les coefficients d'ondelettes permet d'améliorer grandement l'analyse. Ici, l'influence d'une meilleure prise en compte des statistiques

d'erreurs d'observations est positive.
Cependant, lorsque le temps d'observation s'écarte trop du temps initial, on observe qu'utiliser cette distance ne permet plus d'atteindre un état analysé de qualité acceptable. L'amélioration par rapport à l'état initial est dans ce cas très faible voir inexistante.

Origine du problème :

Les résultats dans l'espace des pixels montrent qu'il y a assez d'information pour contrôler le système. Le problème ne vient donc pas du manque d'information. Pour identifier la source du problème, nous pouvons nous servir du fait que la distance utilisée dans l'espace des pixels est identique à celle dans une base d'ondelettes combinée à l'utilisation de la matrice identité. Ainsi, en examinant le poids accordé aux coefficients à différentes échelles, nous pouvons comprendre pourquoi l'algorithme reste bloqué dans un minimum local.
Dans notre cas, le problème vient du fait que les coefficients de petite échelle peu perturbés par le bruit, sont assimilés avec une très forte confiance.

Remarque 5.38

Pour le bruit et la base considérés, la variance des coefficients diagonaux à l'échelle la plus fine est environ 10 000 *fois plus faible que la variance des coefficients d'ondelettes à une échelle plus grossière.*

Lorsque l'écart entre les observations et l'équivalent modèle n'est pas dû au bruit, assimiler avec un trop fort poids ces coefficients bloque la minimisation. La figure 5.30 illustre le phénomène se déroulant pour le signal vivant à petite échelle.

Comment éviter ce problème ? Pour contourner ce problème, on peut employer l'aspect

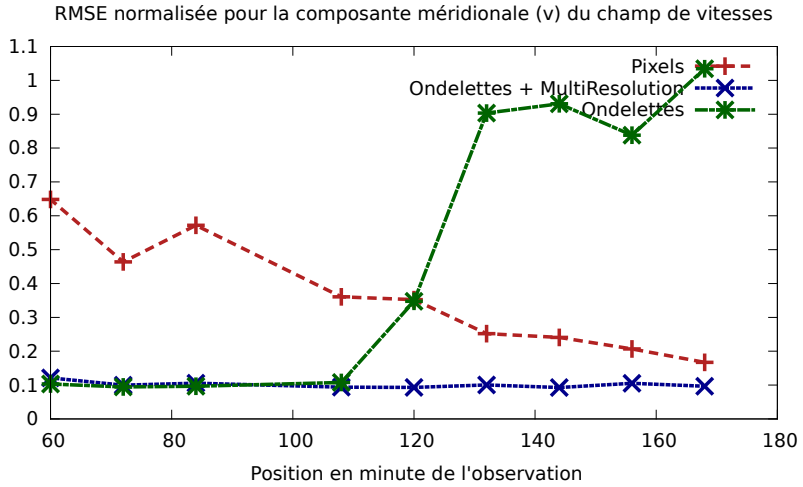


FIGURE 5.31 – Pourcentage d'erreur restant sur la composante méridionale du champ de vitesse en fonction de la différence de temps entre l'ébauche et la première observation. Le bruit d'observation est faible.

multi-échelle de notre transformée. L'idée est ici de recaler au préalable l'information à une échelle grossière puis, lorsque ce recalage est effectué, d'incorporer l'information contenue dans les détails des échelles plus fines. Pour faire ceci dans une base d'ondelette, il suffit en effet de seuiller les coefficients aux échelles les plus fines et de travailler dans un premier temps uniquement avec les coefficients des échelles les plus grossières.
Les expériences précédentes ont été réalisées en utilisant cette nouvelle méthode.

- Une première phase de minimisation (25 itérations) a été réalisée en ne prenant pas en compte les coefficients d'ondelettes à l'échelle la plus fine. Cette phase a permis de recaler les structures du signal à grande échelle.

- Une seconde phase de minimisation (25 itérations) a été réalisée en prenant en compte l'ensemble des coefficients. Cette seconde phase permet d'améliorer le champ de vitesse au niveau du vortex et d'éliminer une partie du bruit créé lors de la première phase.

Les résultats de ces expériences sont présentés en figure 5.31. On observe qu'en appliquant cette méthode, les résultats obtenus dans une base d'ondelettes sont corrects. L'effet d'une trop grande distance entre la première observation et l'image synthétique créée par le modèle est résolu en n'intégrant pas l'information vivant à une échelle fine dans un premier temps.

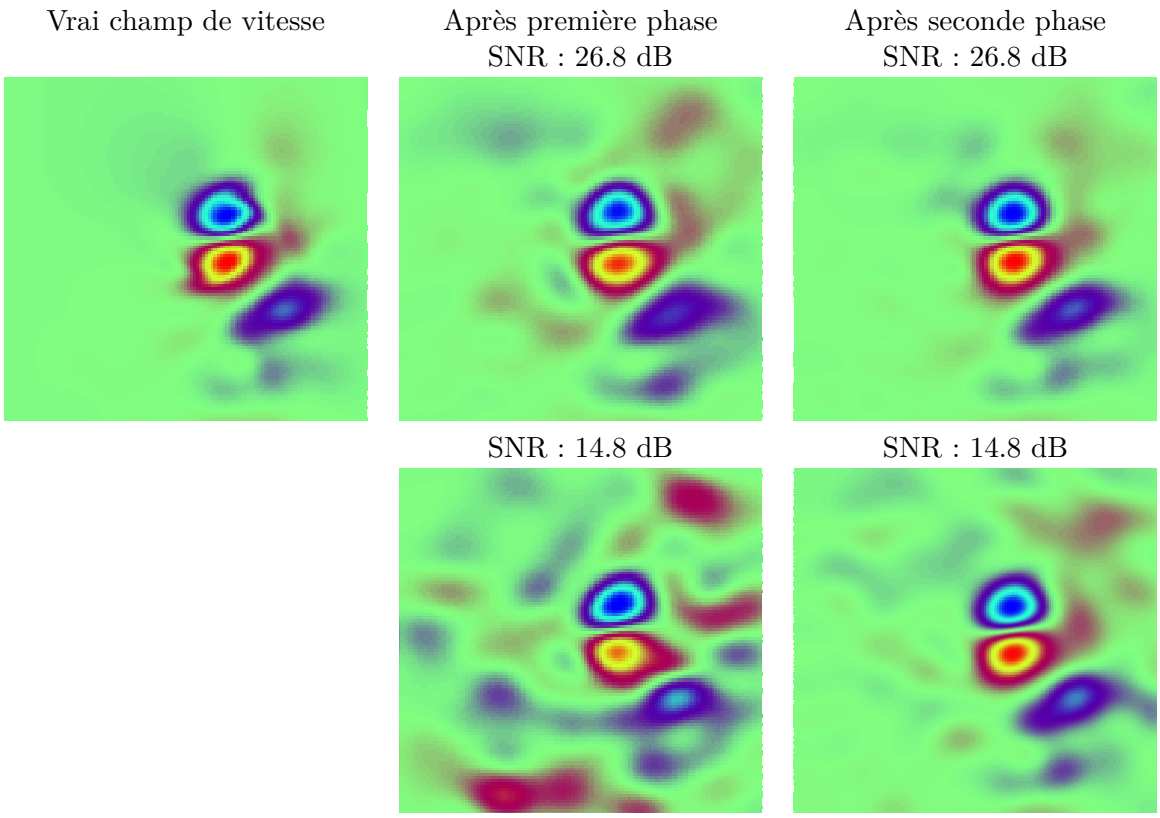


FIGURE 5.32 – Composante zonale de la vitesse analysée après la première phase (au centre) et la seconde phase (à droite). La première ligne correspond à un bruit de faible amplitude (SNR de 26.8 dB). La seconde ligne correspond à un bruit important (SNR de 14.8 dB).

La figure 5.32 illustre l'état obtenu après la première phase et après la seconde pour deux niveaux de bruit. On observe que la seconde phase est utile quelque soit le niveau de bruit. Elle permet de débruiter le champ de vitesse obtenu au préalable et de mieux reproduire la structure du vortex (lorsque l'information est très bruitée).

Bilan 5.10

On a vu au cours de cette expérience qu'améliorer les statistiques d'erreurs d'observations ne conduisait pas nécessairement à une meilleure analyse. Dans le cas présent, les observations étant trop éloignées de leur équivalent modèle, la minimisation ne peut être effectuée correctement lorsque l'ensemble de l'information est utilisée. Cependant l'emploi de la décomposition du signal dans une base multi-échelle permet de contourner ce problème en recalant d'abord les structures à une échelle plus grossière puis en affinant l'analyse via l'incorporation de l'information contenue à de plus fines échelles.

5.8.5 Anisotropie et choix de la représentation

Les expériences précédentes ont toutes été effectuées en présence de bruit isotrope et homogène en espace. Cependant, bien que ce type de bruit soit plus aisé à utiliser, rien ne garantit que le bruit rencontré lors d'expériences réelles respecte ces hypothèses.

Dans ce paragraphe, nous évaluons la performance des différentes représentations diagonales introduites en présence d'un bruit anisotrope et non homogène en espace.

Expériences effectuées

But : Voir plus loin que le cas d'un bruit isotrope homogène.

Observations utilisées : Une séquence de 24 images bruitées. Le bruit est inhomogène et anisotrope. Les erreurs d'observations sont corrélées en espace et décorrélées en temps. Trois niveaux de bruits sont considérés.

Termes d'écarts aux observations : Pixels-scal, **Daub-diag**, **Curv-diag**, **Fourier-diag** et **∇-scal**.

A savoir : Pour chaque niveau de bruit, 10 scénarios de bruits différents sont utilisés. Les résultats présentés dans le tableau 5.5 sont les résultats moyens sur 10 expériences.

Autres paramètres : La longueur de corrélation d'erreur d'ébauche est de 7.65 pixels. La fenêtre d'assimilation est de 2 h 24 en équivalent atmosphère. Le traceur est observé toutes les 6 minutes.

Appartée – Génération du bruit inhomogène anisotrope

Le bruit inhomogène anisotrope utilisé a été généré sous matlab. Ce dernier est une somme de bruits gaussiens anisotrope (chaque bruit est généré via l'utilisation d'un filtre anisotrope). La direction privilégiée, ainsi que le rapport $\frac{L_x}{L_y}$ (où L_x représente la longueur de corrélation dans la direction privilégiè et L_y la longueur de corrélation dans la direction orthogonale à cette première) varient pour chacun des bruits "sources" utilisés. En fin de compte, l'erreur effectuée en chaque point provient donc de différentes "sources de bruit" dont l'intensité varie selon la position spatiale.

Exemple de bruit : La figure 5.33 présente une réalisation d'un bruit gaussien anisotrope non homogène (à gauche) et un impact de bruit sur les observations utilisées (à droite). On observe que la direction privilégiée ainsi que la longueur de corrélation du bruit varie selon la localisation spatiale des pixels. La structure (et l'échelle) du bruit varie fortement d'un pixel à un autre.

Résultats : Le tableau 5.5 présente le pourcentage d'erreur résiduelle pour des modélisations diagonales de la matrice de covariance d'erreur d'observation dans différents espaces.

Comme dans le cas d'un bruit isotrope et homogène en espace, on observe que prendre en compte une partie des corrélations (**Curv-diag**, **Fourier-diag** et **Daub-diag**) permet d'obtenir de meilleurs résultats, et ce quelque soit le niveau de bruit. Comme précédemment, cet impact est d'autant plus grand que le niveau de bruit est important.

Cependant, contrairement au cas isotrope homogène, il y a ici une nette différence entre les différentes représentations du signal dans des espaces multi-échelles. Les résultats obtenus en écrivant la fonction coût dans la trame de curvelettes sont bien meilleurs que ceux obtenus lors de l'utilisation de bases d'ondelettes ou de la transformée de Fourier. Ceci est dû au fait que chaque atome de la famille de curvelettes possède trois caractéristiques importante : une localisation spatiale , une orientation privilégiée et une échelle. La combinaison de ces trois caractéristiques

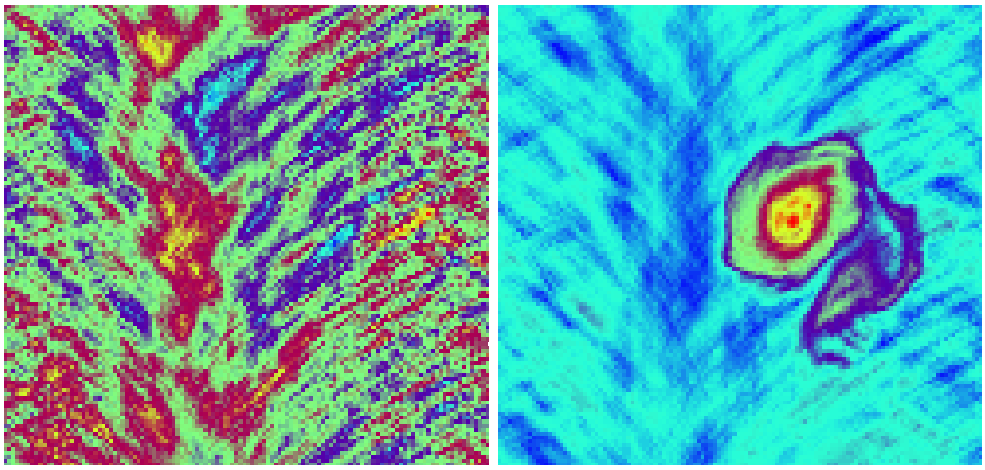


FIGURE 5.33 – Gauche : réalisation d'un bruit anisotrope inhomogène. Droite : impact d'un tel bruit sur les observations.

TABLE 5.5 – Moyenne sur 10 expériences du pourcentage d'erreur résiduelle dans la composante méridionale du champ de vitesse analysée, v_0^a, par rapport à l'erreur présente dans l'ébauche (v_0^b). Ce pourcentage est calculé pour différents niveaux de bruit.

SNR	Bruit fort 12.5 dB	18.5 dB	Bruit faible 26.5 dB
Curv-diag	14%	10%	8%
Fourier-diag	23%	14%	8%
Daub-diag	25%	14%	9%
∇-scal	22%	15%	10%
Pixels-scal	74%	54%	18%

permet de représenter, via des matrices diagonales, des bruits anisotropes et inhomogènes en espace.

Remarque 5.39

Le bruit n'a pas été choisi de telle sorte qu'il soit représenté de manière optimale dans la famille des curvelettes. Toutes les matrices de covariance d'observations utilisées sont uniquement des approximations grossières des véritables matrices. Il est bon de noter que dans ces expériences seules la matrice définie en équation (5.63) a été utilisée dans la trame de curvelettes (nous avons vu en section 5.6.2 que cette matrice n'était pas nécessairement la matrice diagonale optimale dans une trame). Il serait intéressant dans un futur proche de voir si l'utilisation d'une matrice de pondération (5.65) permettrait d'améliorer de manière substantielle les résultats.

Pour ces expériences, nous observons qu'utiliser une partie uniquement de l'information en passant dans l'espace des gradients permet d'obtenir des résultats comparables à ceux obtenus dans des espaces d'ondelettes ou dans l'espace spectral. Ceci souligne le fait que filtrer une partie du bruit (ainsi qu'une partie du signal) peut être dans certains cas aussi bénéfique que d'approcher trop grossièrement les statistiques d'erreurs d'observations.

Bilan 5.11

Au travers de ces expériences, nous avons pu voir que le choix de l'espace dans lequel les observations sont décomposées est guidé par les statistiques du bruit que l'on cherche à

modéliser. Dans le cas d'un bruit anisotrope et inhomogène, la représentation en curvelettes semble mieux adaptée que celle en ondelettes ou dans l'espace spectral. Nous avons aussi vu qu'utiliser une partie seulement de l'information contenue dans le signal (en passant dans l'espace des gradients) permet d'obtenir des résultats comparable à des représentations bien plus complexes (Fourier, Ondelettes).

6

Gestion des occultations dans des bases d'ondelettes

Sommaire

6.1 Données parfaites et nuages 151
6.1.1 Fonction coût dans l'espace des pixels 152
6.1.2 Expériences jumelles 153
6.1.3 Bilan partiel 155
6.2 Impact des occultations sur le modèle de corrélations 156
6.3 Comparaison d'images occultées dans un espace d'ondelettes 158
6.3.1 Approximation de la matrice de covariance sans prise en compte des occultations 160
6.3.2 Que se passe-t-il lorsques certains pixels viennent à manquer ? Illustration dans l'espace des gradients. 162
6.3.3 Pourquoi les coefficients de petite échelle sont peu affectés par un bruit corrélé ? 164
6.4 Modélisation de la variance en présence d'occultation 166
6.4.1 Pourcentage d'information disponible 166
6.4.2 Calcul du terme d'inflation 169
6.4.3 Approximation de la variance de coefficients partiellement occultés . . . 171
6.5 Conclusion 176

Dans la partie précédente, nous avons vu l'intérêt de la représentation du signal dans des espaces multi-échelles lorsque les observations (de type images) sont perturbées par un bruit présentant des corrélations spatiales. Cette représentation permet de prendre en compte de manière simple une partie de l'information sur les statistiques d'un bruit corrélé en espace. Cependant, en plus d'être bruitées, les images observées peuvent être en partie occultées. C'est souvent le cas des observations satellites de l'océan pour lesquelles les nuages constituent une barrière infranchissable.

Dans ce chapitre, nous introduisons une méthode pour adapter, à moindre coût, un modèle de corrélation connu lorsque les observations sont visibles partout, au cas où ces dernières sont partiellement occultées.

6.1 Données parfaites et nuages

Avant de nous attaquer à des données bruitées (et au problème de modélisation de la matrice de covariance d'erreur d'observation), commençons par vérifier que cette assimilation conduit à des résultats tangibles lorsque les données ne sont pas bruitées.

6.1.1 Fonction coût dans l'espace des pixels

Nous nous plaçons dans le contexte où nous connaissons la position des données manquantes. Cette position nous est fournie sous la forme d'un masque, noté $\mathbf{M}$ lorsqu'il s'agit d'une matrice diagonale et $\mathbf{m}$ lorsqu'il s'agit d'une image. Ces opérateurs sont tels que :

$$\mathbf{M}(i,j) = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j \text{ et le pixel } i \text{ est observé} \\ 0 & \text{si } i = j \text{ et le pixel } i \text{ n'est pas observé} \end{cases} \tag{6.1}$$

et

$$\mathbf{m}[k,l] = \begin{cases} 1 \text{ si le pixel situé en position } (k,l) \text{ de l'image est observé} \\ 0 \text{ sinon} \end{cases} \tag{6.2}$$

$\mathbf{M}$ est donc une matrice de taille $N \times N$ tandis que $\mathbf{m}$ est une image de taille $n_x \times n_y$ avec $N = n_x \times n_y$ le nombre de pixels de l'image.

$\mathbf{m}$ comporte juste l'information portée par la diagonale de la matrice $\mathbf{M}$ mais arrangée sous forme d'image : si le pixel est présent le masque vaut 1 et s'il est absent, le masque vaut 0. La matrice $\mathbf{M}$ est utilisé lorsque des des opérations sur la matrice de covariance sont nécessaires. Les deux écritures du masque sont introduite afin de ne pas (trop) abuser des notations.

L'image observée peut s'écrire, lorsque certaines données sont occultées, de la façon suivante :

$$I^o_{\mathbf{M}} = \mathbf{M} I^o. \tag{6.3}$$

où I^o est l'image sans données manquante.

La fonction coût dans l'espace des pixels est définie uniquement entre les pixels observés et leur équivalent modèle. Cette dernière s'écrit sous la forme :

$$J^o_{pix} = (\mathbf{P_M}(I^o_{\mathbf{M}} - I^{synth}))^T \mathbf{R}^{-1}_{pixObs} (\mathbf{P_M}(I^o_{\mathbf{M}} - I^{synth})) \tag{6.4}$$

où $\mathbf{P_M}$ est l'opérateur de projection contraignant les images à vivre dans l'espace des pixels observés (voir figure 6.1).

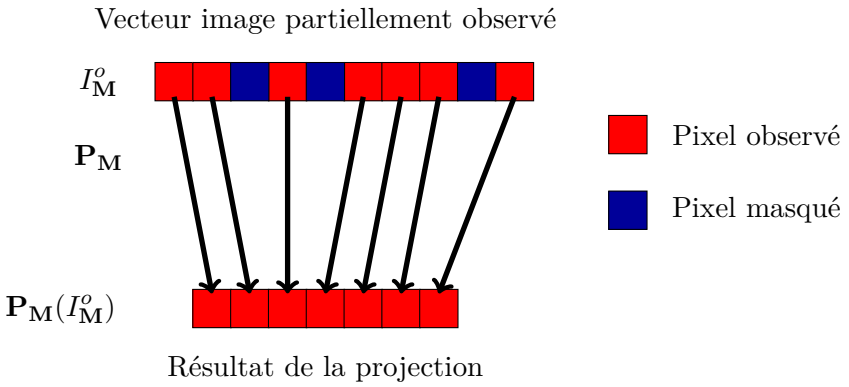


FIGURE 6.1 – Illustration de l'action de l'opérateur de projection

La matrice $\mathbf{R}_{pixObs}$ est la projection de la matrice $\mathbf{R}_{pix}$ dans le sous-espace observé :

$$\mathbf{R}_{pixObs} = \mathbf{P_M} \mathbf{R}_{pix} \mathbf{P}^T_{\mathbf{M}}. \tag{6.5}$$

Cette matrice est donc de dimension $k \times k$ où k est le nombre de pixels observés.

Remarque 6.1

Lorsque la matrice $\mathbf{R}_{pix}$ est inversible, on est assuré que la matrice $\mathbf{R}_{pixObs}$ définie en

équation (6.5) l'est aussi.

6.1.2 Expériences jumelles

Pour évaluer l'impact des nuages, des expériences sont réalisées avec des séquences d'images partiellement observées. Les occultations sont représentées par un nuage se déplaçant au dessus du vortex. Différentes tailles de nuages sont considérées (voir figure 6.2). Lorsque les données ne sont pas corrompues par du bruit, les expériences sont réalisées uniquement avec le terme d'écart aux observations défini entre pixels. Le but de ces premières expériences est simplement de confirmer que suffisament d'information est présente pour effectuer l'analyse.

