



HAL
open science

Assimilation d'images pour les fluides géophysiques

Innocent Souopgui

► **To cite this version:**

Innocent Souopgui. Assimilation d'images pour les fluides géophysiques. Mathématiques [math]. Université Joseph-Fourier - Grenoble I, 2010. Français. NNT: . tel-00539626

HAL Id: tel-00539626

<https://theses.hal.science/tel-00539626>

Submitted on 24 Nov 2010

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

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



UNIVERSITÉ DE GRENOBLE



UNIVERSITE DE YAOUNDE I

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : Mathématiques Appliquées

Préparée dans le cadre d'une cotutelle entre

L'UNIVERSITÉ DE GRENOBLE ET L'UNIVERSITÉ DE YAOUNDÉ I

Arrêtés ministériels : 6 janvier 2005 - 7 août 2006

Présentée et soutenue publiquement par

SOUOPGUI INNOCENT

Le 25 Octobre 2010

ASSIMILATION D'IMAGES POUR LES FLUIDES GEOPHYSIQUES

Thèse dirigée par François-Xavier Le Dimet et codirigée par Emmanuel Kamgnia et Arthur Vidard

JURY

M. Nguetseng Gabriel	Professeur, Université de Yaoundé I,	Président
M. Mémin Etienne	Directeur de Recherche, INRIA Rennes,	Rapporteur
M. Philippe Bernard	Directeur de Recherche, INRIA Rennes,	Rapporteur
M. Blayo Eric	Professeur, Université de Grenoble,	Examineur
M. Le Dimet François-Xavier	Professeur, Université de Grenoble,	Directeur de thèse
M. Kamgnia Emmanuel	Chargé de cours, Université de Yaoundé I,	Codirecteur
M. Tchuenté Maurice	Professeur, ENS Lyon / UYI,	Examineur
M. Vidard Arthur	Chargé de Recherche, INRIA Grenoble,	Codirecteur

Thèse préparée au sein du **Laboratoire Jean Kuntzmann**
et de l'Ecole Doctorale **Mathématiques, Sciences et Technologie de l'Information, Informatique**

*À mon regretté père Fotso Joseph
(il aurait certainement été content de ce travail)*

Remerciements

Je voudrais exprimer ma gratitude à tous les membres du jury qui ont évalué ce travail de thèse, spécialement Étienne Mémin et Bernard Philippe pour leur lourde tâche de rapporteur.

Je remercie particulièrement mes directeurs de thèse : François-Xavier Le Dimet qui a défini le sujet et suivi la thèse d'un bout à l'autre ; Arthur Vidard qui a suivi pas à pas tout mon travail (de fois, je l'ai interrompu plusieurs fois le même jour) et Emmanuel Kamgnia qui a été le premier à m'initier à la recherche.

Un grand merci à Bernard qui a participé à mon initiation à la recherche à travers mon stage de DEA et qui m'a mis en contact avec François-Xavier. Son concours a été d'une importance capitale pour ma formation à travers le programme SARIMA.

Un grand merci à Isabelle sans qui le financement d'une bonne partie de cette thèse n'aurait pas abouti. Qu'elle retrouve dans ce travail le résultat de ma participation au projet ADDISAAF.

Je remercie l'équipe MOISE et tous ses membres pour le bon accueil, la bonne humeur, l'alliance parfaite entre la détente et la recherche. Une mention spéciale aux assistantes Imma et Anne. Grâce à leur dévouement et leur efficacité, j'ai pu me consacrer uniquement au travail de recherche et elles ont fait le reste. Les encouragements et leur bonne humeur y a été pour beaucoup. Peut-on demander plus ?

Je remercie le LJK et son équipe administrative pour tous les bons services rendus, spécialement Claudine pour toute sa disponibilité.

Il sont nombreux, ceux que je voudrais remercier

Olivier pour tous les conseils et la relecture du manuscrit. Vincent pour la relecture.

Tous les thésards du LJK pour ces moments inoubliables (resto, fooball, bar, etc.).

Mes collègues de bureau : Florian, Claire et Claire que j'ai trouvé et qui sont partis, Roland, Bénédicte et Thomas. Ils m'ont supporté tout ce temps.

Madame MAS pour son accueil chaleureux, ses conseils et encouragements.

Tous mes amis de Grenoble : Eric, Linda, Martial, Priscille, Rodrigue sans être exhaustif (la liste est longue et l'espace réservé ne pourrait la contenir entièrement). Ils ont été présents en toute circonstance. S'il fallait les créer à nouveau, je demanderai des copies identiques. Que dire de ceux installés dans d'autres villes de la France : Agnes la limogéode, Roméo et Simplicie les parisiens, Désiré le rennais, Larissa la toulousaine, Fritz le Parisien devenu canadien, Dieudonné le brestois, etc.

Mes amis de Yaoundé avec qui j'ai commencé la thèse : Yanik (déjà docta), Donatien (bientôt docta) et Hamza (futur docta) sans oublier ceux du département d'informatique de l'Université de Yaoundé I et tous les autres. Ils sont nombreux, ceux avec qui j'ai partagé un bout de chemin : des moments inoubliables à renouveler (Germaine, Christiane, Bernard, Paulin, etc. en savent quelque chose). Ils ont dû me supporter même en version peu agréable ; c'est aussi ça le travail des amis.

Toute ma famille : les aînés pour le soutien, les conseils et les encouragements ; les cadets, qu'ils trouvent dans ce travail une source de motivation. Je ne remercierai jamais assez ma mère, et les mots pour décrire le concours de André Pascal me manquent. Guy saura décrire mieux que quiconque ce qu'a été la préparation de cette thèse. François Honoré sait ce que veut dire remonter le moral. Il faudrait une autre thèse pour décrire le concours de tous.