Présentation des données

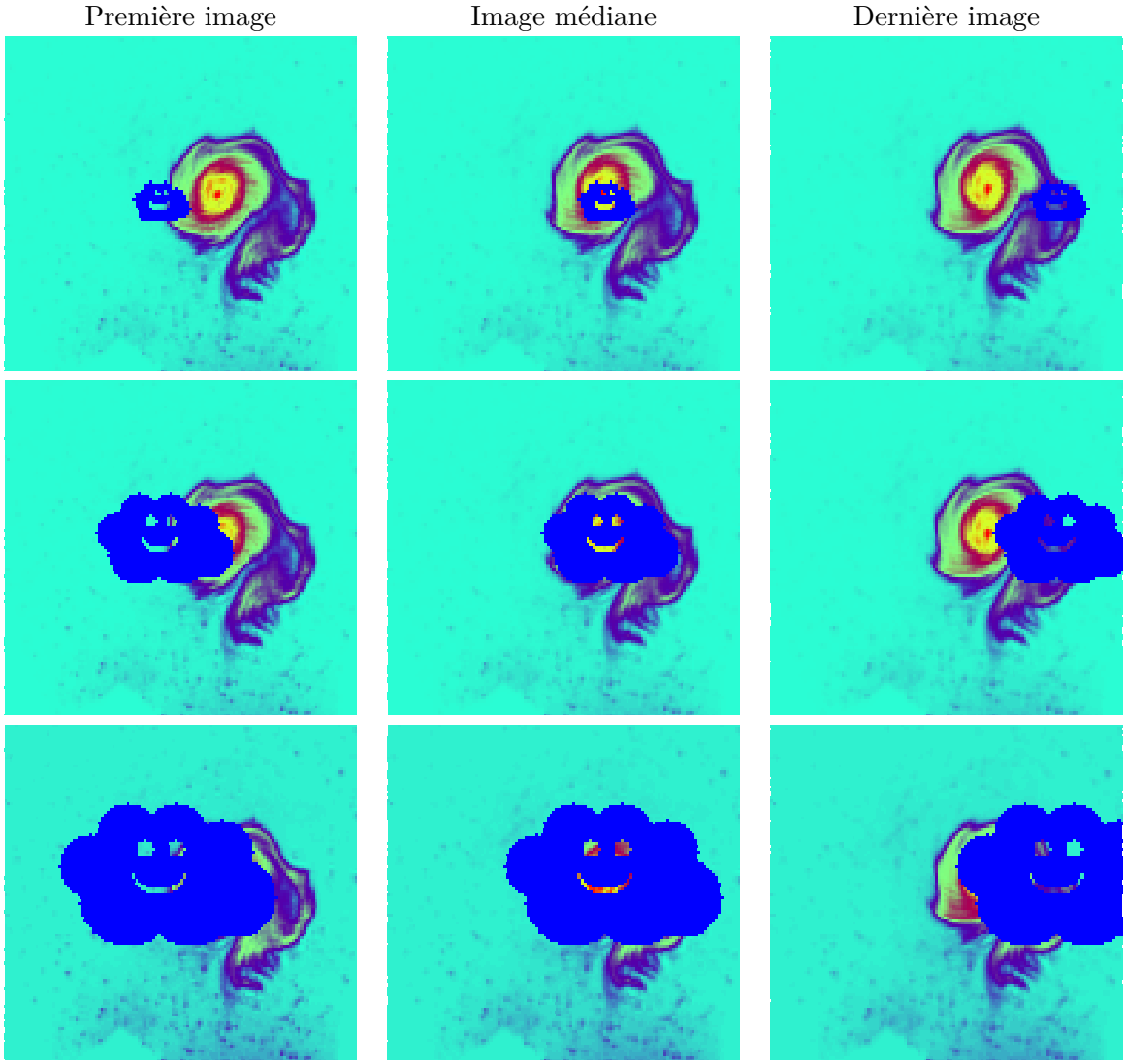


FIGURE 6.2 – Trois images de séquences assimilées.

La figure 6.2 présente, pour trois tailles de nuages, différentes images de la séquence de traceur passifs :

- La première séquence fait intervenir un nuage de toute petite taille. Pour ce nuage, les observations sont presque complètes. Dans la suite, ce nuage correspond au nuage de taille 3 ;

- La seconde séquence fait intervenir un nuage de taille plus importante. Sur chacune des images une faible portion du vortex est visible. Cette séquence correspond à un nuage de taille 7 ;
- La troisième séquence, le cas extrême est présenté. Ici, le coeur du vortex est observé uniquement au travers des yeux et du sourire du nuage. L'information utile est donc relativement faible. Cette séquence correpond à un nuage de taille 11.

Expériences réalisées

But : Déterminer si suffisamment d'information peut être extraite des séquences d'images pour contrôler le champ de vitesses initial.

Observations utilisées : Des données partiellement occultées par la présence de nuage.

Autres paramètres : La longueur de corrélation d'erreur d'ébauche est de 5.4 (pixels). La longueur de la fenêtre d'assimilation est de 2 h 24. Le traceur est observé toutes les 6 minutes.

Résultats obtenus

La figure 6.3 présente le pourcentage d'erreur residuelle de la composante méridionale de la vitesse et de la vorticité analysées. Ces résultat sont obtenus après 50 itérations de minimisation pour les différentes tailles de nuages étudiées.

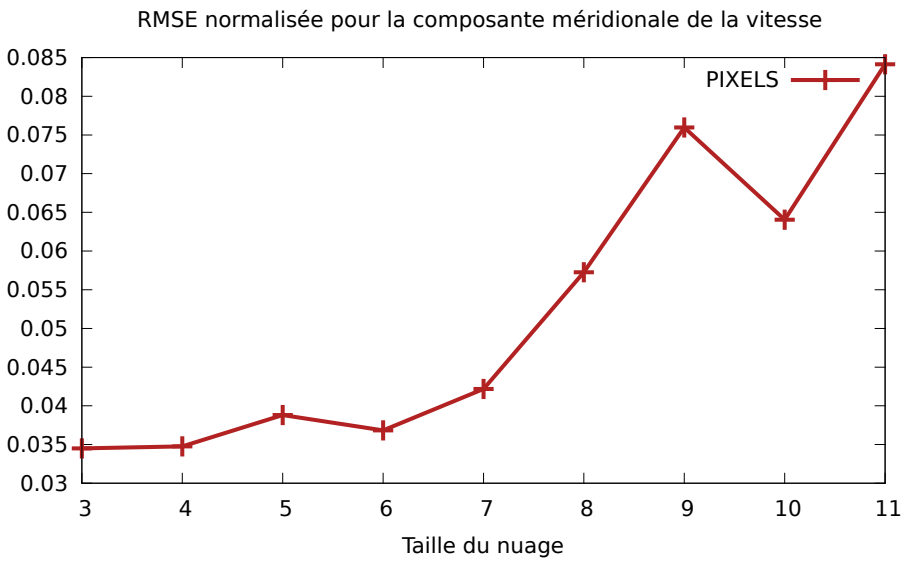


FIGURE 6.3 – Pourcentage d'erreur résiduelle sur la composante méridionale du champ de vitesse.

On observe que la phase d'analyse se déroule bien lorsque les données ne sont pas bruitées, et ce, quelle que soit la taille du nuage. Le pourcentage d'erreur résiduel de la composante méridionale de la vitesse est de 3.5% pour un nuage de petite taille et de 8.5% pour de plus grands nuages. A titre de comparaison, on peut se référer aux expériences réalisées en présence de bruit non corrélé (tableau 4.3) et corrélé (tableau 5.1). Dans l'espace des pixels, pour les données les moins bruitées ce pourcentage est de 4.1% lorsque le bruit est décorrélé et de 12.4% lorsque le bruit est corrélé. L'impact de la présence d'occultations dans les images observées est donc ici assez faible.

Bilan 6.1

Il est donc possible d'extraire assez d'information d'images partiellement masquées pour

contrôler le champ de vitesse initial du système. La quantité d'information contenue dans ces séquence est suffisante, et ce, même lorsque le vortex est simplement observé par les yeux et le sourire du nuage.

Localisation des erreurs

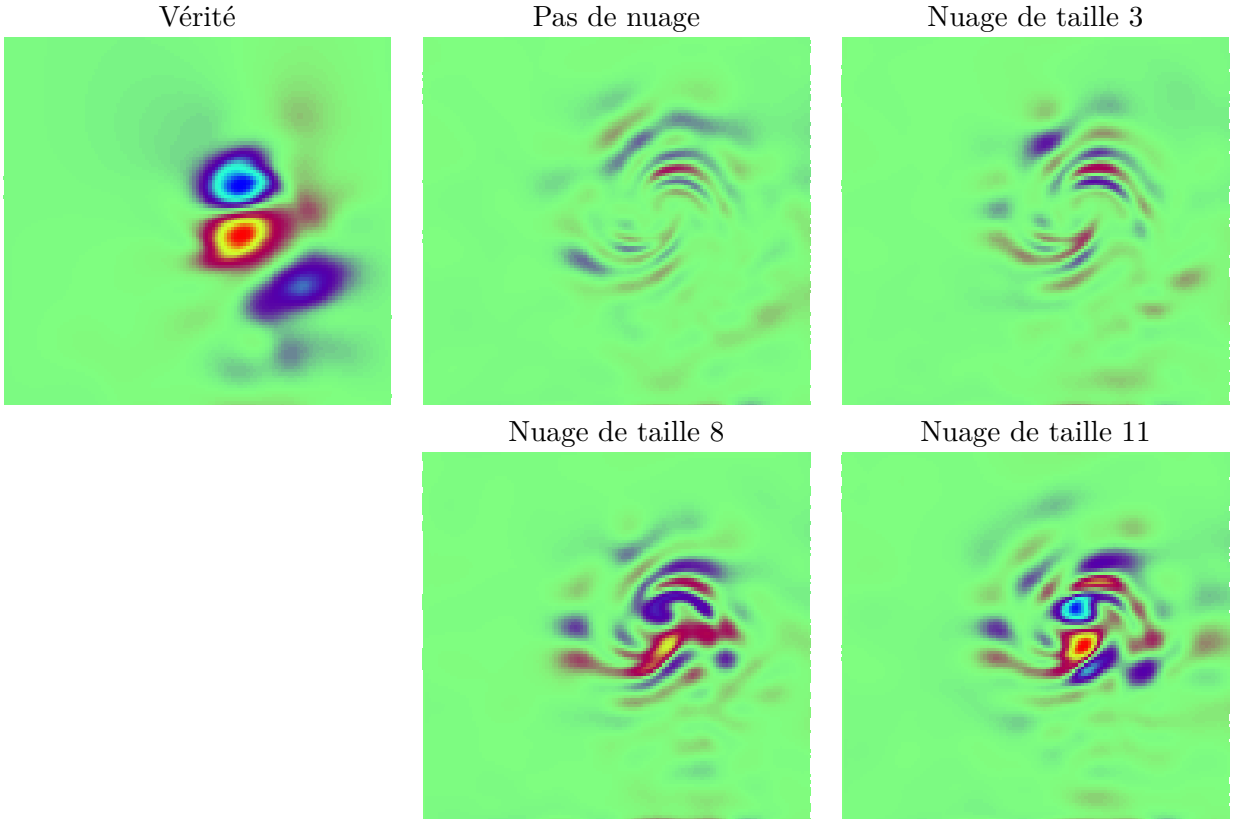


FIGURE 6.4 – En haut à gauche la vraie composante zonale de la vitesse. L'échelle va de $-0.028\ m.s^{-1}$ (en bleu) à $0.028\ m.s^{-1}$ (en rouge). Les autres figures présentent l'erreur effectuée en fonction de la taille du nuage. Pour ces figures, l'échelle va de $-0.004\ m.s^{-1}$ (en bleu) à $0.004\ m.s^{-1}$ (en rouge).

La figure 6.4 présente le champ d'erreur de la composante zonale de la vitesse pour 3 tailles différentes de nuage. L'erreur effectuée lorsque les observations ne sont pas occultées est présentée pour comparaison. Comme attendu, les différences entre ces champs d'erreurs se situent principalement dans la zone du vortex. Lorsque le nuage est de petite taille, la structure et l'amplitude des erreurs sont comparables à celles obtenues avec des images non occultées. De manière globale, l'erreur d'analyse croît avec la taille du nuage. L'erreur effectuée est dans ce cas due à une sous-estimation de l'intensité du vortex.

6.1.3 Bilan partiel

Au cours de cette section, la fonction coût entre une image partiellement observée et son équivalent modèle a été définie dans l'espace des pixels. Les différentes séquences servant de test pour la section suivante ont été présentées. Lorsqu'elles ne sont pas bruitées, ces séquences contiennent suffisamment d'information pour permettre de reconstruire de manière correcte le champ de vitesse initial (voir figure 6.3). Les erreurs d'analyses augmentent en général avec la taille du nuage. Ces erreurs sont dues à une sous-estimation de l'intensité du vortex.

La suite de ce chapitre est dévolue à la prise en compte de l'impact des occultations sur les corrélations modélisées à l'aide d'une matrice diagonale dans un espace d'ondelettes.

6.2 Impact des occultations sur le modèle de corrélations

Avant de passer à l'assimilation d'images corrompues par un bruit corrélé en espace, il est nécessaire d'étudier l'impact des occultations sur le modèle de corrélations. Cette question est en fait plus complexe qu'il n'y paraît au premier abord. Pour pouvoir y répondre, nous devons savoir d'où proviennent les corrélations présentes au sein de notre champ d'erreurs.

Avoir cette connaissance permet de comprendre l'impact d'une occultation sur le modèle de corrélation :

- Si les corrélations sont dues à l'usage d'un même instrument de mesure, alors l'absence d'une observation ne change pas le schéma de corrélation entre les autres observations ;
- Si les corrélations proviennent de l'erreur de représentativité, il en va de même. Le schéma 6.5 illustre ces deux premiers cas ;
- En revanche, si les corrélations sont dues au pré-traitement que subissent les données, le schéma de corrélation peut changer (voir exemple 6.1). Dans ce cas, il est nécessaire de connaître la manière de prétraiter les données pour pouvoir adapter le schéma de corrélation dans l'espace des pixels lorsque certaines données viennent à manquer.

En pratique, tous ces phénomènes sont responsables des corrélations au sein du champ d'erreurs d'observations. Ici, nous nous intéressons uniquement au cas ne nécessitant pas de connaître le pré–traitement effectué lorsqu'une donnée est manquante.

Cas où le modèle de corrélation n'est pas impacté par l'absence d'un pixel

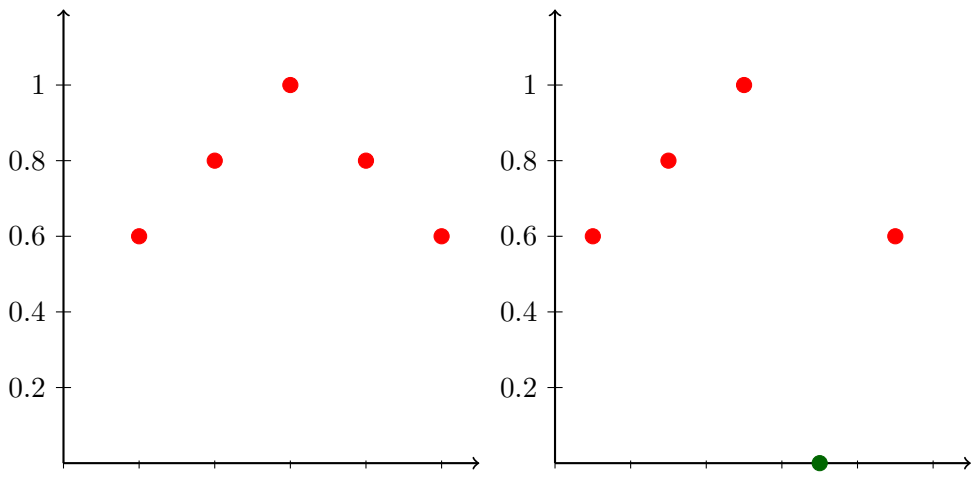


FIGURE 6.5 – Corrélations entre un point et ses voisins. A gauche, le schéma de corrélation lorsque toutes les données sont observées. A droite, le schéma de corrélation lorsqu'une donnée n'est pas observée.

Exemple 6.1

Pour comprendre ce phénomène de "changement de modèle de corrélation" plaçons nous dans un cas 1D et supposons que le prétraitement appliqué consiste à convoluer le signal avec filtre gaussien suivant :

$$g = \begin{pmatrix} 0.065 & 0.18 & 0.51 & 0.18 & 0.065 \end{pmatrix}. \tag{6.6}$$

afin d'obtenir en chaque point une moyenne pondérée. Nous nous plaçons dans le cas où les erreurs sur les données brutes $\mathbf{y}$ *sont décorrélées avant ce traitement (et de même amplitude).*
Les données après traitement sont :

$$\mathbf{y_g} = g \star \mathbf{y} \tag{6.7}$$

A gauche de la figure 6.6 les corrélations entre l'erreur effectuée en un point i *et celles effectuées en des points voisins est présentée lorsque tout les points sont observés.*

Supposons maintenant qu'une donnée vienne à manquer et que nous sommes au courant de cela. Dès lors, lorsque le pré-traitement est effectué, pour certains points la moyenne pondérée ne prend en compte plus que 4 *points au lieu de* 5 *(et est renormalisée pour obtenir une "moyenne"). On suppose de plus qu'on ne calcule pas la moyenne pondérée pour le pixel non observé. Dans ce cas, le modèle de corrélation vient à changer et est présenté en bleu à droite en figure 6.6.*

Cas où le modèle de corrélation est impacté par l'absence d'un pixel

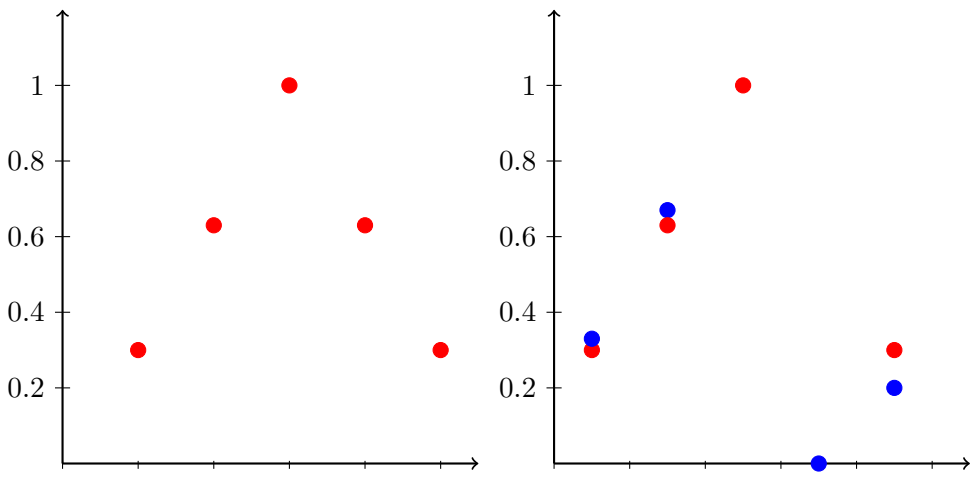


FIGURE 6.6 – Corrélations entre un point et ses voisins. A gauche, le schéma de corrélation lorsque l'ensemble des données sont observées. A droite, les corrélations obtenues lorsqu'une donnée vient à manquer sont présentées en bleu. On observe dans ce cas que l'ensemble du schéma de corrélation est modifié par le fait qu'une donnée n'est plus disponible.

Génération du bruit

Comme nous nous plaçons dans le cas où le modèle de corrélation entre points observés n'évolue pas à cause de l'absence de certains pixels, une manière de générer le bruit consiste

simplement à appliquer un masque $\mathbf{M}$ au vecteur de bruit en fin de processus. Les étapes permettant la génération d'un bruit corrélé partiellement occulté sont donc les suivantes :

- Génération d'un bruit non corrélé dans l'espace des pixels : $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I_n})$;
- Corrélation de ce bruit par l'application de $\mathbf{R}_{pix}^{1/2}$:

$$\boldsymbol{\epsilon}_{cor} = \mathbf{R}_{pix}^{1/2}\boldsymbol{\epsilon}. \tag{6.8}$$

  Ainsi, $\boldsymbol{\epsilon}_{cor} \sim \mathcal{N}(0, \mathbf{R}_{pix})$;
- Application du masque $\mathbf{M}$ au vecteur $\boldsymbol{\epsilon}_{cor}$:

$$\boldsymbol{\epsilon}_{cor}^{\mathbf{M}} = \mathbf{M}\boldsymbol{\epsilon}_{cor}. \tag{6.9}$$

Le bruit ainsi créé suit une loi gaussienne $\boldsymbol{\epsilon}_{cor}^{\mathbf{M}} \sim \mathcal{N}(0, \mathbf{M}\mathbf{R}_{pix}\mathbf{M})$

La matrice de covariance de l'ensemble des observations dans l'espace des pixels, $\mathbf{M}\mathbf{R}_{pix}\mathbf{M}$, est donc une matrice non inversible (car $\mathbf{M}$ est une matrice diagonale non inversible).

Un pixel "occulté" est considéré comme ne comportant pas d'erreur : il n'est tout simplement pas comparé avec son équivalent modèle. La comparaison se fait donc dans un espace d'observation de taille réduite comportant uniquement les pixels observés. C'est cette approche qui a été suivie dans la section précédente lorsque l'image n'était pas corrompue par du bruit.

6.3 Comparaison d'images occultées dans un espace d'ondelettes

Autant il est facile dans l'espace des pixels d'utiliser des images occultées (seule l'information observée est utilisée), autant ceci devient plus complexe dans un espace structuré. En effet, chaque structure est définie sur un certain support de l'image. Lorsqu'un pixel vient à manquer, que doit-on faire ? Peut-on toujours calculer de manière correcte le coefficient correspondant à la structure recherchée ? Quand peut-on considérer qu'une structure est véritablement présente dans l'image non occultée ? Quelle est la structure présente sous l'occultation ?

Ces questions arrivent naturellement lorsqu'on cherche à reconstruire une image complète à partir d'une image occultée. Pour pouvoir y répondre, il est cependant nécessaire d'avoir accès à des informations supplémentaires ([Ba et al., 2011]) et/ou de faire des hypothèses sur la structure du signal ([Bugeau et al., 2009] [Fadili et al., 2009] [Alvera-Azcárate et al., 2005]).

Dans notre cas, nous n'avons fort heureusement pas besoin d'apporter un élément de réponse à ces questions. Nous ne cherchons pas à analyser une séquence observée de manière isolée (pour laquelle la présence ou l'absence d'une structure spatiale ou temporelle en présence d'occultation est discutable) mais *simplement* à comparer cette séquence à son équivalent modèle.

Etant prévenu de la localisation des pixels manquants l'idée est, comme précédemment dans l'espace des pixels (voir équation (6.4)), de comparer une image observée avec son équivalent modèle uniquement dans l'espace observé. Cependant, ici nous ne pouvons pas nous restreindre à travailler uniquement avec les coefficients d'ondelettes entièrement observées [19]. Afin de travailler avec la même source d'information, il est nécessaire *d'occulter* l'équivalent modèle avant d'effectuer la transformée en ondelettes :

$$I_{\mathbf{M}}^{synth} = \mathbf{M}I^{synth} \tag{6.10}$$

19. Un coefficient d'ondelettes est calculé en prenant en compte l'information en provenance d'un nombre important de pixels. Inversement, l'information contenue dans un pixel est utilisée par un nombre important de coefficients d'ondelettes. N'utiliser que les coefficients d'ondelettes ayant accès à l'ensemble de l'information sur leur support revient donc à n'utiliser qu'un faible pourcentage de l'information disponible dans la séquence d'image.

Après application de ce masque à l'image, la fonction coût s'écrit de la manière suivante :

$$J^o_{wav} = (\mathbf{W}(I^o_{\mathbf{M}} - I^{synth}_{\mathbf{M}}))^T \mathbf{R}^{-1}_{\mathbf{WM}} (\mathbf{W}(I^o_{\mathbf{M}} - I^{synth}_{\mathbf{M}})). \tag{6.11}$$

Cette fonction coût est similaire à celle définie dans les sections précédentes à quelques différences près. Tout d'abord, un masque est appliqué à l'image synthétique. Une seconde différence provient de la définition de la matrice de covariance $\mathbf{R_{WM}}$ qui change. Et c'est cette définition qui est problèmatique.

Théoriquement, cette matrice doit s'écrire sous la forme $\mathbf{R_{WM}} = \mathbf{WMR}_{pix}\mathbf{MW}^T$ et n'est donc pas inversible (la matrice $\mathbf{M}$ n'étant pas inversible). Deux approximations diagonales différentes de cette matrice sont présentées dans la suite.

Remarque 6.2

Lorsque $\mathbf{R}_{pixObs} = \mathbf{I_k}$, où k correspond au nombre de pixels non occultés, il est aisé de définir une matrice $\mathbf{R_{WM}}$ permettant de résoudre le même système que dans l'espace des pixels.
En effet, dans ce cas nous avons :

$$\begin{aligned}
J^o_{pix} &= (\mathbf{P}_M(I^o_{\boldsymbol{M}} - I^{synth}))^T \mathbf{R}^{-1}_{pixObs} (\mathbf{P}_{\boldsymbol{M}}(I^o_{\boldsymbol{M}} - I^{synth})) && (6.12)\\
&= (\mathbf{P}_M(I^o_{\boldsymbol{M}} - I^{synth}))^T (\mathbf{P}_{\boldsymbol{M}}(I^o_{\boldsymbol{M}} - I^{synth})) && (6.13)\\
&= (I^o_{\boldsymbol{M}} - I^{synth}_{\boldsymbol{M}})^T (I^o_{\boldsymbol{M}} - I^{synth}_{\boldsymbol{M}}) && (6.14)\\
&= (I^o_{\boldsymbol{M}} - I^{synth}_{\boldsymbol{M}})^T \mathbf{W}^T \mathbf{W} (I^o_{\boldsymbol{M}} - I^{synth}_{\boldsymbol{M}}) && (6.15)\\
&= (\mathbf{W}(I^o_{\boldsymbol{M}} - I^{synth}_{\boldsymbol{M}}))^T (\mathbf{W}(I^o_{\boldsymbol{M}} - I^{synth}_{\boldsymbol{M}})) && (6.16)\\
&= J^o_{wav} && (6.17)
\end{aligned}$$

La matrice de pondération dans l'espace d'ondelettes peut donc être définie par $\mathbf{R_{WM}} = \mathbf{I_n}$, où n est le nombre de pixels (occultés ou non) de l'image (avec $n \gg k$).

Travailler dans un espace d'ondelettes ne présente alors que peu d'intérêt : l'ensemble de l'information dans cet espace étant utilisé, la taille du vecteur d'observation peut être grandement accrue (voir figure 6.7).

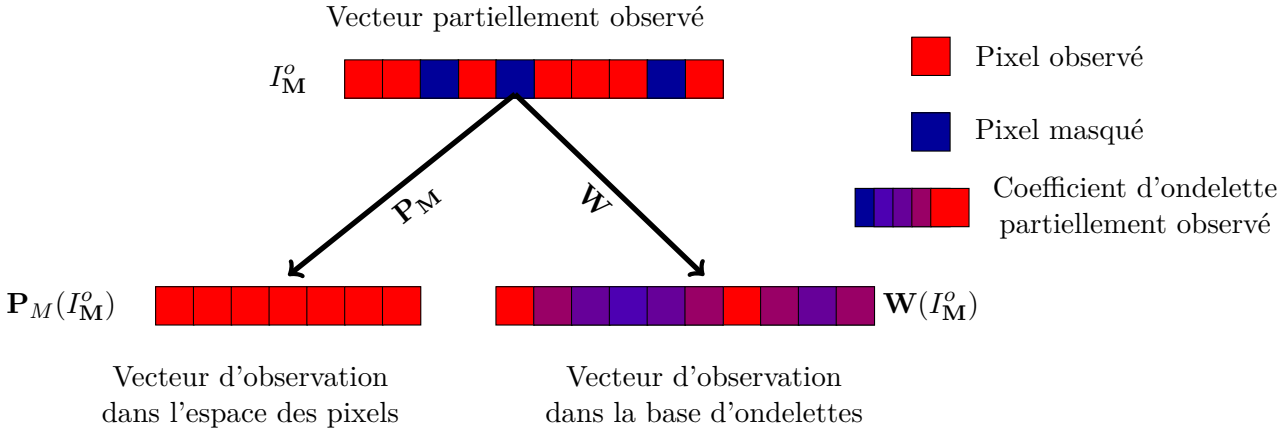


FIGURE 6.7 – Travailler dans une base d'ondelettes lorsque l'image est partiellement masquée conduit à travailler dans un espace de plus grande taille (mais contenant la même information).

6.3.1 Approximation de la matrice de covariance sans prise en compte des occultations

Une première idée consiste à supposer que la présence d'observations masquées n'est pas un problème en soi. Comme présenté en figure 6.5, le schéma de corrélation est peu perturbé lorsque peu de pixels viennent à manquer. Il est donc envisageable d'essayer d'utiliser la même matrice que celle construite en section 5.4.

Remarque 6.3

Utiliser cette matrice revient à prendre en compte le fait que certains pixels sont occultés uniquement lors du calcul de l'écart entre le modèle et les observations. Lors de la définition des coefficients pondérant la contribution de ces différents termes, l'information que certains pixels sont occultés n'est pas prise en compte.

Expériences effectuées

But : Observer l'effet de la non prise en compte des observations manquantes lors de la définition de la matrice $\mathbf{R}$ dans un espace d'ondelettes.

Observations utilisées : Des séquences d'images perturbées par un bruit corrélé en espace partiellement occultées.

A savoir : Une seule séquence est utilisée par taille de masque et niveau de bruit.

Autres paramètres : La longueur de corrélation de l'erreur d'ébauche est de 7.65 (pixels). La longueur de la fenêtre d'assimilation est de $2h24$. Une observation est utilisée toutes les 6 minutes.

La figure 6.8 présente les résultats obtenus lorsque cette matrice est utilisée. Le bruit corrompant les données est ici isotrope et homogène. Il correspond au bruit utilisé dans les expériences jumelles du chapitre précédent (voir section 5.8).

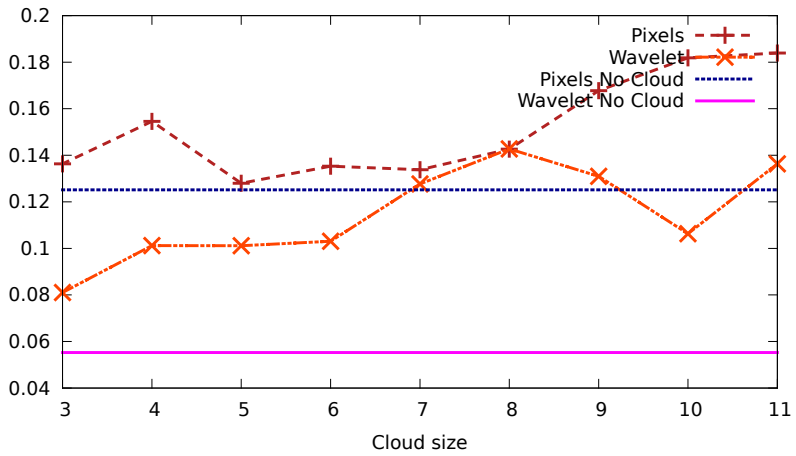
On observe que les résultats sont meilleurs lorsque la comparaison entre l'équivalent modèle et les observations est effectuée dans une base d'ondelettes. La différence entre ces deux approches provenant de la modélisation des erreurs, on en déduit qu'utiliser la matrice de covariance construite dans un cas non occulté peu s'avérer payant. De l'information supplémentaire sur la distribution du bruit est intégrée, et ce, notamment lorsque les coefficients d'ondelettes sont calculés sur des zones entièrement observées.

L'erreur d'analyse effectuée en présence d'un nuage de petite taille est cependant bien plus importante que celle lorsque l'image est entièrement observée. Le vortex étant presque entièrement observé à certains instants, ces résultats peuvent sembler *curieux* en comparaison de ceux obtenus lors des expériences portant sur le nombre d'observations nécessaires au contrôle du vortex en section 5.8.3. Nous avions notamment vu dans ces expériences qu'en utilisant seulement 3 observations l'erreur résiduelle sur composante méridionale du champ de vitesse était relativement faible (envrion 14%). Ici, l'erreur résiduelle est plus élevée ($\sim 20\%$) bien que nous ayons accès à bien plus d'information sur la dynamique du vortex (24 images peu occultées sont utilisées).

La figure 6.9 présente l'erreur effectuée sur la composante zonale du champ de vitesse pour différentes tailles de nuage. Les champs d'erreurs lorsque l'ensemble de l'image est observée sont présentés en première ligne.

Dans l'espace des pixels, la différence entre l'analyse effectuée lorsque l'ensemble des données sont observées et lorsqu'un petit nuage vient à passer réside uniquement dans l'amplitude des erreurs.

Séquence d'image peu bruitée (SNR = 26.8 dB)



Séquence d'image fortement bruitée (SNR = 14.8 dB)

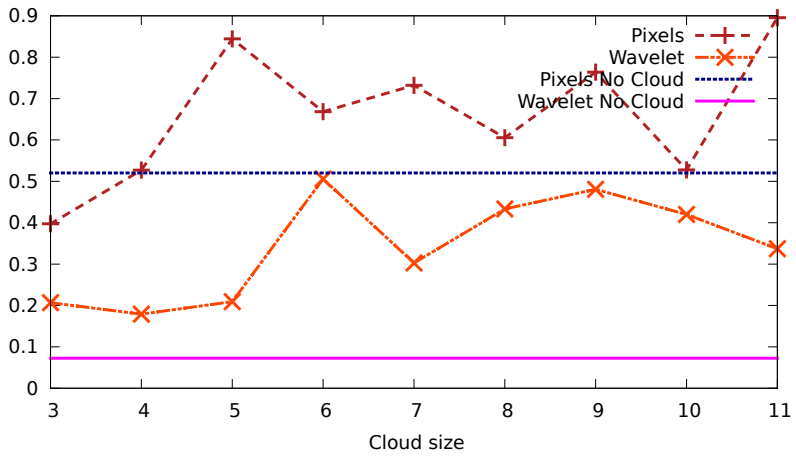


FIGURE 6.8 – Pourcentage d'erreur résiduelle sur la composante méridionale de la vitesse en fonction de la taille du nuage. Le pourcentage d'erreur restant dans l'analyse pour ces même séquences lorsque les images ne sont pas occultées est présenté à titre de comparaison .

Dans une base d'ondelettes nous n'assistons pas au même phénomène : même lorsque le nuage est de petite taille, il y a une nette différence entre la présence et l'absence de nuage. Ceci est contraire à ce qu'il advient lorsque 3 images sont assimilées en absence d'occultations (figure en haut au centre). En effet, dans ce cas, les erreurs présentent grossièrement les mêmes structures.

Bilan 6.2

Au vu des résultats présentés dans cette section, nous pouvons tirer plusieurs conclusions :
- *l'apport d'information induit par l'utilisation d'une matrice diagonale dans un espace d'ondelette peut être utile, et ce, même si les corrélations modélisées ne prennent pas en compte la présence de pixels manquants,*
- *cet apport d'information n'est cependant pas satisfaisant : à proximité du nuage, il y a apparition de nombreuses erreurs d'analyses bien que ces zones soient à d'autres moments parfaitement observées.*

Dans cette section, l'approximation de la matrice de covariance effectuée ne prend pas en compte l'absence de certains pixels avant d'être inversée. Comme présenté en figure 6.10, tenir compte de cette information avant d'inverser la matrice est crucial pour effectuer une approxi-

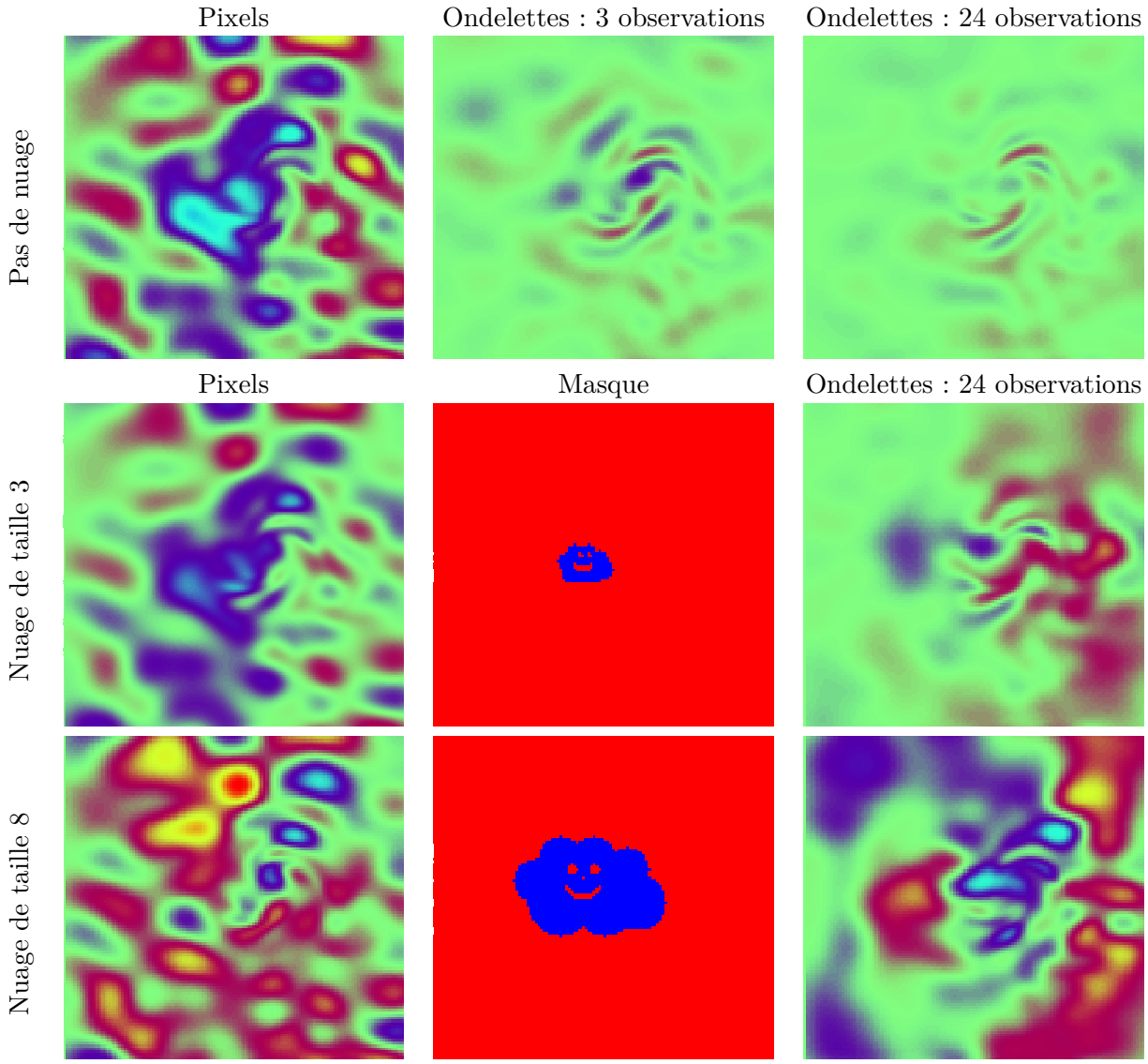


FIGURE 6.9 – Erreur effectuée sur la composante zonale de la vitesse pour le type de séquence le plus bruitées (SNR = 14.8 dB). Echelle : $-0.01\ m.s^{-1}$ (en bleu) à $0.01\ m.s^{-1}$ (en rouge). A titre de comparaison les valeurs de cette composante vont de $-0.028\ m.s^{-1}$ à $0.028\ m.s^{-1}$.

mation convenable de la matrice inverse.

Dans les sections qui suivent, nous allons voir comment, à partir de l'approximation diagonale de la matrice de covariance dans une base d'ondelettes, nous pouvons tenir compte de l'effet de ces occultations avant d'effectuer l'inversion en modifiant notre approximation.

6.3.2 Que se passe-t-il lorsques certains pixels viennent à manquer ? Illustration dans l'espace des gradients.

Savoir comment mieux prendre en compte le fait que certains pixels soient manquants dans une base d'ondelettes n'est pas chose aisée. Il faut comprendre pourquoi certains éléments sont peu affectés par un bruit corrélé et savoir pourquoi lorsque certains pixels viennent à manquer, ces coefficients d'ondelettes se retrouvent plus perturbés par le bruit.

Afin de mieux appréhender le problème rencontré lorsque la définition de la variance n'est pas adaptée à la présence d'occultations, nous pouvons évaluer l'effet de ces dernières sur la variance dans l'espace des gradients.

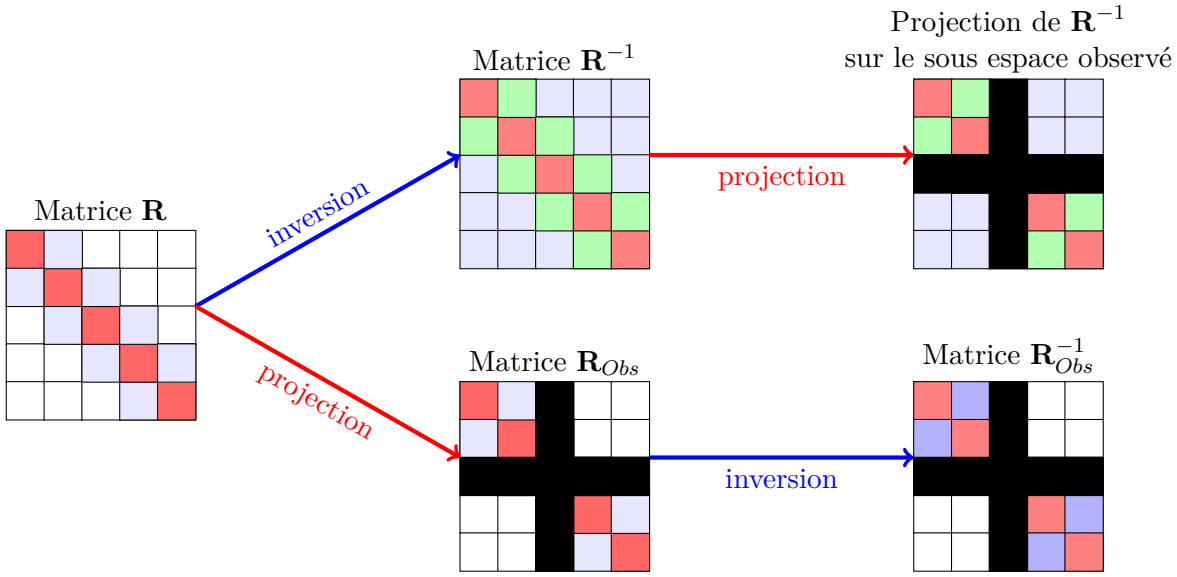


FIGURE 6.10 – Illustration du problème dû à l'utilisation d'une matrice inverse en présence d'occultation. Projeter la matrice inversée dans l'espace des pixels observés peut ne pas conduire à considérer une bonne approximation de l'inverse de la matrice de covariance dans cet espace.

Commençons par rappeler que le gradient dans la direction x au pixel (i, j) est calculé par un schéma de différence finie centrée :

$$I_x(i,j) = \frac{I(i+1,j) - I(i-1,j)}{2}. \tag{6.18}$$

Comme présenté en section 5.2.2, lorsque les pixels $I(i+1,j)$ et $I(i-1,j)$ sont observés, la variance de $I_x(i,j)$ est donnée par :

$$\begin{aligned} \mathrm{var}\left(I_x(i,j)\right) &= \mathrm{var}\left(\frac{I(i+1,j)}{2}\right) + \mathrm{var}\left(\frac{I(i-1,j)}{2}\right) - 2\frac{\mathrm{cov}\left(I(i+1,j), I(i-1,j)\right)}{4} \\ &= \sigma^2 \frac{1-\alpha}{2} \end{aligned}$$

où σ^2 est la variance de l'erreur au pixels $I(i+1,j)$ et $I(i-1,j)$ et α est la corrélation entre ces deux erreurs.

Lorsque le pixel $I(i-1,j)$ est masqué, ce pixel n'est pas affecté par une erreur. En effet, son équivalent modèle est lui aussi masqué. Par conséquent, la distance entre ces deux pixels est nulle. Dès lors, la variance de l'erreur de $I_x(i,j)$ est donnée par :

$$\begin{aligned} \mathrm{var}\left(I_x(i,j)\right) &= \mathrm{var}\left(\frac{I(i+1,j)}{2}\right) \\ &= \frac{\sigma^2}{4} \end{aligned}$$

La figure 6.11 présente la variance de l'erreur d'observation effectuée sur $I_x(i,j)$ en fonction de la corrélation entre les erreurs situées au pixel $I(i+1,j)$ et $I(i-1,j)$. Lorsque les deux pixels sont observés, cette variance décroît (courbe bleue). Lorsqu'un seul pixel est observé, la variance est toujours la même (courbe verte).

Deux forces s'opposent lors du calcul de la variance dans l'espace des gradients :

1. moins il y a d'information (pixels non occultés) et moins la variance est grande,

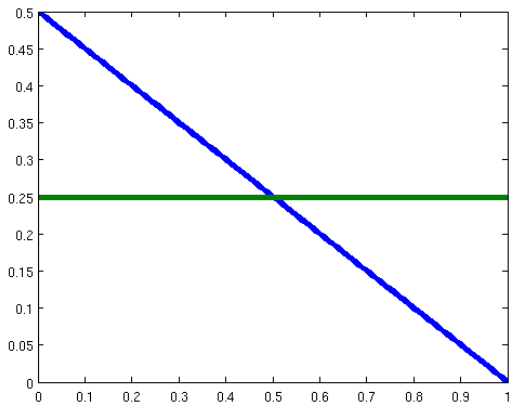


FIGURE 6.11 – Variance de l'erreur sur $I_x(i,j)$ en fonction de la corrélation entre l'erreur effectuée en $I(i+1,j)$ et $I(i-1,j)$. En bleu la variance est donnée lorsque $I(i+1,j)$ et $I(i-1,j)$ sont observées. En vert lorsque un seul des deux pixels est observé.

2. lorsqu'un seul pixel est visible, la réduction de variance due aux corrélations entre pixels est annulée. Il en résulte que la variance de l'erreur effectué en $I_x(i,j)$ est plus importante en utilisant un pixel entâché d'erreur et un pixel "artificiel" (ne portant pas d'erreur) que deux pixels entâchés d'erreur lorsque $\alpha > 1/2$.

Dans l'espace des gradients, les calculs pour tenir compte des occultations sont simples : le nombre de cas est limité à trois (0, 1 ou 2 pixels manquants). Dans une base d'ondelettes, ces calculs deviennent bien plus complexes. Le support d'une ondelette dépend en effet :
- de la base choisie,
- de l'échelle,
- de l'orientation.

De plus, le poids des différents pixels contribuant à la création du coefficient d'ondelette n'est pas le même. Le nombre de cas à prendre en compte est trop important pour être listé de manière exhaustive.

Une possibilité (non employée ici) pourrait être de réeffectuer l'ensemble des opérations conduisant à l'approximation diagonale dans une base d'ondelettes pour chaque image partiellement observée. Cependant, au vu du coût induit par ces opérations (le changement d'espace coûte cher), cela ne semble pas envisageable en pratique.
Dès lors, dans la section suivante, nous allons chercher à modéliser l'effet des occultations sur la variance des coefficients d'ondelettes. Pour cela, nous allons tenir compte de l'effet de la déflation de la variance (dû au fait que moins d'erreur est intégrée au calcul du coefficient d'ondelette) et de l'inflation de la variance (dû au fait que les erreurs fortement corrélées en espace ne se compensent plus).
La différence entre ces deux approches est présenté en figure 6.12.

6.3.3 Pourquoi les coefficients de petite échelle sont peu affectés par un bruit corrélé ?

Afin d'évaluer les changements dus à la non présence de certains coefficients, il est nécessaire de comprendre pourquoi les coefficients des petites échelles (capturant des discontinuités sur un petit support) sont peu perturbés par un bruit corrélé (de manière positive).

Cette bonne propriété des coefficients de détail dans des bases d'ondelettes est liée au nombre de moments nuls de l'ondelette (et à la taille du support de celle-ci).

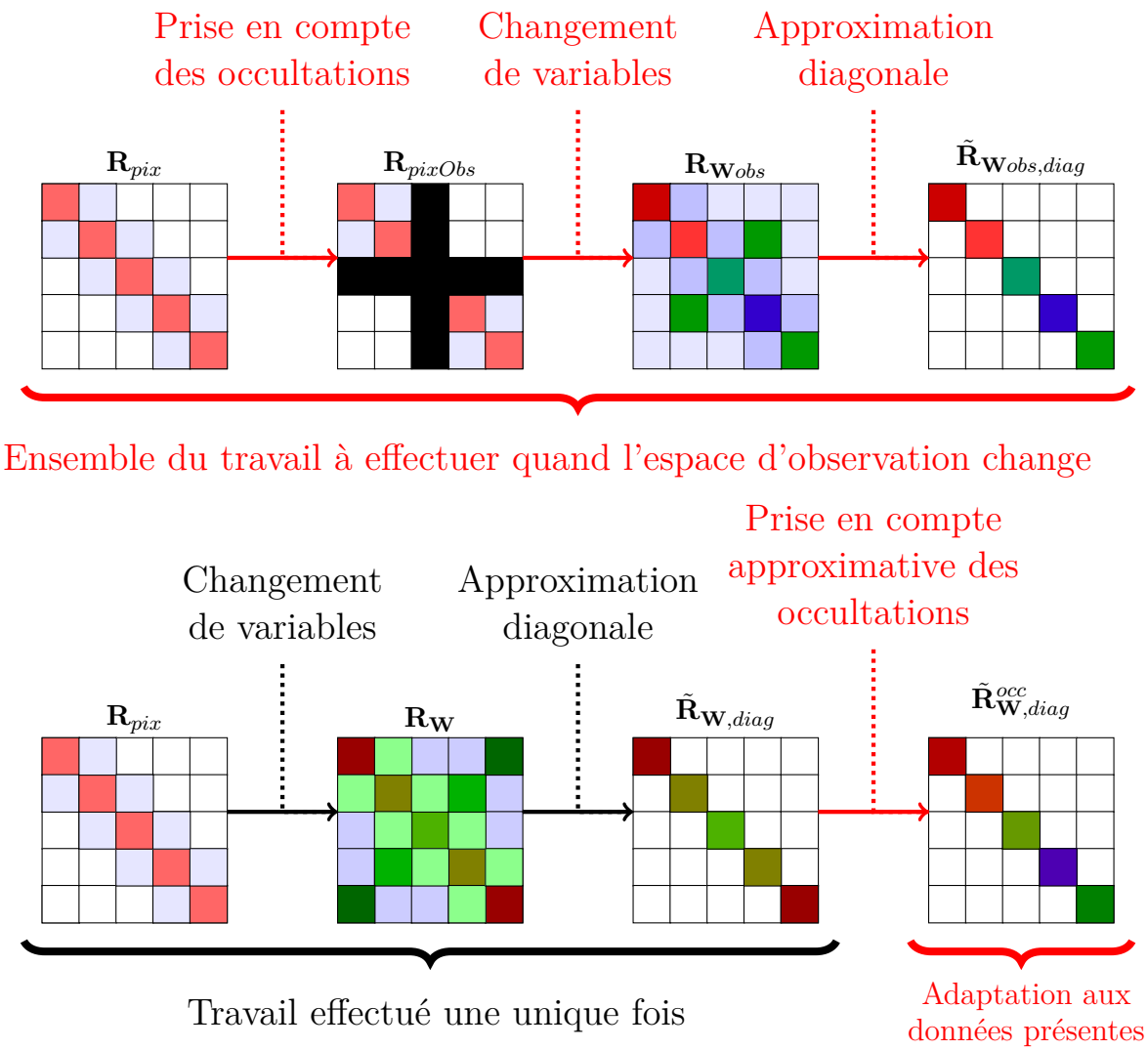


FIGURE 6.12 – Illustration de la méthode de détermination de la matrice de poids diagonale dans une base d'ondelettes lorsque des nuages viennent occulter la séquence d'images.

Pour comprendre ceci, rappelons la formule (en $1D$) permettant de calculer la variance d'un coefficient de détail à l'échelle j :

$$\begin{aligned}
\operatorname{var}\left(d^j[k]\right) &= \operatorname{cov}\left(\sum_n g[n-2k]c^{j+1}[n], \sum_m g[m-2k]c^{j+1}[m]\right) \\
&= \sum_{n,m} g[n-2k]g[m-2k] \operatorname{cov}\left(c^{j+1}[n], c^{j+1}[m]\right) \qquad (6.19) \\
&= \sum_n g[n-2k] \sum_m g[m-2k] \operatorname{cov}\left(c^{j+1}[n], c^{j+1}[m]\right) \qquad (6.20)
\end{aligned}$$

Le filtre g, permettant l'extraction des détails à une certaine échelle, possède la propriété d'être de moyenne nulle :

$$\sum_n g[n] = 0. \qquad (6.21)$$

En fait, ce filtre a même des propriétés encore plus intéressantes : si p est le nombre de moments nuls de l'ondelette, la propriété vérifiée par ce filtre est :

$$\sum_n g[n]n^k = 0 \quad \forall\, 0 \leq k \leq p-1 \qquad (6.22)$$

Le filtre g est donc orthogonal à tout signal s'exprimant (sur son support) comme un polynôme de degré $p-1$.

Lorsque les corrélations, autour d'un point n donné, sont bien approchées par un polynôme de degré $p-1$, cette propriété induit que :

$$\sum_m g[m-2k] \operatorname{cov}\left(a^{j+1}[n], a^{j+1}[m]\right) \simeq 0. \tag{6.23}$$

Par conséquent, si pour tous les points n de l'équation (6.20) cette propriété est vérifiée, il en résulte que la variance du coefficient de détail $d^j[k]$ est faible.

Remarque 6.4

Le filtre correspondant au calcul du gradient dans une direction est aussi de moyenne nulle. C'est cette propriété qui permet de filtrer une grande partie du bruit en passant dans l'espace des gradients lorsque le bruit est corrélé en espace.

Effet des occultations

Dès lors que certains pixels viennent à manquer, le bruit ne peut plus être approché de manière correcte par un polynôme sur le support de l'ondelette : les pixels occultés ne sont en effet pas perturbés.

Comme dans le cas des gradients, l'absence d'un pixel produit deux effets contradictoires lors du calcul d'un coefficient de détail :

- Un pixel occulté ne porte pas plus de signal que d'erreur (au vu de la définition de la fonction coût (6.11)). Le coefficient d'ondelette calculé en tenant compte de ce coefficient peut donc voir la variance de son erreur décroître car un certain pourcentage d'information est "parfaite" ;
- Les corrélations entre l'erreur effectuée en ce pixel et celles de ses voisins ne se compensent plus lors du calcul du coefficient de détail. Selon les cas, il en résulte que la variance de l'erreur du coefficient d'ondelette peut fortement augmenter.

6.4 Modélisation de la variance en présence d'occultation

Pour prendre en compte l'action de ces deux effets contradictoires, nous proposons une correction de la variance des coefficients dont le support est partiellement occulté. Celle-ci s'écrit sous la forme :

$$\sigma_{occ}^2 = \left(\sigma^2 + \beta\sigma_m^2\right) I^2 \tag{6.24}$$

avec $0 \leq \beta \leq 1$ et $0 \leq I \leq 1$.

1. Le terme β correspond à un terme d'inflation de la variance. Il est multiplié par la variance moyenne, notée σ_m^2, correspondant à la variance du coefficient d'ondelette lorsque le bruit est décorrélé en espace. C'est donc un terme de "rappel" vers l'écriture de la variance lorsque le signal est décorrélé. Lorsque le bruit est "lisse", ce terme de rappel est dû au fait que les propriétés liées aux moments nuls de l'ondelette ne sont plus respectées.
2. Le terme I correspond à un terme de déflation de la variance. Ce terme est lié au pourcentage de pixels visibles lors du calcul du coefficient d'ondelette.

6.4.1 Pourcentage d'information disponible

Pour modéliser le terme I de l'équation (6.24), nous pouvons nous baser sur le pourcentage d'information disponible lors du calcul d'un coefficient d'ondelettes.
Différentes définitions de ce pourcentage d'information peuvent être envisagées. Le terme I peut :

- correspondre au pourcentage du support observé. Ce pourcentage prend uniquement en compte la présence ou l'absence d'un pixel. Il ne tient pas compte de l'importance de ce pixel dans le calcul du coefficient d'ondelette. Ce pourcentage est nommé I_0.
- tenir compte que la présence de certains pixels est plus importante que celle d'autres lors du calcul d'un coefficient d'ondelette particulier. Pour cela, un terme de pondération basé sur l'importance (en norme L^1) de chaque pixel dans le calcul du coefficient d'ondelette peut être utilisé. Ce pourcentage d'information est noté I_1.

Remarque 6.5

Ces deux pourcentages peuvent être approchés de manière simple en tirant profit de l'algorithme permettant d'effectuer la transformée en ondelettes.

Mémo 13 – Calcul des coefficients d'ondelettes

Les formules de décompositions dans des bases d'ondelettes sont les suivantes (en $1D$) :

$$c^k[j-1] = \sum_{n\in\mathbb{Z}} h[n-2k]c^n[j] \tag{6.25}$$

$$d^k[j-1] = \sum_{n\in\mathbb{Z}} g[n-2k]c^n[j] \tag{6.26}$$

où h (resp. g) est un filtre dont le support est $[n_1; n_2]$ (resp. $[m_1; m_2]$).

Pourcentage d'information basé sur le support de l'ondelette

Ici, chaque pixel a le même poids, qu'il contribue de manière importante ou faible au calcul du coefficient d'ondelette.

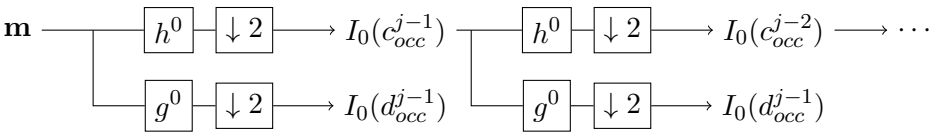


FIGURE 6.13 – Illustration de la manière d'obtenir le pourcentage d'information I_0. Ce pourcentage est calculé par une cascade de convolution décimation.

Une manière simple d'obtenir le pourcentage du support "observé" lors du calcul d'un coefficient d'ondelette est d'effectuer une transformée du masque **m**. Cette transformation est basée sur l'algorithme de la transformée en ondelettes (voir figure 6.13) mais met en jeux des filtres différents. Ces filtres, nommés h^0 et g^0, s'écrivent de la manière suivante :

$$\begin{aligned} h^0[n] &= \frac{1}{(n_2-n_1+1)} \quad \forall n \in [n_1; n_2] \\ g^0[n] &= \frac{1}{(m_2-m_1+1)} \quad \forall n \in [m_1; m_2] \end{aligned}$$

Remarque 6.6

Les supports de ces filtres sont identiques à ceux utilisés dans la décomposition en ondelettes (6.25) et (6.26). Cependant, ici, chaque élément des filtres h^0 et g^0 a le même poids. La somme des éléments de chacun des filtres vaut 1.

Illustration

En $1D$, lorsque le signal vit dans V^{j+1}, à partir de la connaissance du masque **m** (défini

en équation (6.2)), nous pouvons écrire le pourcentage du support observé d'un coefficient de détail dans V^j de la manière suivante :

$$I_0(d^j_{occ}[k]) = \sum_{n\in\mathbb{Z}} g^0[n-2k]\mathbf{m}[n]. \tag{6.27}$$

Celui d'un coefficient d'approximation peut s'écrire comme :

$$I_0(c^j_{occ}[k]) = \sum_{n\in\mathbb{Z}} h^0[n-2k]\mathbf{m}[n]. \tag{6.28}$$

Une fois ces pourcentages calculés, le pourcentage de support observé des coefficients de détails dans V^{j-1} se calcule alors sous la forme :

$$I_0(d^{j-1}_{occ}[k]) = \sum_{n\in\mathbb{Z}} g^0[n-2k]I_0(c^j_{occ}[n]). \tag{6.29}$$

Et ainsi de suite ...

Remarque 6.7

On peut remarquer au travers de ces calculs qu'un pixel occulté peut être comptabilisé plusieurs fois dans le support d'un coefficient à l'échelle $j-1$. Ceci se produit quand un pixel est utilisé dans le calcul d'au moins deux coefficients d'approximation à l'échelle précédente (par exemple dans $c^j_{\mathbf{m}}[n]$ et de $c^j_{\mathbf{m}}[n-1]$). Ce comptage multiple est lié à la taille du support des ondelettes : plus le support est grand, plus le nombre de comptages multiples est important. Pour toutes les bases d'ondelettes, à l'exception de celle de Haar, ce phénomène intervient. Ceci est dû au fait que seule la base de Haar possède des fonctions d'échelle ϕ dont le support est disjoint.

Pourcentage d'information L^1

Dans les faits, certains pixels contribuent plus que d'autres au calcul d'un coefficient d'ondelette. Pour prendre en compte ceci, il est possible de définir deux nouveaux filtres basés sur la norme L^1 des coefficients des filtres h et g. Ces filtres, notés h^1 et g^1, sont définis par :

$$h^1[n] = \frac{|h[n]|}{\sum_{n_1}^{n_2}|h[n]|} \quad \forall n \in [n_1; n_2] \tag{6.30}$$

$$g^1[n] = \frac{|g[n]|}{\sum_{m_1}^{m_2}|g[n]|} \quad \forall n \in [m_1; m_2] \tag{6.31}$$

Ils permettent, en effectuant la transformée du masque $\mathbf{m}$ par l'algorithme présenté en figure 6.13, de connaître le pourcentage d'information en norme L^1 disponible pour chaque coefficient.

Différence entre ces deux pourcentages d'information

La figure 6.14 présente les deux pourcentages d'information définis précédemment pour un nuage de taille 10 lorsqu'une décomposition de l'image sur deux niveaux est effectuée dans la base d'ondelettes de Daubechies à 4 moments nuls.

On observe que le pourcentage I_0 est beaucoup plus lisse que I_1. Le pourcentage en norme L^1 est bien plus concentré sur quelques points. Bien que visualisable aussi sur les bords du nuage, cette différence de comportement est particulièrement flagrante pour les yeux et la bouche du

nuage. En effet, ces derniers ne sont plus distinguables dès la seconde échelle des coefficients de détail pour le pourcentage I_0.

Certains coefficients (décrivant le signal proche des yeux du nuage) ont un très faible pourcentage d'information I_1 tandis que leur pourcentage d'information I_0 est élevé. Cette différence est due à la décroissance rapide en norme L^1 des coefficients des filtres h et g.

Contrairement au pourcentage d'information I_0, le pourcentage d'information I_1 varie selon l'orientation des coefficients de détails (à une même échelle et position).

Le pourcentage I_1 permet donc de mieux prendre en compte la spécificité de chaque coefficient d'ondelette. Cependant la variance induite par l'utilisation de ce pourcentage peut varier très rapidement entre coefficients voisins ce qui peut poser quelques problèmes de convergence.

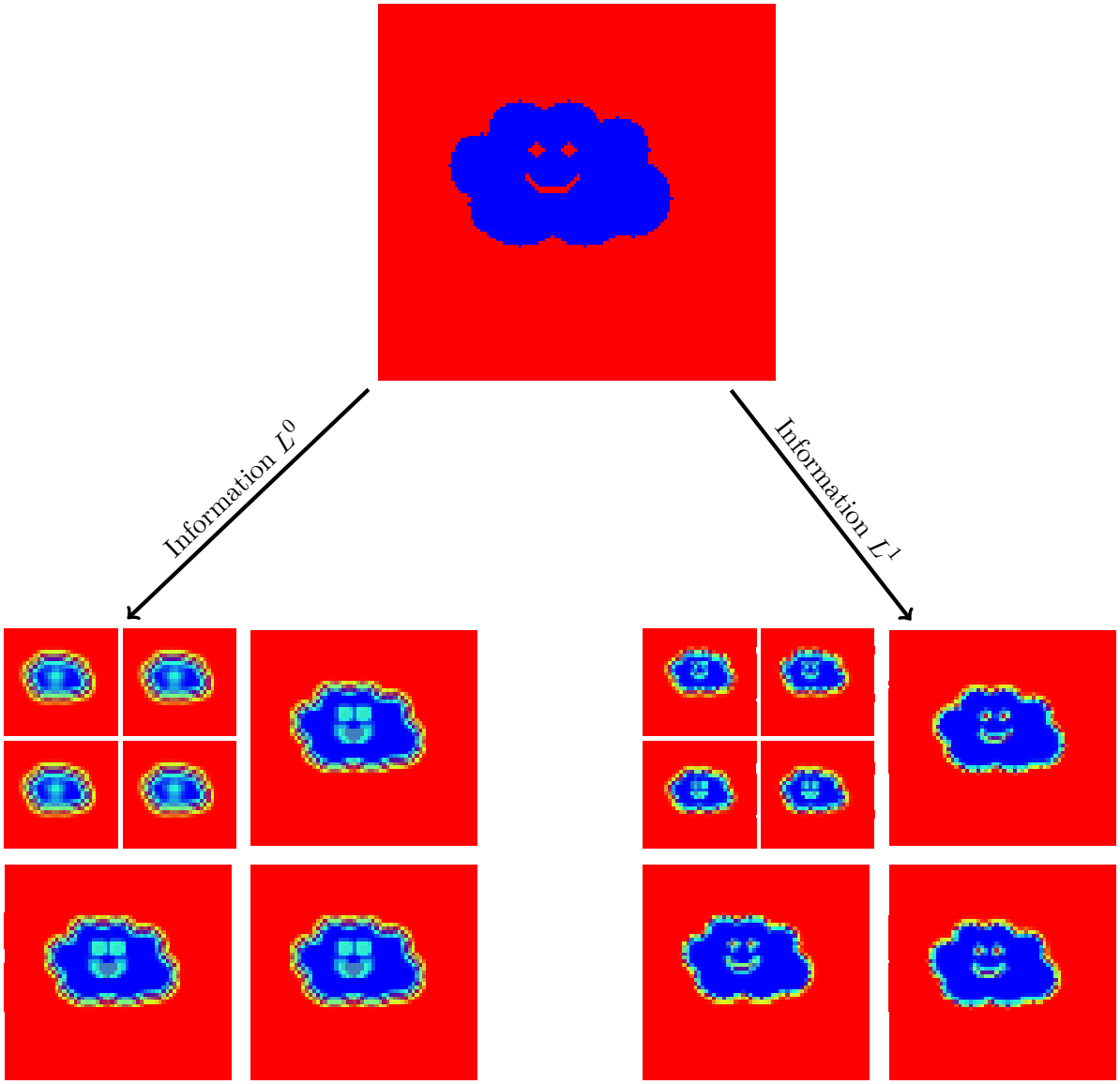


FIGURE 6.14 – Pourcentages d'information I_0 et I_1 calculés à partir du nuage de taille 10. Ces pourcentages vont de 0% (bleu) à 100% (rouge) .

6.4.2 Calcul du terme d'inflation

Pour tenir compte du fait que le schéma de corrélation n'est plus lisse mais comporte des discontinuités, nous pouvons tenter d'évaluer l'inflation à appliquer à la variance de chaque

coefficient de détail. Cette inflation est due au fait que les erreurs affectant les pixels utilisés lors du calcul d'un coefficient d'ondelette ne se compensent plus.

Lorsque les coefficients ne sont pas tous observés, effectuer la transformée en ondelettes d'une image masquée revient à utiliser des fonctions d'ondelettes n'ayant plus aucun moment nul sur certaines zones de l'image. En effet, le calcul d'un coefficient de détail, noté $d^j_{occ}[k]$, à l'échelle la plus fine s'écrit sous la forme :

$$d^j_{occ}[k] = \sum g[n-2k]c^{j+1}_{occ}[n] \tag{6.32}$$

où $c^{j+1}_{occ}[n]$ est défini par :

$$c^{j+1}_{occ}[n] = \begin{cases} c^{j+1}[n] & \text{si le pixel en position } n \text{ est observé} \\ 0 & \text{autrement} \end{cases} \tag{6.33}$$

Ce coefficient peut donc tout aussi bien s'écrire sous la forme :

$$d^j_{occ}[k] = \sum g^{occ}[n-2k]c^{j+1}[n] \tag{6.34}$$

où le filtre g^{occ} est tel que :

$$g^{occ}[n-2k] = \begin{cases} g[n-2k] & \text{si le pixel en position } n \text{ est observé} \\ 0 & \text{autrement} \end{cases} \tag{6.35}$$

Dès lors, pour savoir si une partie des erreurs d'observations se compense toujours, il est nécessaire de savoir si le filtre g^{occ} possède encore un certain nombre de moments nuls.

Remarque 6.8

En pratique, dès qu'un coefficient vient à manquer, ce filtre ne possède bien souvent plus aucun moment nul. Cependant, savoir si ce filtre est proche ou non d'avoir un moment nul d'ordre 0 (et dans l'absolu d'ordres plus élevés) permet d'évaluer grossièrement l'influence des occultations sur la compensation des erreurs d'observations.

Pour savoir si notre ondelette possède encore au moins un moment nul, il faut donc évaluer $\sum g^{occ}[n-2k]$. Si cette somme vaut 0, l'ondelette utilisée possède encore un moment nul.

Le coefficient d'inflation β intervenant dans l'équation (6.24) est choisi pour refléter l'écart à la condition de moment nul d'ordre 0 du filtre g^{occ} (équation (6.21)).
Pour un coefficient de détail à l'échelle la plus fine, ce coefficient est donné par :

$$\beta^j[k] = \frac{|\sum g^{occ}[n-2k]|}{\sum |g^{occ}[n-2k]|} \tag{6.36}$$

Remarque 6.9

Cette définition suppose que le support de l'ondelette comporte au moins un pixel observé. Si ce n'est pas le cas, alors ce coefficient n'est pas assimilé et nous n'avons donc pas besoin de définir sa variance.

Le coefficient d'inflation $\beta^j[k]$ est construit de manière à prendre en compte le fait que l'occultation de certains pixels peut se compenser. Il est nul lorsque l'ensemble des coefficients est observé ou lorsque le filtre g^{occ} possède un moment nul d'ordre 0. Son maximum est atteint lorsque seuls les coefficients positifs (ou négatifs) du filtre sont utilisés dans (6.34). Dans ce cas, il vaut 1.

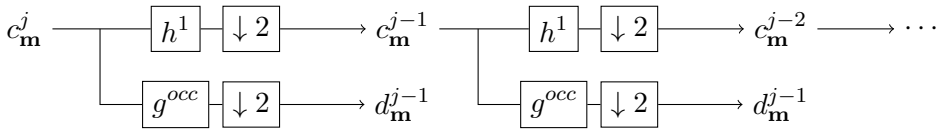


FIGURE 6.15 – Illustration de la manière d'obtenir les coefficients d'inflation $\beta^j[k]$. Ils sont calculés à partir du masque par une cascade de convolution décimation.

Obtention du facteur d'inflation aux échelles plus grossières

Les calculs précédents sont valables pour l'échelle la plus fine. Pour calculer ce facteur d'inflation nous allons, comme pour le pourcentage d'information, utiliser l'algorithme de la transformée en ondelettes avec d'autres filtres.

Le filtre permettant d'obtenir les coefficients "d'approximation" est h^1 (voir équation (6.30)). Il permet de savoir le pourcentage d'information I_1 disponible dans un tel coefficient. Le filtre permettant d'obtenir l'inflation à appliquer aux coefficients de détails est noté g^{occ} et est défini par :

$$g^{occ}[n] = \frac{g[n]}{\sum |g[n]|} \tag{6.37}$$

La figure 6.15 illustre les opérations à effectuer pour obtenir les coefficients d'inflations.

Les coefficients $\beta^j[k]$ sont donnés pour chaque coefficient de détail de l'image originale $d^j_{occ}[k]$ par :

$$\beta^j[k] = |d^j_{\mathbf{m}}[k]|. \tag{6.38}$$

6.4.3 Approximation de la variance de coefficients partiellement occultés

En résumé, la variance d'un coefficient de détail partiellement occulté , $d^j_{occ}[k]$, est définie via le modèle proposé en équation (6.24) :

$$\text{var}\left(d^j_{occ}[k]\right) = (\sigma^2 + \beta^j[k]\sigma^2_m) I(d^j_{occ}[k])^2 \tag{6.39}$$

Ce modèle combine le terme de déflation I dont différentes définitions sont présentées en section 6.4.1 et un terme d'inflation $\beta^j[k]$ présenté en section 6.4.2.
Concernant les coefficients d'approximation, seul le terme de déflation est utilisé. Leur variance est donc définie par :

$$\text{var}\left(c^j_{occ}[k]\right) = \sigma^2 I(c^j_{occ}[k])^2. \tag{6.40}$$

Aucun terme d'inflation n'est utilisé (la variance de ces coefficients est surestimée lorsque les images sont occultées).

Cependant, la modélisation proposée de l'effet des occultations reste une modélisation grossière. Par conséquent, pour éviter un problème de trop mauvaise estimation de la variance, lorsqu'un coefficient contient trop peu d'information un seuillage des coefficients est effectué.

Ce seuillage, ne prend pas en compte la valeur du coefficient de détail mais simplement son pourcentage d'information en norme L^1 :

$$c^j_{occ}[k] = \begin{cases} c^j_{occ}[k] & \text{si } I_1(c^j_{occ}[k]) > 10\% \\ 0 & \text{autrement} \end{cases} \tag{6.41}$$

Expérience effecutée

But : Evaluer l'impact du changement de modélisation de la variance dans un espace

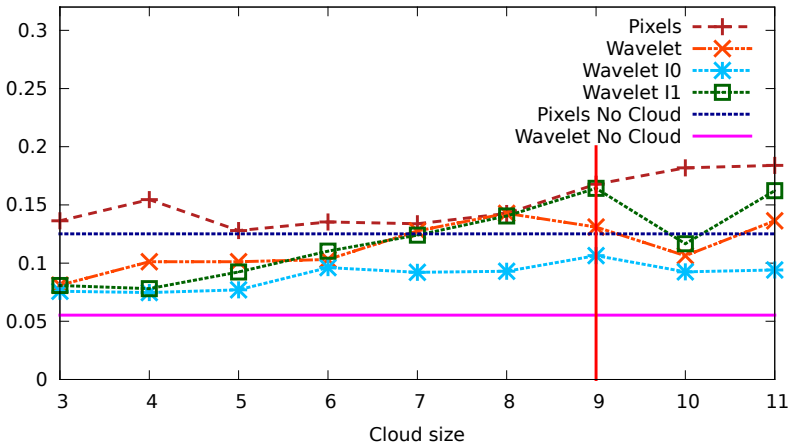
d'ondelettes en présence d'occultation.

Matrices utilisées : Matrice non modifiée, modifiée en utilisant l'information I_0, modifiée en utilisant l'information I_1 dans un espace d'ondelettes.

Observation utilisées : Des séquences d'images perturbées par un bruit corrélé en espace partiellement occultées.

Autres paramètres : La longueur de corrélation de l'erreur d'ébauche est de 7.65 (pixels). La longueur de la fenêtre d'assimilation est de $2h24$. Une observation est utilisée toutes les 6 minutes.

Séquence peu bruitée (SNR = 26.8 dB)



Séquence fortement bruitée (SNR = 14.8 dB)

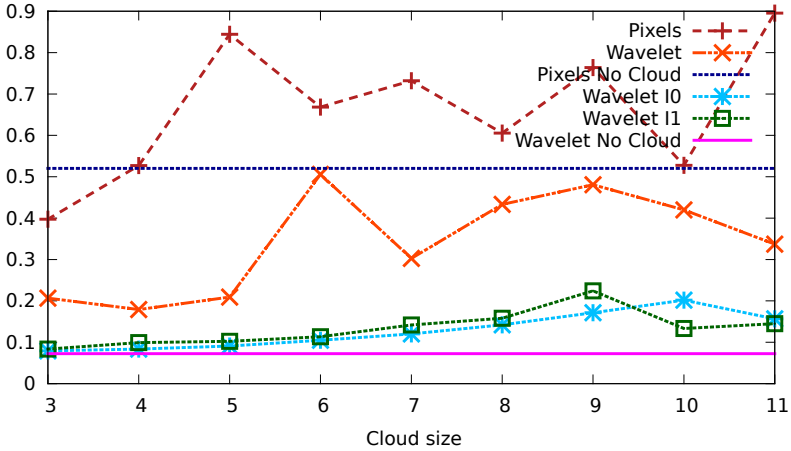


FIGURE 6.16 – Pourcentage d'erreur résiduelle sur la composante méridionale du champ de vitesse. La décroissance de cette erreur au cours des itérations de minimisation est présentée en figure 6.18 pour les expériences situées sur la ligne rouge.

La figure 6.16 présente les résultats obtenus en utilisant ces nouvelles définitions de la variance lors de l'utilisation d'une séquence d'images peu bruitée (en haut) et fortement bruitée (en bas).

Les résultats utilisant le pourcentage d'information I_0 sont présentés en bleu tandis que ceux utilisant le pourcentage I_1 sont présentés en vert.

– Lorsque l'amplitude du bruit est importante, les résultats obtenus par ces deux modélisa-

tions sont bien meilleurs. La prise en compte d'une partie de l'impact dû aux occultations est donc positive. Les deux manières de définir le pourcentage d'information fournissent des résultats de qualité comparable.

– Lorsque l'amplitude du bruit est faible, l'impact de la prise en compte des occultations lors de la définition de la variance est moins importante (car l'impact du bruit l'est aussi). L'impact du changement de variance est toujours positif lorsque le pourcentage d'information I_0 est utilisé mais peut être négatif pour l'autre pourcentage d'information (I_1).

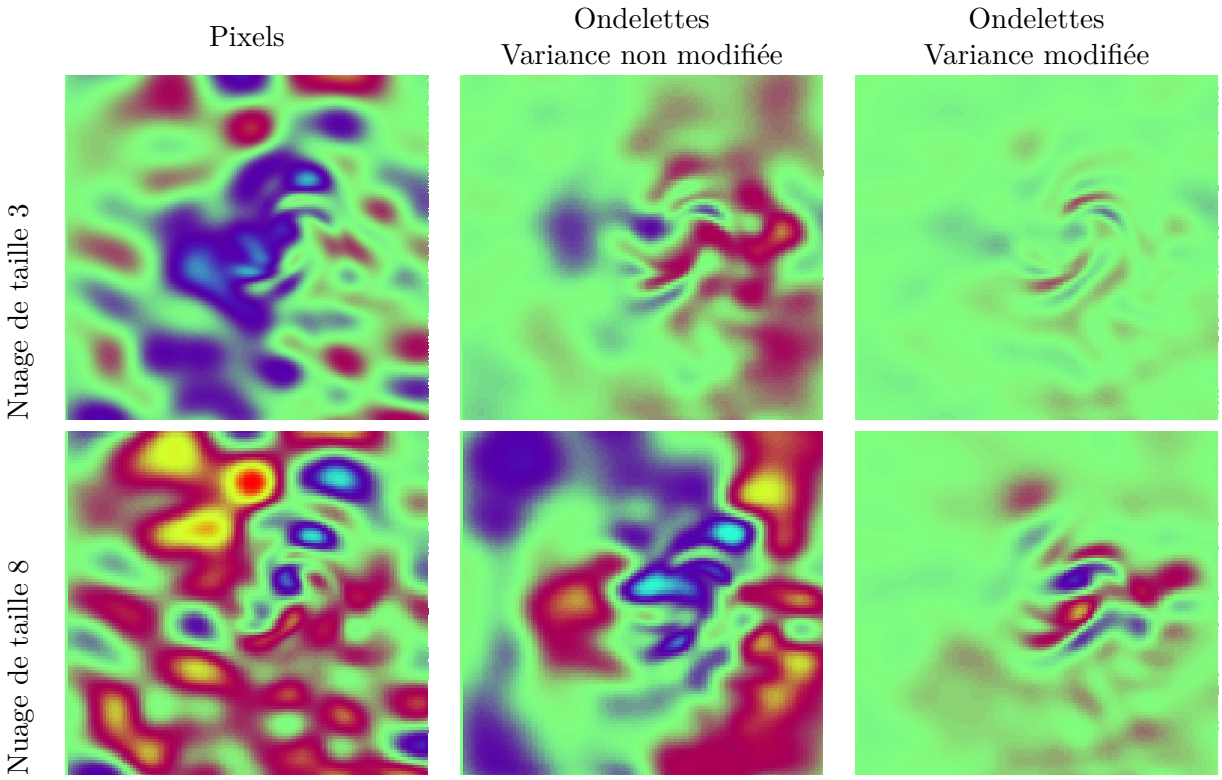


FIGURE 6.17 – Erreur sur la composante zonale de la vitesse lorsque la séquence d'images est fortement bruitée (SNR =14.8 dB). Echelle : $-0.01\ m.s^{-1}$ (en bleu) à $0.01\ m.s^{-1}$ (en rouge). A titre de comparaison les valeurs de cette composante vont de $-0.028\ m.s^{-1}$ à $0.028\ m.s^{-1}$. La modification de la variance est effectuée via l'équation (6.24). Les résultats présentés sont ceux utilisant le pourcentage d'information I_0.

La figure 6.17 présente l'erreur effectuée sur la composante zonale du champ de vitesse selon si la variance dans une base d'ondelettes est modifiée ou non. On observe que la répartition spatiale des erreurs dans le cas où la variance est modifiée est plus conforme à celle attendue au vu des résultats obtenus avec des données parfaites (figure 6.4). Lorsque le nuage est petit, cette dernière présente la structure observée avec des données non occultées. Lorsque le nuage grandit, l'erreur se concentre sous le nuage et est due à une sous-estimation de l'intensité du vortex.

Vitesse de convergence de la minimisation

Lorsque le bruit est de faible amplitude, l'impact mitigé de la modification de la définition de la variance est en partie dû à un problème de convergence de la minimisation.

Expériences effectuées

But : Illustrer le problème de convergence.

Observations utilisées : Une séquence d'image peu perturbée (SNR = 26.8 dB). Le nuage est de taille 9. Ces expériences sont celles situées sur la ligne rouge de la figure 6.16.

A noter : Les erreurs à convergence ne correspondent pas à celles observées en figure 6.17. En effet, le niveau de bruit considéré n'est pas le même : ici il est faible alors que précédemment il était important.

La figure 6.18 présente la décroissance de l'erreur résiduelle de la composante méridionale de la vitesse en fonction de l'itération de minimisation.

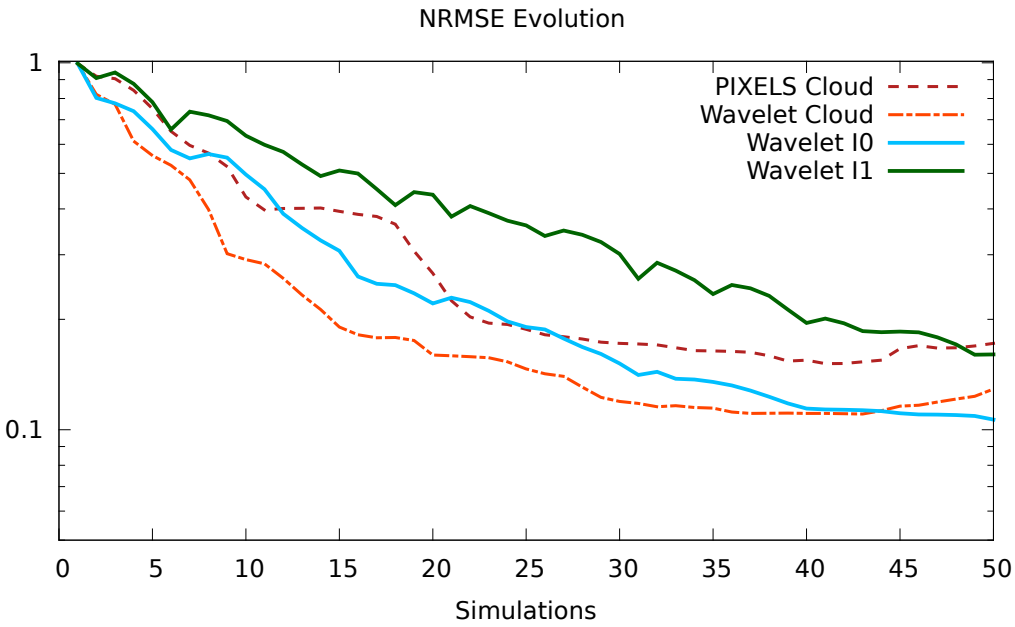


FIGURE 6.18 – Evolution du pourcentage d'erreur résiduelle de la composante méridionale du champ de vitesse en fonction des itérations de minimisation. La nuage occultant le vortex est de taille 9. Le niveau de bruit est faible (SNR = 26.8 dB).

Lorsque la variance est modifiée en employant le pourcentage d'information I_1, la vitesse de convergence est bien plus lente. Ceci est moins le cas lorsque le pourcentage I_0 est employée.

Remarque 6.10

Nous pouvons aussi observer le problème rencontré par la distance entre pixels : après un certain nombre d'itérations où l'erreur décroît, cette dernière tend à augmenter. Ce problème est aussi présent pour la distance dans un espace d'ondelettes ne tenant pas compte de la présence d'occultations lors de la définition de la variance. Ces deux distances tendent donc à considérer qu'une partie du bruit présent dans les images fait partie intégrante du signal et essaient donc de s'en approcher. Ce n'est pas le cas des des distances modifiant la variance des coefficients d'ondelettes en présence d'occultation.

Accélération de la vitesse de convergence

Afin d'accélérer la convergence de la minimisation, nous pouvons effectuer une première étape de minimisation en ne tenant pas compte du signal aux échelles les plus fines de la décomposition d'ondelettes. Puis, après s'être approché du véritable état, effectuer une seconde étape de minimisation en prenant en compte l'ensemble de l'information.

La figure 6.19 présente la décroissance de l'erreur résiduelle en fonction du nombre d'itérations. Les minimisations correspondant aux courbes orange et rose n'utilisent pas, dans un premier temps, l'information contenue à l'échelle la plus fine. Cette information n'est intégrée qu'après 30 itérations.

On observe que ne pas utiliser l'information contenue à l'échelle la plus fine permet, dans un premier temps, d'accélerer fortement la rapidité de la minimisation. Lorsque le changement de variance fait intervenir le pourcentage d'information en norme L^1, effectuer 20 itérations sans utiliser l'information à l'échelle fine permet d'obtenir des résultats équivalents à ceux obtenus après 55 itérations en utilisant l'ensemble de l'information. Il y a donc un gain important en terme de rapidité de minimisation.

Pour la pseudo-métrique utilisant le terme de pondération issue du pourcentage d'information en norme L^0, le bénéfice de la non incorporation des détails à l'échelle la plus fine lors des premières itérations est plus mitigé. En effet, on observe une décroissance plus rapide de l'erreur sur la composante méridionale de la vitesse au cours des 15 premières itérations. Cependant, au bout de 30 itérations le résultat est moins bon qu'en utilisant l'ensemble de l'information. L'analyse effectuée avec ces termes de pondération est moins robuste au bruit, ce qui induit une croissance des erreurs durant les dernières itérations de cette première phase. Ces erreurs ne sont pas bien corrigées pendant la seconde phase incorporant l'information aux échelles les plus fines. Pour cette manière de définir la variance, l'incorporation de l'information en deux phases a pour résultat une analyse de moins bonne qualité au bout de 100 itérations que celle obtenue avec 50 itérations en incorporant l'ensemble de l'information.

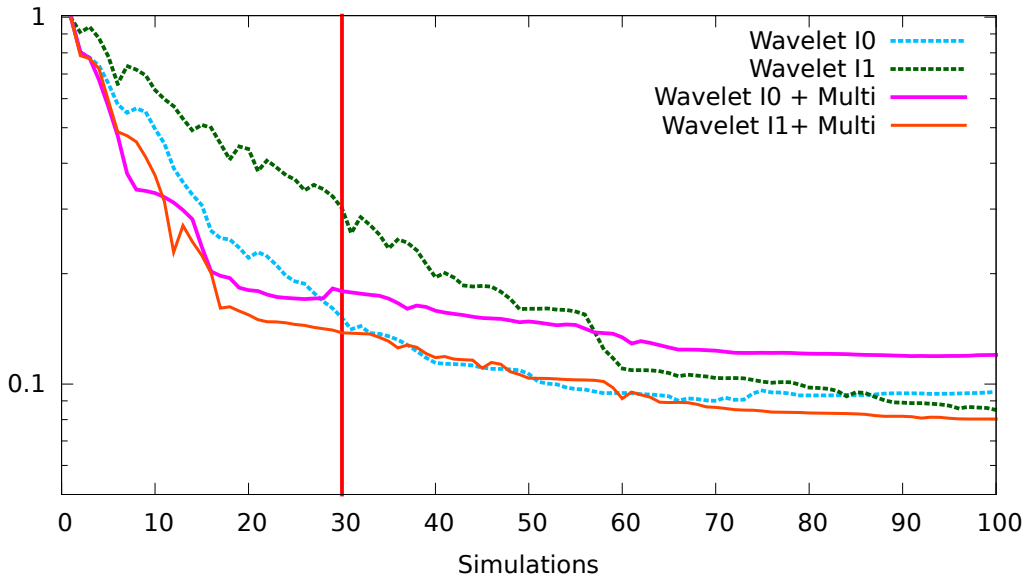


FIGURE 6.19 – Evolution du pourcentage d'erreur résiduelle de la composante méridionale du champ de vitesse en fonction des itérations. Le nuage occultant le vortex est de taille 9. Le niveau de bruit est faible (SNR = 26.8 dB). La ligne rouge présente la dernière itération effectuée sans prendre en compte l'information à l'échelle la plus fine dans une base d'ondelettes.

L'ensemble des expériences précédentes a été effectué une nouvelle fois en incorporant l'information en deux phases lorsque le pourcentage d'information d'un coefficient d'ondelette est définie en norme L^1.

Effectuer l'analyse en deux phases avec ce nombre d'itérations (50) n'améliore pas, en moyenne, l'erreur résiduelle sur la composante méridionale du champ de vitesse.

Cependant, en figure 6.20, on observe que la répartition des erreurs n'est pas la même lorsque ces deux manières différentes d'intégrer l'information sont utilisées. En effet, lorsque l'information est intégrée en deux phases, une erreur importante apparaît après la première phase en dehors de la zone du vortex. L'erreur en fin de seconde phase est moins importante au coeur du vortex que lorsque l'information est intégrée en une phase unique. Cependant, cette erreur est plus importante à l'extérieur du vortex : l'erreur effectuée en première phase n'a pas eu le temps d'être corrigée dans cette région.

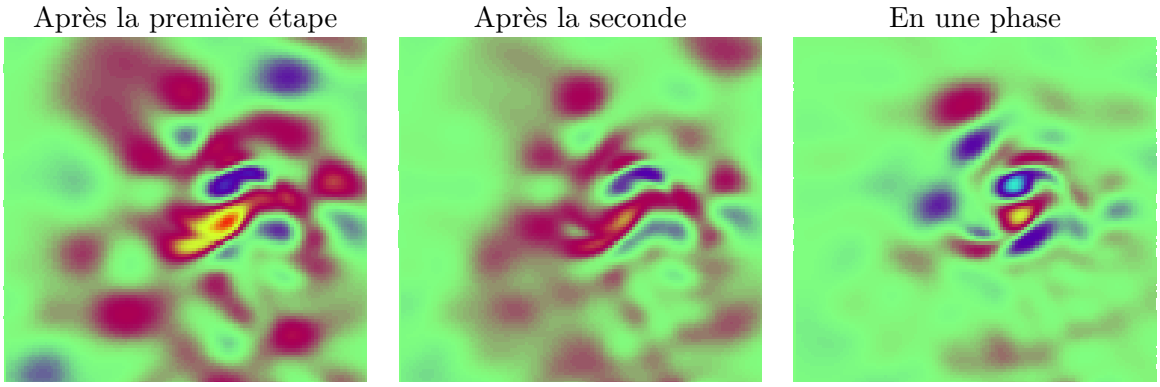


FIGURE 6.20 – Répartition spatiale de l'erreur sur la composante zonale de la vitesse. L'image de gauche présente l'erreur après une première phase d'assimilation (25 itérations) au cours de laquelle l'information à petite échelle n'a pas été utilisée. Celle du centre présente l'erreur après la seconde phase (25 itérations) prenant en compte l'information à l'échelle la plus fine. L'image de droite présente l'erreur lorsque l'assimilation est effectuée en une phase unique (50 itérations).

6.5 Conclusion

Dans ce chapitre, nous avons abordé le problème d'assimilation d'images occultées. Nous avons pu voir que, comme rapporté par différents auteurs [Papadakis, 2007, Korotaev et al., 2008, Beyou, 2013], l'absence de certains pixels ne pose pas particulièrement de problème lorsque les données sont parfaites ou perturbées par un bruit décorrélé en espace (les résultats concernant ces dernières expériences n'ont pas été présentés). Dans ce cas, les images peuvent en effet être utilisées dans l'espace des pixels en association avec une matrice de corrélation proportionnelle à l'identité ou dans un espace d'ondelettes (mais dans ce cas, si aucun seuillage n'est effectué, la taille du vecteur d'observation est plus importante).

Cependant, dès lors qu'on souhaite modéliser les corrélations entre les erreurs effectuées en différents pixels, on doit faire face à un nouveau problème : il est nécessaire d'adapter la matrice de corrélations aux données réellement observées.

Dans ce chapitre, le cas de l'adaptation de l'hypothèse d'une matrice de covariance diagonale dans un espace d'ondelettes au cas où les images sont partiellement observées a été abordé. Nous avons vu qu'il est nécessaire de prendre en compte l'absence de certains pixels lors de la définition de la variance de l'erreur associée à chaque coefficient d'ondelettes. Si cela n'est pas fait, les résultats obtenus sont assez décevant (en comparaison de ceux obtenus dans le chapitre 5). Afin de ne pas avoir à recalculer l'ensemble des variances en partant de la vraie matrice de covariance dans l'espace des pixels, une modélisation grossière de l'effet des occultations a été proposée. Cette modélisation est peu coûteuse (deux transformées en ondelettes pour chaque image) mais permet d'obtenir des résultats satisfaisants.

7

Conclusion et perspectives

7.1 Conclusion

Au cours de ces travaux portant sur l'assimilation d'images, nous avons montré l'utilité de représenter le signal dans des espaces multi–échelles afin d'exploiter plus facilement les caractéristiques de ces derniers.

Les points abordés directement on été :

- **La représentation les erreurs :**
  En travaillant dans un autre espace, il est en effet possible de décrire les erreurs de manière différente. Ces dernières ne sont plus décrites au niveau du pixel mais au niveau de la structure. Au chapitre 5 nous avons vu, comme cela est connu depuis assez longtemps [Johnstone and Silverman, 1997], qu'un bruit corrélé en espace n'affecte pas toutes les fonctions dans des bases d'ondelettes de la même manière. Utiliser cette connaissance permet de représenter de manière simple une partie des corrélations présentes au sein du bruit. Nous avons aussi aperçu l'intérêt et les difficultés liées à l'utilisation d'espaces dans lesquels l'information est stockée de manière redondante (les curvelettes). Dans ce cas, différentes manières de définir le poids de chacune des structures dans la fonction coût ont été envisagées. La reformulation du problème envisagée en section 5.6 (on ne cherche pas à inverser la matrice de covariance des coefficients dans la trame des curvelettes mais à décrire de manière optimale l'inverse de la matice de covariance dans l'espace des pixels en travaillant dans un espace de plus grande dimension) ouvre des pistes possibles pour une meilleure utilisation de tels espaces.
- **La gestion des occultations :**
  En assimilation de données pour la prévision numérique du temps, un des problèmes vient de la taille des espaces considérés. Pour les observations, à cause du fait que l'espace observé change d'un temps d'observation à un autre, il n'est pas simple de proposer une modélisation de la matrice inverse de covariance. En effet, cette modélisation doit pouvoir être adaptable aisément au fil du temps pour prendre en compte les informations réellement récoltées. Au chapitre 6, nous avons discuté quelque peu de ce problème et proposé une solution pour adapter une approximation de la matrice de covariance réalisée lorsque l'ensemble des observations sont présentes au cas où certaines viennent à manquer. Cette adaptation de la matrice de covariance d'erreur d'observation est peu coûteuse. Elle est effectuée dans une base d'ondelettes en utilisant une matrice diagonale et peu donc être aisément inversée.
- **Capacité de compression :**
  La capacité de compression des transformées en ondelettes a été plusieurs fois abordée. Cette compression est possible aussi bien dans le cadre où les observations sont perturbées par un bruit décorrélé en espace que corrélé en espace. Cependant, comme présenté au chapitre 5, la manière de compresser l'information doit tenir compte aussi bien des caractéristiques du signal que du bruit.

Un autre avantage lié à l'utilisation de transformée mutli-échelles apparait au cours de ce manuscrit. Il s'agit d'une plus **grande maniabilité** du signal dans ces espaces que dans l'espace des pixels. Cette dernière a par exemple été utilisée au chapitre 6 pour accélérer la convergence de la minimisation. Au chapitre 5, elle a été utilisée afin d'outrepasser le problème d'une trop grande distance entre les observations et les images synthétiques dû à une ébauche de trop mauvaise qualité.

Dans ces espaces multi–échelles, il est ainsi plus aisé, en fonction de nos connaissances sur l'ébauche, de savoir quel type d'information va jouer un rôle important lors de la minimisation et quand ce rôle important va avoir lieu. Par exemple, dans les expériences de la section 5.8.4, incorporer les détails sur les structures à une échelle fine s'est avéré contre–productif lorsque les structures à plus haute–échelle n'étaient pas bien recalées. En travaillant dans des espaces mutli–échelles, il est ainsi plus simple de prendre en compte ceci afin de ne pas embarrasser le processus de minimisation avec des informations non pertinentes pour le problème qu'il doit corriger en premier lieu.

7.2 Perspectives

7.2.1 Vers l'utilisation d'autres espaces multi–échelles

Passage en revue des transformées multi–échelles existantes

Ces quinze dernières années, un grand nombre d'espaces multi–échelles ont été introduit dans le domaine du traitement du signal et des images dans le but d'analyser, de manière plus aisée, un signal dense. Chaque espace possède des propriétés qui lui sont propres et permet de représenter de manière parcimonieuse différents types de signaux. Les bases d'ondelettes et les trames de curveletttes utilisées au cours cette thèse ne sont que la partie visible de l'iceberg. Savoir quels types de bruits peuvent être représentés par des matrices diagonales dans ces différents espaces est un enjeux important dans l'optique d'utiliser des matrices de covariance d'erreur d'observation aisément inversibles.

La majeure partie des algorithmes permettant d'effectuer ces transformées sont disponibles, soit sous forme de fonction matlab, soit en c++. Effectuer un travail de review sur les possibilités fournies par chacune de ces transformées est une continuité logique des travaux développés dans cette thèse. Au chapitre 5, pour les transformées en curvelettes, une formulation du problème à résoudre afin d'obtenir une matrice diagonale permettant de représenter au mieux une matrice de covariance pleine dans l'espace des pixels a été introduite. Cette formulation, à laquelle des termes de régularité spatiale sur la diagonale de la matrice peuvent être ajoutés aisément, peut contribuer à la création d'un cadre générique afin d'étudier les propriétés de l'ensemble de ces transformées sans pour autant avoir à les maîtriser dans le détail.

Combinaison de plusieurs représentations

En traitement du signal, la complémentarité des propriétés de ces différents espaces a permis d'aborder le problème de séparation d'un signal en différentes composantes selon leurs morphologies. De fait, un signal ne pouvant s'écrire de manière parcimonieuse dans aucun des espaces à disposition peut se révéler représentable de manière simple dans l'union de ces espaces. Ceci est illustré en figure 7.1 où un signal complexe s'écrit à partir de 4 fonctions s'il est représenté dans une union de base.

Différents travaux [Starck et al., 2004, Elad et al., 2005, Fadili et al., 2010b] sur des images réelles ou synthétiques ont montré l'intérêt d'utiliser des espaces complémentaires afin d'ex-

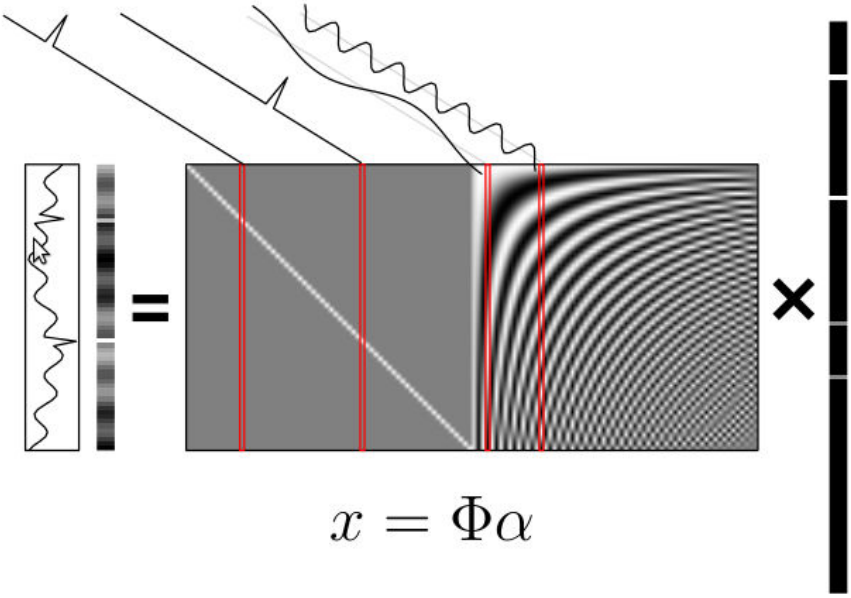


FIGURE 7.1 – Exemple d'un signal parcimonieux lorsque celui ci est décomposé dans l'union de deux bases(*dictionnaire*) $\phi = [I\ F]$ étant la concaténation de l'identité et de Fourier. Le signal exemple a quatre coefficients non nuls dans ce *dictionnaire*, alors que sa représentation serait pleine en Fourier comme dans l'espace direct. A gauche est présenté le signal considéré, au centre le dictionnaire, à droite, les éléments actifs de ce dictionnaire. Source : [Woiselle, 2010].

traire de manière parcimonieuse certains types de signaux [20]. La figure 7.2 présente un exemple de décompostion d'image en utilisant la parcimonie de l'écriture du signal dans deux espaces différents.

FIGURE 7.2 – A gauche l'image originale, au centre première composante morphologique (extraite dans une trame de curvelette), à droite seconde composante morphologique (extraite dans une base de cosinus engendrée par la transformée en cosinus discrète). Source : [Bobin et al., 2007].

Cette écriture parcimonieuse du signal dans des espaces redondants (engendrés par l'union d'espaces complémentaires) doit aussi pouvoir être exploitée afin de représenter à base de matrices diagonales certains bruits plus complexes. La fonction coût à minimiser peut en effet pouvoir être approchée correctement de la manière suivante :

$$(\mathbf{y}^{\mathbf{o}o} - H(x))^T \mathbf{R}_{pix}^{-1} (\mathbf{y}^{\mathbf{o}o} - H(x)) \simeq (\mathbf{y}^{\mathbf{o}o} - H(x))^T \left(\mathbf{A}_1^T \mathbf{D}_{\mathbf{A}_1} \mathbf{A}_1 + \mathbf{A}_2^T \mathbf{D}_{\mathbf{A}_2} \mathbf{A}_2\right) (\mathbf{y}^{\mathbf{o}o} - H(x)) \tag{7.1}$$

20. Un logiciel (en C++ ou matlab) décrit dans [Fadili et al., 2010a] est d'ailleurs à disposition sur la page de l'auteur afin d'appréhender l'intérêt de ces différentes transformées et de leur combinaisons.

où $\mathbf{A}_1$ et $\mathbf{A}_2$ sont deux transformées aux propriétés complémentaires. $\mathbf{D}_{\mathbf{A}_1}$ et $\mathbf{D}_{\mathbf{A}_2}$ sont dans ce cas deux matrices de pondérations diagonales.

Dans un sens, ce travail a déjà été initié durant cette thèse lorsque les transformées en curvelettes ont été employées. En effet, ces transformées décrivent les propriétés du signal de manière redondante et peuvent être vues comme l'union de différents dictionnaires.

7.2.2 Prise en compte des corrélations temporelles

La prise en compte des corrélations temporelles n'a pas été étudié dans ces travaux. Ces corrélations sont pourtant potentiellement présentes, notamment dans les images satellites. Elles peuvent provenir des erreurs de mesures (et du fait que l'information est récoltée par un même instrument) aussi bien que des erreurs de représentativité : si le modèle employé n'est pas capable de représenter un phénomène au temps t_k il est envisageable qu'il ne puisse pas représenter son évolution. Dès lors, l'erreur de représentativité au temps t_{k+1} dépend en partie de l'erreur de représentativité au temps t_k. Les erreurs de représentativité semblent donc par nature corrélées temporellement.

La figure 7.3 présente l'erreur résiduelle de l'analyse en fonction du nombre d'observations utilisées dans le cas où le bruit est décorrélé en espace (à gauche) et corrélé en espace (à droite). On observe que l'impact des corrélations temporelles est important. L'apport d'information des images supplémentaires est fortement réduit à cause des corrélations temporelles au sein du bruit.

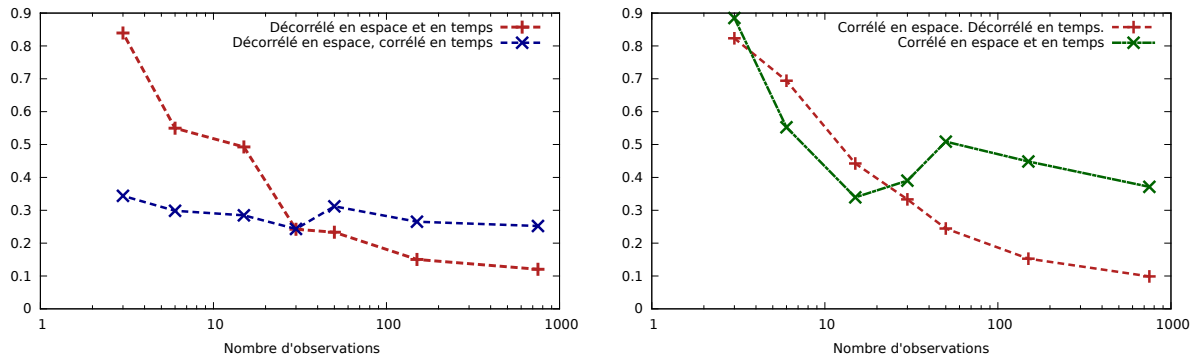


FIGURE 7.3 – Erreur résiduelle sur la composante méridionale de la vitesse en fonction du nombre d'observations utilisées. A gauche, le bruit est décorrélé en espace (le niveau de bruit est moins important lorsque celui–ci est corrélé en temps). A droite, le bruit est corrélé en espace (le niveau de bruit est le même que le bruit soit ou non corrélé en temps).

Une extension naturelle des travaux présentés dans ce manuscrit consiste donc à considérer des transformations comportant une composante temporelle afin de prendre en compte les corrélations en temps présentes au sein des champs d'erreurs. L'utilisation de transformées séparables (telles les ondelettes) en 3 dimensions combinée avec des matrices diagonales peut fournir une première approche peu coûteuse pour intégrer une partie de ces corrélations.
Utiliser des transformées directionnelles en plusieurs dimensions (telles les curvelettes $3D$ [Ying et al., 2005]) est certainement plus complexe à mettre en place. Cependant, ces espaces redondants doivent permettre de représenter "correctement" une plus grande variété de statistique d'erreur d'observation lorsqu'on se restreint à utiliser des matrices de pondération diagonales. Notamment, comme en deux dimensions, ces transformées doivent permettre de représenter des corrélations possèdant une direction privilégiée non orientée dans la direction des axes "naturels". Bien que le coût de calcul de telles transformées soit important, le fait que ces transformées soient localisées en espace permet d'envisager l'utilisation d'algorithmes parallélisés. Dans l'article introduisant

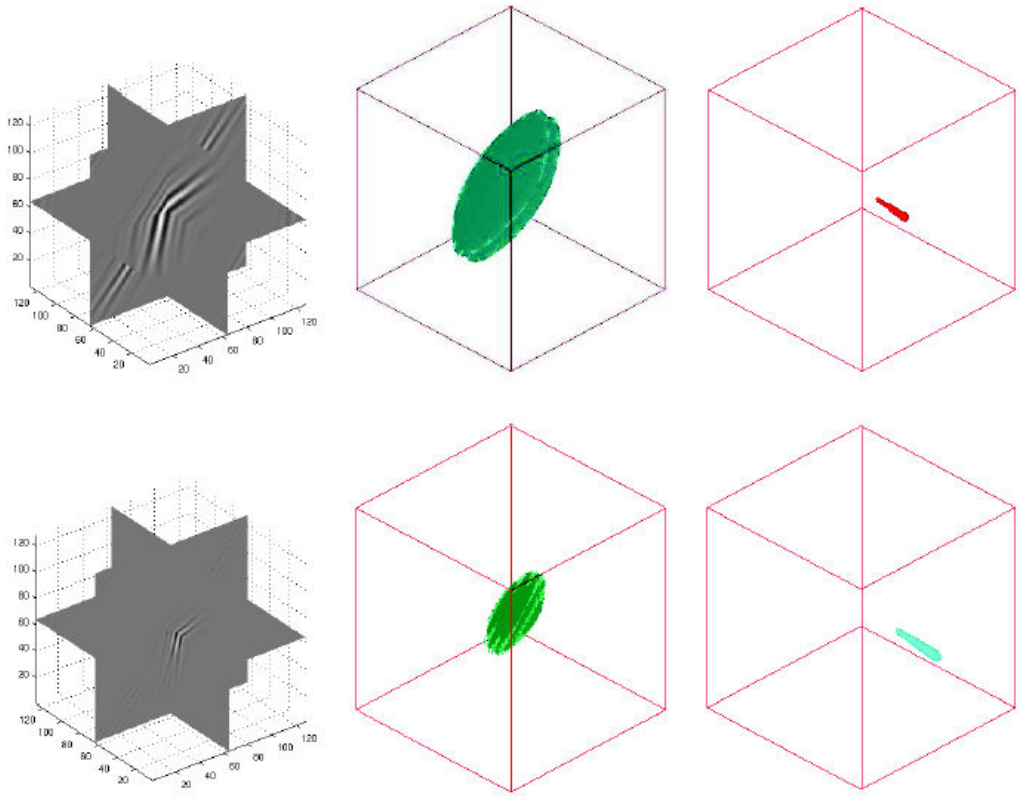


FIGURE 7.4 – A gauche un élément de curvelettes en 3 dimensions à l'échelle grossière (en haut) et à une échelle plus fine (en bas). Leur support dans le plan et en fréquence est présenté à droite. Source : [Ma et al., 2008]

les curvelettes en 3 dimensions, la transformée d'une image $1024 \times 1024 \times 1024$ pixels (soit environ 1 milliard de pixels) s'effectue en "seulement" 94 seconde si elle est distribuée sur 128 processeurs de 1.5 Ghz [Ying et al., 2005].

7.2.3 Gestion des occultations

Un des problèmes inhérents aux observations réside dans le fait que ces dernières ne sont pas distribuées dans le temps de manière homogène. En assimilation de données, prendre en compte les corrélations existant entre leurs erreurs est alors complexe : il est en effet nécessaire d'accéder à l'inverse de la matrice de covariance d'erreur d'observation, matrice dont les éléments changent d'un temps d'observation à un autre. Ce problème lié à l'inversion est renforcé par la taille du vecteur d'observations ($\sim 10^7$) rendant toute forme de stockage et d'inversion de la véritable matrice de covariance inenvisageable.

Le travail d'approximation réalisé pour inverser aisément une matrice de covariance d'erreur d'observation spécifique doit pouvoir être adapté à moindre frais à un espace d'observation changeant au gré du temps et des conditions climatiques.
Cette adaptation n'est cependant pas triviale :

- lorsque l'approximation est effectuée directement sur la matrice inverse afin d'avoir une matrice inverse creuse, comme cela est proposé dans [Stewart et al., 2013b], la projection de cette matrice inverse dans le sous–espace observé au temps t peut complètement modifier le modèle de corrélation direct.
- lorsque l'approximation est effectuée via un changement d'espace, comme proposé ici, prendre en compte l'absence de certains pixels n'est pas nécessairement aisé bien qu'une

méthode afin d'adapter grossièrement l'approximation de la matrice de covariance ait été proposée.

Plusieurs solutions peuvent cependant être envisagées :

1. La première consiste à poursuivre dans la voie ouverte par le chapitre 6. Elle consiste à adapter une matrice de covariance construite dans un espace de plus grande dimension au cas présent. Pour ce faire, l'impact des observations manquantes sur l'approximation effectuée peut par exemple être modélisé à moindre en coût en employant des coefficients d'inflations et/ou de déflations.
2. La seconde, proposée dans [Fisher, 2014], consiste à contourner le problème en utilisant d'autres méthodes d'assimilation de données ne nécessitant pas l'inversion de la matrice de covariance d'erreur d'observation[21].

Le problème de la gestion d'un espace d'observation évoluant au cours du temps est donc encore largement ouvert. Il est nécessaire de s'y intéresser de près si nous souhaitons être capable de prendre en compte les corrélations spatiales (et temporelles) présentes au sein des observations disponibles et ainsi cesser d'utiliser uniquement un très faible pourcentage de ces dernières.

7.2.4 Approximation par blocs diagonaux

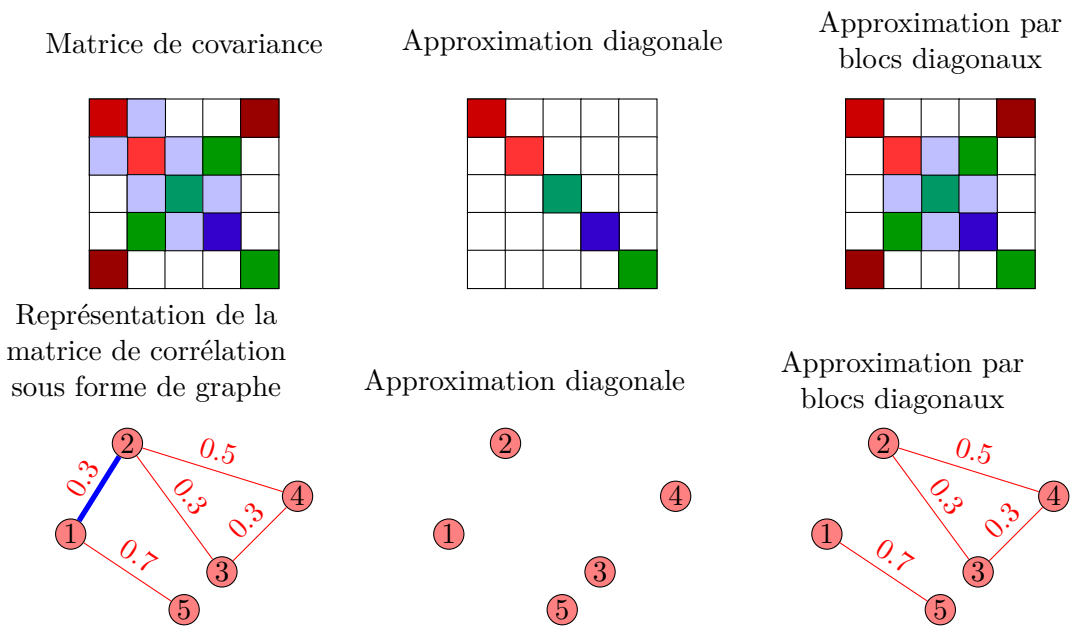


FIGURE 7.5 – Représentation des corrélations entres les différentes erreurs sous forme de graphe. A gauche, la matrice de corrélation complète, au centre l'approximation diagonale, à droite l'approximation de la matrice en supprimant seulement une arête du graphe.

Les approximations des matrices de covariances employées dans ces travaux sont des approximations diagonales effectuées après un changement de variables. Or une approximation diagonale, même effectuée dans d'autres espaces (base ou trame), reste bien souvent une approximation très grossière de la véritable matrice de corrélation.
Prendre en compte une partie de ces corrélations pourrait s'avérer bénéfique. Cependant, ces matrices sont trop importantes pour pouvoir considérer l'ensemble des corrélations. Une idée

21. A savoir le 4D-PSAS ou la méthode de point selle qui ne nécessitent pas l'inversion de la matrice de covariance d'erreur d'observation. Le préconditionement usuellement employée avec ces méthodes fait cependant apparaître l'inverse de la matrice $\mathbf{R}$.

pourrait être d'utiliser des approximations par blocs diagonaux (d'une taille assez petite pour qu'ils soient aisément inversibles) de ces matrices de covariance.

Une porte d'entrée pour définir ces blocs consiste à se référer au travail effectué en théorie des graphes. Une matrice peut en effet être vue sous la forme d'un graphe où les coefficients hors de la diagonale définissent les *arêtes* entre les différents *sommets*. Le problème posé, afin de définir les meilleurs blocs diagonaux possibles, peut alors être vu comme le problème du meilleur partitionnement possible d'un graphe en k sous–graphes indépendants (voir figure 7.5).

Ce problème est courramment traité dans la littérature de la théorie des graphes. On le rencontre par exemple quand on cherche à résoudre un système de type $Ax = b$ avec A une matrice creuse en parallélisant les tâches. Dans ce cas, on cherche en effet à "découper" le problème en sous–problèmes devant communiquer le moins possible.

7.2.5 Vers des problèmes réalistes

Une étape importante de la suite de ce travail consiste à intégrer l'assimilation directe de séquences d'images dans des modèles réalistes. Cette tâche est actuellement en exploration avec le modèle NEMO.

Cependant, dans des applications océaniques, la gestion des régions côtières peu néanmoins poser certains problèmes. En effet, les transformées mutli–échelles sont en général définies sur des domaines rectangulaires et il peut être complexe de les adapter à des domaines particuliers.

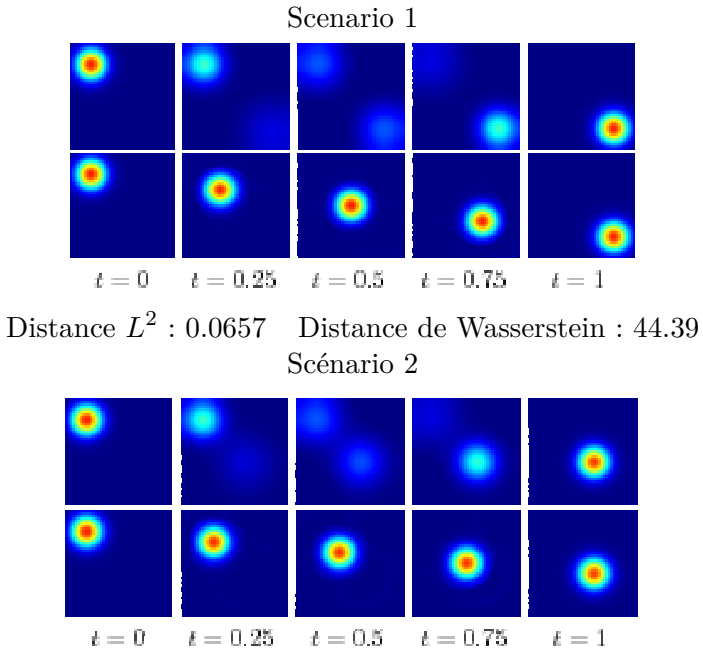


FIGURE 7.6 – Interpolation entre deux densités. Pour les deux scénarios, en haut est présenté l'interpolation dans l'espace des pixels, en bas celle en utilisant la distance de Wasserstein. La distance entre la densité originale (image de gauche) et finale (image de droite) est présentée pour chaque scénario. Source : [Papadakis et al., 2013].

Une manière pour tenir compte de la présence des côtes consiste à les considérer comme des occultations. Il est alors possible de les gérer comme proposé au chapitre 6. Cependant, ces "occultations" n'évoluant pas au cours du temps, le problème peut certainement être résolu de

manière "plus propre". Il n'est en effet pas nécessaire de refaire l'ensemble du travail pour prendre en compte l'existance de côtes à chaque nouvelle image. En passant plus de temps lors de la définition du poids à accorder aux fonctions ayant une partie de leur support dans l'océan et une partie sur la côte, il est certainement possible de mieux tenir compte de la géométrie. Il est aussi possible d'intégrer quelques éléments hors de la diagonale de la matrice de covariance d'erreur d'observation près des côtes afin de mieux modéliser l'interaction entre les structures et ces dernières.

Même si cela demande certainement un travail théorique et de développement conséquent, il est aussi envisageable d'adapter les transformées multi–échelles utilisées à la géométrie de chacune des régions. Cette méthode serait sans doute beaucoup plus propre, mais beaucoup plus complexe à mettre en oeuvre et spécifique à chaque transformée.

7.2.6 Autres définitions de la distance entre deux images

Il est aussi possible d'utiliser des notions de distance entre deux images radicalement différentes. Une idée en cours de développement au sein de l'équipe MOISE consiste à utiliser la distance de Wasserstein. En considérant les images comme des densités, la distance de Wasserstein mesure le transport minimal permettant de déplacer une densité vers une autre. Des avancées théoriques et numériques récentes permettent d'envisager d'utiliser cette métrique pour comparer deux images.

La distance de Wasserstein possède des propriétés pertinentes pour le calcul de la distance entre deux images. Par exemple, contrairement à une métrique du type pixel à pixel, la distance de Wasserstein donne une indication précise sur la distance séparant le support de deux densités dont les supports sont disjoints. Ceci est présenté en figure 7.6.

Par conséquent, en utilisant la distance de Wasserstein, on peut espérer diminuer le nombre de minimum locaux dû à un éloignement conséquent entre l'image observée et l'image synthétique. Cette distance n'est en effet pas sensible au problème rencontré en section 5.8.4 et illustré par le schéma 7.7.

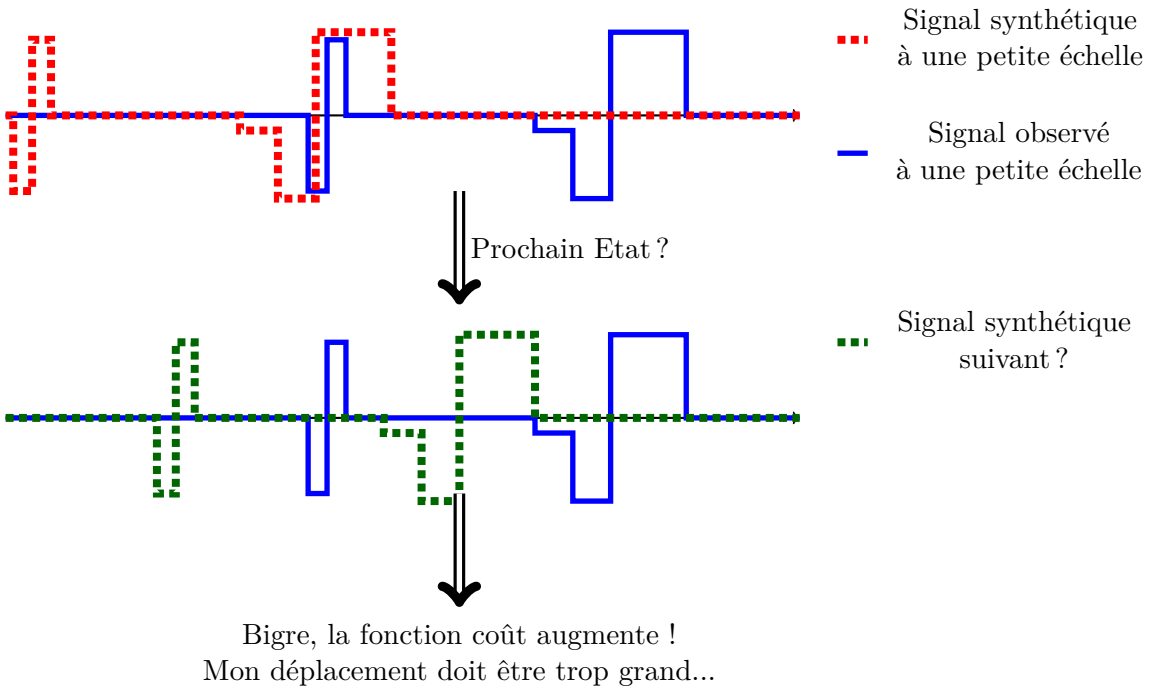


FIGURE 7.7 – Illustration du problème rencontré pour le signal à petite échelle en section 5.8.4 avec une métrique de type pixel à pixel.

Cependant utiliser cette distance n'est pas nécessairement une chose aisée :

- D'un point de vue numérique, le calcul de cette distance est encore très cher (hors cas $1D$) ;
- D'un point de vue théorique il est nécessaire d'accéder à la dérivée de cette distance par rapport aux images pour l'intégrer à une méthode d'assimilation variationnelle.

Des études théoriques et numériques pouvant conduire à l'utilisation de cette distance en assimilation de données sont actuellement menées au sein de l'équipe MOISE dans la thèse de Nelson Feyeux.

A

Pourquoi la hauteur d'eau n'est-elle pas corrigée ?

Tout d'abord rappelons que la hauteur d'eau n'est pas observée directement (on observe uniquement la concentration de traceur passif) et n'entre pas en jeu dans la définition de l'opérateur d'observation (équation (3.36)). Son contrôle est donc, en théorie, assuré par l'intégration rétrograde du modèle (c'est un des avantages théorique du $4D$–Var sur le $3D$–Var et les méthodes de filtrage). Cependant dans les expériences effectuées cette intégration ne suffit pas à contrôler de manière efficace la hauteur d'eau.

Quelle solution peut être employée ?

Une stratégie, afin de transmettre l'information obtenue pour des variables bien observées vers des variables non observées, consiste à utiliser des opérateurs d'équilibre dans un changement de variable.

Concernant l'équilibre entre la hauteur d'eau et le champ de vitesse, cet opérateur est bâti en effectuant l'hypothèse que le déplacement s'effectue uniquement dans la direction tangente aux isobares (lignes de pression constante). Effectuer cette hypothèse signifie qu'un certain équilibre entre les forces normales aux isobares est respecté.

Ces forces comprennent :
- L'accélération centripète (due à la courbure du mouvement),
- La force de pression qui émerge de la différence de pression autour d'un objet,
- La force de Coriolis qui est une force fictive due à la rotation du référentiel dans lequel les équations de mouvement sont étudiées (ici la Terre).

La figure A.1 résume la contribution de ces forces lorsque la circulation s'effectue autour des basses pressions (engendrant des cyclones) et des hautes pressions (engendrant des anticyclones).

L'équation d'équilibre entre ces forces s'écrit :

$$\frac{V^2}{R} + fV = -\frac{1}{\rho}\frac{\partial p}{\partial n} \tag{A.1}$$

où
- ρ est la densité de la masse d'eau
- $\dfrac{\partial p}{\partial n}$ la composante de la force de pression perpendiculaire à la trajectoire,
- fV la force de Coriolis,
- R le rayon de courbure (qui est par convention positif lorsque la circulation s'effectue autour des basses pressions et négatif dans l'autre cas),
- $\dfrac{V^2}{R}$ l'accélération centripète,

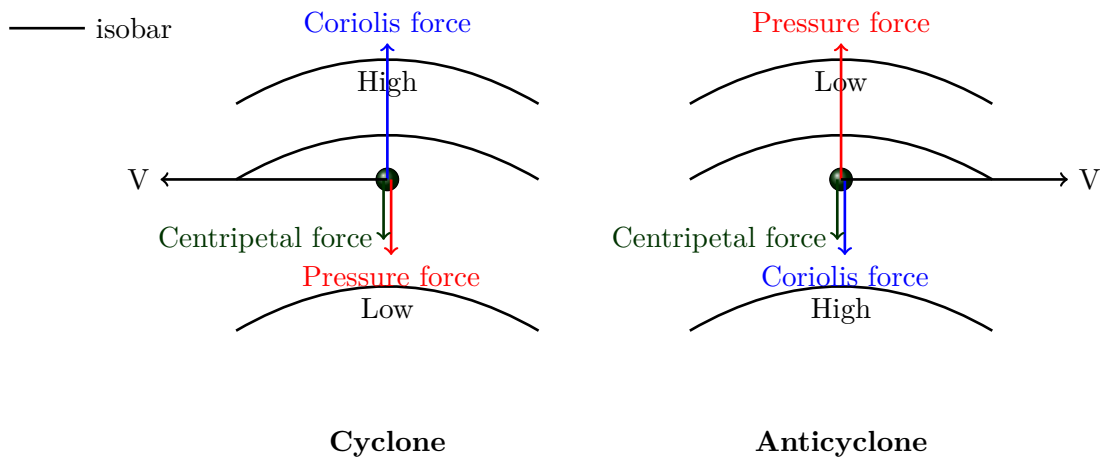


FIGURE A.1 – Forces normales au déplacement. Dans l'hémisphère nord, la force de Coriolis induit une déviation vers la droite des objets. De fait la circulation autour des dépressions est anti-horaire (figure de gauche) et horaire autour des anti-cyclones (figure de droite).

Lorsque la courbure du flot est très faible (c'est à dire que les isobares sont approximativement des lignes droites), l'accélération centripète peut être négligée. Cette approximation, couramment utilisée, se nomme l'équilibre géostrophique :

$$fV_g = -\frac{1}{\rho}\frac{\partial p}{\partial n} \tag{A.2}$$

Cependant dans le cadre des expériences effectuées, l'accélération centripète ne peut pas être négligée (on étudie la dérive d'un vortex). Il n'est de plus pas possible de négliger la force de Coriolis (ce qui aurait pu conduire à utiliser l'équilibre cyclogéostrophique) ou la force de pression. L'ensemble des termes de l'équation (A.1) doivent donc être utilisés. Les solutions de cette équation s'écrivent :

$$V_{gr} = \frac{fR}{2} \pm \sqrt{\frac{f^2R^2}{4} - \rho R\frac{\partial p}{\partial n}}. \tag{A.3}$$

Pourquoi ne pas utiliser cet opérateur d'équilibre ?

Dans la suite nous n'utilisons pas d'opérateur d'équilibre. Utiliser l'opérateur d'équilibre géostrophique conduit à fournir un mauvais opérateur d'équilibre au système et n'améliore dès lors pas l'analyse effectuée sur la surface libre.

L'utilisation de l'opérateur d'équilibre prenant en compte l'ensemble des forces est quant à lui mal défini : la solution de l'équation (A.3) n'est pas définie partout (lorsqu'on étudie certains anticyclones par ex.). De plus, cette solution nécessite d'accéder à une "bonne" estimation du rayon de courbure ce qui n'est pas une chose aisée. Par exemple, dans [Patoux and Brown, 2002] ce dernier est calculé de la manière suivante :

$$\frac{1}{R} \simeq \frac{1}{V}\frac{p_x p_{yt} - p_{xt} p_y}{p_x^2 + p_y^2} + \frac{p_{xx}p_y - 2p_x p_y + p_x p_{yy}}{(p_x + p_y)^{3/2}} \tag{A.4}$$

où p est la pression (qui est justement la variable non équilibrée que l'on cherche à modifier).

Dans cet article, afin d'obtenir à partir d'un champ de vitesse la champ de pression correspondant, un algorithme itératif est utilisé, en remarquant que

$$fV_g = fV_{gr}\left(1 + \frac{V_{gr}}{fR}\right). \tag{A.5}$$

Pour cela la pression est tout d'abord évaluée en résolvant par minimisation ([Brown and Levy, 1986]) le problème (A.2) en supposant l'équilibre géostrophique. Le rayon de courbure est alors obtenu à partir de ce champ de pression via l'équation (A.4). A partir du champ de vent et du rayon de courbure calculé précédemment on en déduit une nouvelle approximation du vent géostrophique *équivalent* par l'équation (A.5). La procédure est ensuite itérée.

Cependant cet équilibre, permettant de corriger le champ de pression, est utilisé uniquement lorsque le champ de vitesse est lisse (car le rayon de courbure peut-être approché de manière "correcte" dans ce cas) ce qui n'est pas nécessairement vérifié durant la phase d'assimilation.

La non-définition d'une solution dans certain cas, en plus de la difficulté liée au calcul du rayon de courbure (lié au gradient de pression) notamment dans les régions où le champ de vitesse n'est pas lisse (entrainant un champ de pression non lisse) entraîne le fait que nous n'utilisons pas cet opérateur d'équilibre lors de la phase d'assimilation.

Remarque A.1

L'opérateur d'équilibre présenté précédemment peut être utilisé en fin d'assimilation afin d'obtenir une estimation de la surface libre analysée cohérente avec le champ de vitesse lorsque ce dernier est assez lisse. Cependant la surface libre ainsi reconstruite dépend dès lors uniquement du champ de vitesses analysé et de la qualité de l'opérateur d'équilibre.

Bilan A.1

Pour ces raisons, dans la suite de ces travaux, nous nous concentrons uniquement sur les diagnostics liés au champ de vitesses et nous ne présentons plus de diagnostics sur la qualité de la surface libre.

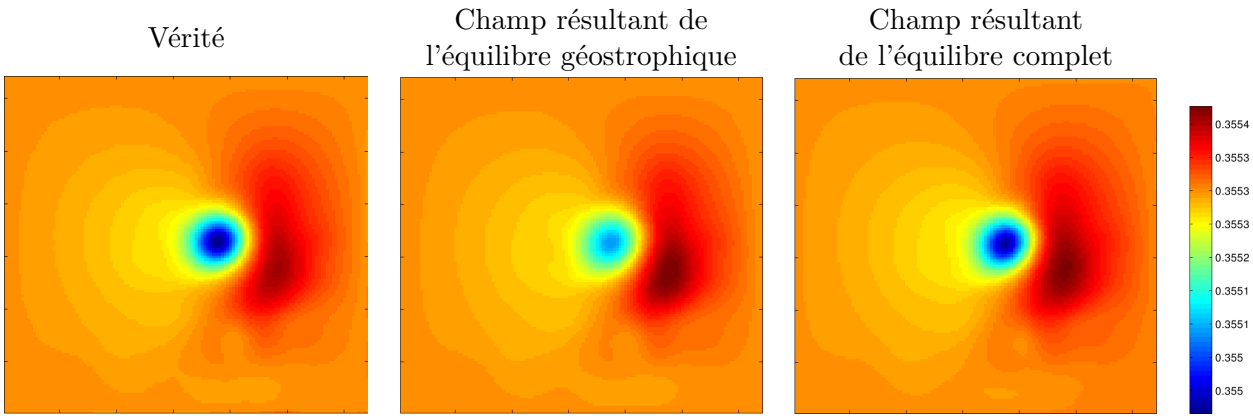


FIGURE A.2 – Exemple d'utilisation d'opérateurs d'équilibre. Les opérateurs d'équilibre sont ici appliqués à partir du vrai champ de vitesse.

Bibliographie

[Alvera-Azcárate et al., 2005] Alvera-Azcárate, A., Barth, A., Rixen, M., and Beckers, J.-M. (2005). Reconstruction of incomplete oceanographic data sets using empirical orthogonal functions : application to the adriatic sea surface temperature. *Ocean Modelling*, 9(4) :325–346.

[Antoniadis et al., 2007] Antoniadis, A. et al. (2007). Wavelet methods in statistics : Some recent developments and their applications. *Statistics Surveys*, 1 :16–55.

[Antoniadis and Fan, 2001] Antoniadis, A. and Fan, J. (2001). Regularization of wavelet approximations. *Journal of the American Statistical Association*, 96(455).

[Auroux and Fehrenbach, 2011] Auroux, D. and Fehrenbach, J. (2011). Identification of velocity fields for geophysical fluids from a sequence of images. *Experiments in Fluids*, 50(2) :313–328.

[Azimifar et al., 2012] Azimifar, Z., Amiri, M., Fieguth, P., and Jernigan, E. (2012). Empirical study of wavelet domain image joint statistics and proposition of an efficient correlation map. *Journal of Mathematical Imaging and Vision*, 43(1) :24–38.

[Ba et al., 2011] Ba, S. O., Corpetti, T., and Fablet, R. (2011). Multi-resolution missing data interpolation in sst image series. In *18th IEEE International Conference on Image Processing*, pages 1517–1520. IEEE.

[Barron et al., 1994] Barron, J. L., Fleet, D. J., and Beauchemin, S. S. (1994). Performance of optical flow techniques. *International journal of computer vision*, 12(1) :43–77.

[Beyou, 2013] Beyou, S. (2013). *Estimation de la vitesse des courants marins à partir de séquences d'images satellitaires*. PhD thesis, Université de Rennes 1.

[Beyou et al., 2013a] Beyou, S., Corpetti, T., Gorthi, S., and Mémin, E. (2013a). Fluid flow estimation with multiscale ensemble filters based on motion measurements under location uncertainty. *Numerical Mathematics : Theory, Methods & Applications*, 6(1).

[Beyou et al., 2013b] Beyou, S., Cuzol, A., Gorthi, S. S., Mémin, E., et al. (2013b). Weighted ensemble transform kalman filter for image assimilation. *Tellus A*, 65.

[Blayo and Nodet, 2012] Blayo, E. and Nodet, M. (2012). Introduction à l'assimilation de données variationnelle. Notes de cours.

[Bobin et al., 2007] Bobin, J., Starck, J.-L., Fadili, J. M., Moudden, Y., and Donoho, D. L. (2007). Morphological component analysis : An adaptive thresholding strategy. *Image Processing, IEEE Transactions on*, 16(11) :2675–2681.

[Bocquet, 2011] Bocquet, M. (2011). Introduction aux principes et méthodes de l'assimilation de données en géophysique. Notes de cours.

[Bonan, 2013] Bonan, B. (2013). *Assimilation de données pour l'initialisation et l'estimation de paramètres d'un modèle d'évolution de calotte polaire*. PhD thesis, Université de Grenoble.

[Bormann et al., 2014] Bormann, N., Hernandez-Carrascal, A., Borde, R., Lutz, H.-J., Otkin, J. A., and Wanzong, S. (2014). Atmospheric motion vectors from model simulations. part i : Methods and characterization as single-level estimates of wind. *Journal of Applied Meteorology & Climatology*, 53(1).

[Bormann et al., 2003] Bormann, N., Saarinen, S., and Kelly, G. (2003). The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. *Mon. Wea. Rev.*, 131(4) :706–718.

[Brown and Levy, 1986] Brown, R. A. and Levy, G. (1986). Ocean surface pressure fields from satellite-sensed winds. *Monthly weather review*, 114(11) :2197–2206.

[Bugeau et al., 2009] Bugeau, A., Bertalmio, M., et al. (2009). Combining texture synthesis and diffusion for image inpainting. *International Conference on Computer Vision Theory and Applications.*, pages 26–33.

[Cai, 1999] Cai, T. T. (1999). Adaptive wavelet estimation : a block thresholding and oracle inequality approach. *Annals of statistics*, pages 898–924.

[Cai and Silverman, 2001] Cai, T. T. and Silverman, B. W. (2001). Incorporating information on neighbouring coefficients into wavelet estimation. *Sankhyā : The Indian Journal of Statistics, Series B*, pages 127–148.

[Candes et al., 2006] Candes, E., Demanet, L., Donoho, D., and Ying, L. (2006). Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5(3) :861–899.

[Candes and Donoho, 2000] Candes, E. J. and Donoho, D. L. (2000). Curvelets : A surprisingly effective nonadaptive representation for objects with edges. Technical report, DTIC Document.

[Cardinali, 2009] Cardinali, C. (2009). Monitoring the observation impact on the short-range forecast. *Quarterly Journal of the Royal Meteorological Society*, 135(638) :239–250.

[Casazza and Leonhard, 2008] Casazza, P. G. and Leonhard, N. (2008). Classes of finite equal norm parseval frames. *Contemporary Mathematics*, 451 :11–32.

[Charney et al., 1950] Charney, J. G., Fjörtoft, R., and Neumann, J. v. (1950). Numerical integration of the barotropic vorticity equation. *Tellus*, 2(4) :237–254.

[Chaux, 2006] Chaux, C. (2006). *Analyse en ondelettes M-bandes en arbre dual; Application à la restauration d'images.* PhD thesis.

[Christensen, 2003] Christensen, O. (2003). *An introduction to frames and Riesz bases.* Springer.

[Corpetti et al., 2009] Corpetti, T., Héas, P., Mémin, E., and Papadakis, N. (2009). Pressure image assimilation for atmospheric motion estimation. *Tellus Series A : Dynamic Meteorology and Oceanography*, 61(1) :160–178.

[Corpetti and Mémin, 2012] Corpetti, T. and Mémin, E. (2012). Stochastic uncertainty models for the luminance consistency assumption. *Image Processing, IEEE Transactions on*, 21(2) :481–493.

[Courtier, 1997] Courtier, P. (1997). Dual formulation of four-dimensional variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, 123(544) :2449–2461.

[Courtier et al., 1998] Courtier, P., Andersson, E., Heckley, W., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., Fisher, M., and Pailleux, J. (1998). The ecmwf implementation of three-dimensional variational assimilation (3d-var). i : Formulation. *Quarterly Journal of the Royal Meteorological Society*, 124(550) :1783–1807.

[Dando et al., 2007] Dando, M., Thorpe, A., and Eyre, J. (2007). The optimal density of atmospheric sounder observations in the met office nwp system. *Quarterly Journal of the Royal Meteorological Society*, 133(629) :1933–1943.

[Daubechies, 1988] Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on pure and applied mathematics*, 41(7) :909–996.

[Daubechies et al., 1992] Daubechies, I. et al. (1992). *Ten lectures on wavelets*, volume 61. SIAM.

[Deckmyn and Berre, 2005] Deckmyn, A. and Berre, L. (2005). A wavelet approach to representing background error covariances in a limited-area model. *Monthly weather review*, 133(5).

[Desroziers et al., 2009] Desroziers, G., Berre, L., Chabot, V., and Chapnik, B. (2009). A posteriori diagnostics in an ensemble of perturbed analyses. *Monthly Weather Review*, 137 :3420–3436.

[Do, 2001] Do, M. N. (2001). *Directional multiresolution image representations*, volume 2500. Citeseer.

[Do and Vetterli, 2002] Do, M. N. and Vetterli, M. (2002). Contourlets : a directional multiresolution image representation. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I–357. IEEE.

[Do and Vetterli, 2003] Do, M. N. and Vetterli, M. (2003). The finite ridgelet transform for image representation. *Image Processing, IEEE Transactions on*, 12(1) :16–28.

[Donoho and Johnstone, 1994] Donoho, D. L. and Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3) :425–455.

[D'Ovidio et al., 2009] D'Ovidio, F., Taillandier, V., Mortier, L., and Taupier-Letage, I. (2009). Lagrangian validation of the mediterranean mean dynamic topography by extraction of tracer frontal structures. *Mercator Ocean Quarterly Newsletter*, 32 :24–32.

[Duffin and Schaeffer, 1952] Duffin, R. J. and Schaeffer, A. C. (1952). A class of nonharmonic fourier series. *Transactions of the American Mathematical Society*, pages 341–366.

[Durand et al., 2010] Durand, S., Fadili, J., and Nikolova, M. (2010). Multiplicative noise removal using l1 fidelity on frame coefficients. *Journal of Mathematical Imaging and Vision*, 36(3) :201–226.

[Elad et al., 2005] Elad, M., Starck, J.-L., Querre, P., and Donoho, D. L. (2005). Simultaneous cartoon and texture image inpainting using morphological component analysis (mca). *Applied and Computational Harmonic Analysis*, 19(3) :340–358.

[Fadili et al., 2010a] Fadili, J., Starck, J.-L., Elad, M., Donoho, D., et al. (2010a). Mcalab : Reproducible research in signal and image decomposition and inpainting. *IEEE Computing in Science and Engineering*, 12(1) :44–63.

[Fadili et al., 2010b] Fadili, M.-J., Starck, J.-L., Bobin, J., and Moudden, Y. (2010b). Image decomposition and separation using sparse representations : an overview. *Proceedings of the IEEE*, 98(6) :983–994.

[Fadili et al., 2009] Fadili, M.-J., Starck, J.-L., and Murtagh, F. (2009). Inpainting and zooming using sparse representations. *The Computer Journal*, 52(1) :64–79.

[Fisher, 2014] Fisher, M. (2014). Accounting for correlated observation error in variational data assimilation. In *Workshop on correlated observation errors in data assimilation*, Reading, United Kingdom. ESA - University of Reading.

[Flór and Eames, 2002] Flór, J.-B. and Eames, I. (2002). Dynamics of monopolar vortices on a topographic beta-plane. *Journal of Fluid Mechanics*, 456(1) :353–376.

[Forsythe et al., 2008] Forsythe, M., Cotton, J., and Saunders, R. (2008). Improving amv impact in nwp. *Terra*.

[Frigo and Johnson, 2005] Frigo, M. and Johnson, S. G. (2005). The design and implementation of fftw3. *Proceedings of the IEEE*, 93(2) :216–231.

[Gaultier, 2013] Gaultier, L. (2013). *Couplage des observations spatiales dynamiques et biologiques pour la restitution des circulations océaniques : une approche conjointe par assimilation de données altimétriques et de traceurs*. PhD thesis, Université de grenoble.

[Gaultier et al., 2012] Gaultier, L., Verron, J., Brankart, J.-M., Titaud, O., and Brasseur, P. (2012). On the inversion of submesoscale tracer fields to estimate the surface ocean circulation. *Journal of Marine Systems*.

[Gilbert and Lemaréchal, 1989] Gilbert, J. C. and Lemaréchal, C. (1989). Some numerical experiments with variable-storage quasi-newton algorithms. *Mathematical programming*, 45(1-3) :407–435.

[Hascoet and Pascual, 2013] Hascoet, L. and Pascual, V. (2013). The tapenade automatic differentiation tool : principles, model, and specification. *ACM Transactions on Mathematical Software (TOMS)*, 39(3) :20.

[Herlin et al., 2012] Herlin, I., Huot, E., et al. (2012). Monitoring surface currents from uncertain image observations. In *XX International Conference on Problems of Decision Making under Uncertainties (PDMU)*.

[Holmlund, 2000] Holmlund, K. (2000). The atmospheric motion vector retrieval scheme for meteosat second generation. In *Proc. Fifth Int. Winds Workshop*, pages 201–208.

[Horn and Schunck, 1981] Horn, B. and Schunck, B. (1981). Determining optical flow. *Artificial Intelligence*, 17 :185–203.

[Huot et al., 2010] Huot, E., Herlin, I., Mercier, N., and Plotnikov, E. (2010). Estimating apparent motion on satellite acquisitions with a physical dynamic model. In *International Conference on Pattern Recognition (ICPR'10)*, volume 1, pages 41–44.

[Jacques et al., 2011] Jacques, L., Duval, L., Chaux, C., and Peyré, G. (2011). A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity. *Signal Processing*, 91(12) :2699–2730.

[Johnstone and Silverman, 1997] Johnstone, I. M. and Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 59(2) :319–351.

[Kingsbury, 1999] Kingsbury, N. (1999). Image processing with complex wavelets. *Philosophical Transactions of the Royal Society of London. Series A : Mathematical, Physical and Engineering Sciences*, 357(1760) :2543–2560.

[Korotaev et al., 2008] Korotaev, G., Huot, E., Le Dimet, F.-X., Herlin, I., Stanichny, S., Solovyev, D., and Wu, L. (2008). Retrieving ocean surface current by 4-d variational assimilation of sea surface temperature images. *Remote Sensing of Environment*, 112(4) :1464 – 1475.

[Le Cadet, 2004] Le Cadet, O. (2004). *Méthodes d'ondelettes pour la segmentation d'images : Applications à l'imagerie médicale et au tatouage d'images*. PhD thesis, PhD Thesis at the Polytechnic Institute of Grenoble.

[LeDimet, 1982] LeDimet, F.-X. (1982). Une application des methodes de côntrole, optimal à l'analyse variationnelle. *Rapport scientifique LAMP, Universite Blaise Pascal*.

[LeDimet and Talagrand, 1986] LeDimet, F.-X. and Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations : theoretical aspects. *Tellus A*, 38(2) :97–110.

[Lepoittevin et al., 2013] Lepoittevin, Y., Herlin, I., Béréziat, D., et al. (2013). Object's tracking by advection of a distance map. In *ICIP-IEEE International Conference on Image Processing*, pages 3612–3616.

[Liu and Rabier, 2002] Liu, Z.-Q. and Rabier, F. (2002). The interaction between model resolution, observation resolution and observation density in data assimilation : A one-dimensional study. *Quarterly Journal of the Royal Meteorological Society*, 128(582) :1367–1386.

[Lucas and Kanade, 1981] Lucas, B. and Kanade, T. (1981). An iterative image registration technique with an application to stereovision. In *Int. Joint Conf. on Artificial Intel. (IJCAI'81)*, pages 674–679.

[Ma et al., 2008] Ma, J., Titaud, O., Vidard, A., Le Dimet, F.-X., et al. (2008). Spatio-temporal structure extraction and denoising of geophysical fluid image sequences using 3d curvelet transforms. Technical report, INRIA.

[Makris and Papadakis, 2012] Makris, A. and Papadakis, N. (2012). Data assimilation with state alignment using high-level image structures detection. In *Computer Vision in Remote Sensing (CVRS), 2012 International Conference on*, pages 78–83. IEEE.

[Mallat, 1989] Mallat, S. (1989). A theory for multiresolution signal decomposition : The wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7) :674–693.

[Mallat, 2008] Mallat, S. (2008). *A wavelet tour of signal processing : the sparse way.* Academic press.

[Michel, 2011] Michel, Y. (2011). Displacing potential vorticity structures by the assimilation of pseudo-observations. *Monthly Weather Review*, 139(2).

[Mirouze, 2010] Mirouze, I. (2010). *Régularisation de problèmes inverses à l'aide de l'équation de diffusion, avec application à l'assimilation variationnelle de données océaniques.* PhD thesis, Université Paul Sabatier-Toulouse III.

[Morlet and Grossman, 1984] Morlet, J. and Grossman, A. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis*, 15 :723–736.

[Neveu, 2011] Neveu, É. (2011). *Application des méthodes multigrilles à l'assimilation variationnelle de données en géophysique.* PhD thesis, Université de Grenoble.

[Nodet, 2013] Nodet, M. (2013). *Problèmes inverses pour l'environnement : outils, méthodes et applications.* PhD thesis, Université de Grenoble.

[Pannekoucke, 2009] Pannekoucke, O. (2009). Heterogeneous correlation modeling based on the wavelet diagonal assumption and on the diffusion operator. *Monthly Weather Review*, 137(9) :2995–3012.

[Pannekoucke et al., 2007] Pannekoucke, O., Berre, L., and Desroziers, G. (2007). Filtering properties of wavelets for local background-error correlations. *Quarterly Journal of the Royal Meteorological Society*, 133(623) :363–379.

[Pannekoucke et al., 2013] Pannekoucke, O., Raynaud, L., and Farge, M. (2013). A wavelet-based filtering of ensemble background-error variances. *Quarterly Journal of the Royal Meteorological Society.*

[Papadakis, 2007] Papadakis, N. (2007). *Assimilation de données image : application au suivi de courbes et de champs de vecteurs.* PhD thesis, Université de Rennes I.

[Papadakis et al., 2013] Papadakis, N., Chabot, V., Makris, A., Nodet, M., Vidard, A., et al. (2013). Assimilation d'images et de structures. In *24ème colloque GRETSI*, Brest, France.

[Papadakis and Mémin, 2007] Papadakis, N. and Mémin, E. (2007). A variational framework for spatio-temporal smoothing of fluid motions. In *Scale Space and Variational Methods, SSVM'07*, pages 603–615.

[Papadakis and Mémin, 2008] Papadakis, N. and Mémin, E. (2008). Variational assimilation of fluid motion from image sequence. *SIAM Journal on Imaging Sciences*, 1 :343–363.

[Patoux and Brown, 2002] Patoux, J. and Brown, R. A. (2002). A gradient wind correction for surface pressure fields retrieved from scatterometer winds. *Journal of Applied Meteorology*, 41(2).

[Schmetz et al., 1993] Schmetz, J., Holmlund, K., Hoffman, J., Strauss, B., Mason, B., Gaertner, V., Koch, A., and van De, B. (1993). Operational cloud-motion winds from Meteosat infrared images. *Journal of applied meteorology*, 32(7) :1206–1225.

[Souopgui, 2010] Souopgui, I. (2010). *Assimilation d'images pour les fluides géophysiques.* PhD thesis, Université de Grenoble.

[Starck et al., 2004] Starck, J.-L., Elad, M., and Donoho, D. (2004). Redundant multiscale transforms and their application for morphological component separation. *Advances in Imaging and Electron Physics*, 132(82) :287–348.

[Stewart et al., 2013a] Stewart, L., Dance, S., Nichols, N., Eyre, J., and Cameron, J. (2013a). Estimating interchannel observation error correlations for iasi radiance data in the met office system. *Quarterly Journal of the Royal Meteorological Society.*

[Stewart, 2010] Stewart, L. M. (2010). *Correlated observation errors in data assimilation.* PhD thesis, University of Reading.

[Stewart et al., 2013b] Stewart, L. M., Dance, S., and Nichols, N. K. (2013b). Data assimilation with correlated observation errors : experiments with a 1-d shallow water model. *Tellus A*, 65.

[Thomas et al., 2010] Thomas, C., Corpetti, T., and Memin, E. (2010). Data assimilation for convective-cell tracking on meteorological image sequences. *Geoscience and Remote Sensing, IEEE Transactions on*, 48(8) :3162–3177.

[Titaud et al., 2011] Titaud, O., Brankart, J.-M., and Verron, J. (2011). On the use of finite-time lyapunov exponents and vectors for direct assimilation of tracer images into ocean models. *Tellus Series A : Dynamic Meteorology and Oceanography*, 63(5) :1038–1051.

[Titaud et al., 2010] Titaud, O., Vidard, A., Souopgui, I., and Le Dimet, F.-X. (2010). Assimilation of image sequences in numerical models. *Tellus Series A : Dynamic Meteorology and Oceanography*, 62(1) :30–47.

[Truchetet, 1998] Truchetet, F. (1998). *Ondelettes pour le signal numérique.* Hermes.

[Vannucci and Corradi, 1999] Vannucci, M. and Corradi, F. (1999). Covariance structure of wavelet coefficients : theory and models in a bayesian perspective. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 61(4) :971–986.

[Vidakovic and Mueller, 1994] Vidakovic, B. and Mueller, P. (1994). Wavelets for kids. *Instituto de Estadística, Universidad de Duke.*

[Vidard, 2001] Vidard, A. (2001). *Vers une prise en compte des erreurs modèles en assimilation de données 4D-variationnelle : application à un modèle réaliste d'océan.* PhD thesis, Grenoble 1.

[Vidard et al., 2008] Vidard, A., Le Dimet, F.-X., Souopgui, I., and Titaud, O. (2008). Assimilation of Images in Numerical Models in Geophysics. In Herskovits, J., editor, *EngOpt 2008 - International Conference on Engineering Optimization*, Rio de Janeiro, Brésil. COPPE/UFRJ. ISBN 978-85-7650-152-7.

[Weaver and Courtier, 2001] Weaver, A. and Courtier, P. (2001). Correlation modelling on the sphere using a generalized diffusion equation. *Quarterly Journal of the Royal Meteorological Society*, 127 :1815–1846.

[Weston, 2011] Weston, P. (2011). Progress towards the implementation of correlated observation errors in 4d-var. *Met Office Forecasting Research Technical Report*, 560.

[Weston et al., 2014] Weston, P., Bell, W., and Eyre, J. (2014). Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society.*

[Woiselle, 2010] Woiselle, A. (2010). *Parcimonie et ses applications à des données de flot optique.* PhD thesis, Paris 7.

[Ying et al., 2005] Ying, L., Demanet, L., and Candes, E. (2005). 3d discrete curvelet transform. In *Optics & Photonics 2005*, pages 591413–591413. International Society for Optics and Photonics.

Résumé

Les dernières décennies ont vu croître en quantité et en qualité les données satellites. Au fil des ans, ces observations ont pris de plus en plus d'importance en prévision numérique du temps. Ces données sont aujourd'hui cruciales afin de déterminer de manière optimale l'état du système étudié, et ce, notamment car elles fournissent des informations denses et de qualité dans des zones peu observées par les moyens conventionnels. Cependant, le potentiel de ces séquences d'images est encore largement sous–exploité en assimilation de données : ces dernières sont sévèrement sous–échantillonnées, et ce, en partie afin de ne pas avoir à tenir compte des corrélations d'erreurs d'observation.

Dans ce manuscrit, nous abordons le problème d'extraction, à partir de séquences d'images satellites, d'information sur la dynamique du système durant le processus d'assimilation variationnelle de données. Cette étude est menée dans un cadre idéalisé afin de déterminer l'impact d'un bruit d'observations et/ou d'occultations sur l'analyse effectuée.

Lorsque le bruit est corrélé en espace, tenir compte des corrélations en analysant les images au niveau du pixel n'est pas chose aisée : il est nécessaire d'inverser la matrice de covariance d'erreur d'observation (qui se révèle être une matrice de grande taille) ou de faire des approximations aisément inversibles de cette dernière. En changeant d'espace d'analyse, la prise en compte d'une partie des corrélations peut être rendue plus aisée. Dans ces travaux, nous proposons d'effectuer cette analyse dans des bases d'ondelettes ou des trames de curvelettes. En effet, un bruit corrélé en espace n'impacte pas de la même manière les différents éléments composants ces familles. En travaillant dans ces espaces, il est alors plus aisé de tenir compte d'une partie des corrélations présentes au sein du champ d'erreurs. La pertinence de l'approche proposée est présentée sur différents cas tests.

Lorsque les données sont partiellement occultées, il est cependant nécessaire de savoir comment adapter la représentation des corrélations. Ceci n'est pas chose aisée : travailler avec un espace d'observation changeant au cours du temps rend difficile l'utilisation d'approximations aisément inversibles de la matrice de covariance d'erreur d'observation. Dans ces travaux ,une méthode permettant d'adapter, à moindre coût, la représentation des corrélations (dans des bases d'ondelettes) aux données présentes dans chaque image est proposée. L'intérêt de cette approche est présenté dans un cas idéalisé.

Recent decades have seen an increase in quantity and quality of satellite observations . Over the years , those observations has become increasingly important in numerical weather forecasting. Nowadays, these datas are crucial in order to determine optimally the state of the studied system. In particular, satellites can provide dense observations in areas poorly observed by conventionnal networks. However, the potential of such observations is clearly under–used in data assimilation : in order to avoid the management of observation errors, thinning methods are employed in association to variance inflation.

In this thesis, we adress the problem of extracting information on the system dynamic from satellites images data during the variationnal assimilation process. This study is carried out in an academic context in order to quantify the influence of observation noise and of clouds on the performed analysis.

When the noise is spatially correlated, it is hard to take into account such correlations by working in the pixel space. Indeed, it is necessary to invert the observation error covariance matrix (which turns out to be very huge) or make an approximation easily invertible of such a matrix. Analysing the information in an other space can make the job easier. In this manuscript, we propose to perform the analysis step in a wavelet basis or a curvelet frame. Indeed, in those structured spaces, a correlated noise does not affect in the same way the differents structures. It is then easier to take into account part of errors correlations : a suitable approximation of the covariance matrix is made by considering only how each kind of element is affected by a correlated noise. The benefit of this approach is demonstrated on different academic tests cases.

However, when some data are missing one has to address the problem of adapting the way correlations are taken into account. This work is not an easy one : working in a different observation space for each image makes the use of easily invertible approximate covariance matrix very tricky. In this work a way to adapt the diagonal hypothesis of the covariance matrix in a wavelet basis, in order to take into account that images are partially hidden, is proposed. The interest of such an approach is presented in an idealised case.