Résumé

La compréhension et la prévision de l'évolution des fluides géophysiques sont d'une importance capitale et constituent un domaine de recherche scientifique aux enjeux conséquents. Une bonne prévision est basée sur la prise en compte de toutes les informations disponibles sur le système considéré. Ces informations incluent les modèles, les observations et les connaissances *a priori*. L'assimilation de données permet de les combiner de façon optimale pour déterminer les entrées du modèle. Les dernières décennies ont vu croître en densité et en qualité la couverture satellitaire produisant, entre autres, des séquences d'images montrant l'évolution dynamique de certains phénomènes géophysiques tels que les dépressions et les fronts. Ces séquences d'images sont jusqu'à présent sous-utilisées en assimilation de données. Cette thèse propose une extension de l'assimilation variationnelle de données aux observations de type séquence d'images. Après avoir présenté les images, leur utilisation actuelle et ses limites, nous introduisons les notions de niveau d'interprétation, d'espaces et d'opérateur image. Ces notions sont utilisées pour formuler l'assimilation directe de séquences d'images. Nous proposons également une nouvelle approche de régularisation par diffusion généralisée pour les problèmes inverses. Les résultats préliminaires en traitement d'images et en assimilation directe de séquences d'images montrent une méthode prometteuse qui résout la plupart des problèmes rencontrés avec les approches classiques de régularisation.

Mots clés : assimilation de données, assimilation d'images, fluide géophysique, régularisation.

Abstract

Understanding and forecasting the evolution of geophysical fluids is a major scientific and societal challenge. A good forecast must take into account all available information on the studied system. These informations include models, observations and *a priori* knowledge. Data assimilation techniques combine all these informations in a consistent way to produce model inputs. During the last decades, many satellites were launched to increase the knowledge of earth. They produce, among others, image sequences showing the dynamical evolution of geophysical processes such as depressions and fronts. These images sequences are currently under-utilized in data assimilation.

This thesis presents a consistent approach for taking into account image sequences in variational data assimilation. After a presentation of images, their current use and its limitation, we introduce the concepts of interpretation level, image space and image operator used for direct image sequences assimilation. We also propose a new approach of regularization based on generalized diffusion for ill-posed inverse problems. Preliminary results on image processing and image sequences assimilation show a promising approach that solve most of the problems encountered with classical approaches of regularization.

Keywords : data assimilation, images assimilation, geophysical fluid, regularization.

Table des matières

1	Introduction générale	7
1.1	Introduction	7
1.2	Vocabulaire	11
I	Assimilation de données : concepts fondamentaux	15
2	Assimilation variationnelle de données	17
2.1	Introduction à la théorie de l'estimation	17
2.1.1	Estimation probabiliste	17
2.1.2	Estimation par moindre carré	20
2.2	De l'estimation probabiliste à l'assimilation variationnelle	21
2.2.1	Généralisation de l'approche du maximum de vraisemblance	21
2.2.2	Assimilation variationnelle 4D	22
2.3	Assimilation variationnelle sous contraintes fortes	23
2.3.1	Le modèle direct	23
2.3.2	La fonction coût	24
2.4	Méthodes de minimisation de la fonction coût	25
2.4.1	Théorie de l'optimisation	25
2.4.2	Rappel sur les notions d'optimisation	25
2.4.3	Principe général des méthodes de descente	29
2.4.4	Préconditionnement pour les problèmes inverses	30
2.5	Calcul du gradient : le modèle adjoint	34
2.5.1	Adjoint continu	35
2.5.2	Adjoint discret	36
2.5.3	Le système d'optimalité	38
2.5.4	Utilité de l'adjoint	38
2.6	Pratique des méthodes basées sur l'adjoint	39
2.6.1	Principe de différentiation algorithmique	39
2.6.2	Outils de différentiation automatique	40
3	Modèles et données : cas de l'atmosphère	43
3.1	Variables du système	43
3.2	Observations conventionnelles	43
3.2.1	Mesures <i>in situ</i>	44
3.2.2	Télémesures	45

3.2.3	Ingrédients pour l'assimilation des télémessures	47
3.3	Observation de type image	49
3.3.1	Images brutes	49
3.3.2	Présentation des images satellitaires	50
3.3.3	Principe d'acquisition des images satellite	50
3.3.4	Couverture géographique et résolution des images par satellite	52
3.3.5	Images et vision par ordinateur	53
3.3.6	Images en assimilation de données : état de l'art	56
3.3.7	Ingrédients pour l'assimilation directe d'images	59
3.4	Conclusion	59
II	Vers une prise en compte des observations de type image	61
4	Assimilation directe de séquences d'images	63
4.1	Observations de type image et leurs caractéristiques	63
4.1.1	Interprétation des images	63
4.1.2	Représentation mathématique des images	65
4.1.3	Séquences d'images	69
4.1.4	Utilisation classique des images en assimilation de données	69
4.2	Cadre général de l'assimilation d'images	71
4.2.1	Contexte	71
4.2.2	Prise en compte des images dans la fonction coût	72
4.2.3	Opérateurs d'observation pour les images	75
4.2.4	Régularisation	77
4.3	Application à un modèle de Saint-Venant bidimensionnel	78
4.3.1	Cadre général de l'application	78
4.3.2	Processus d'assimilation d'images	79
4.4	Résultats des expérimentations numériques	83
4.4.1	Expériences jumelles	83
4.4.2	Expériences avec des images réelles	92
4.5	Conclusion	106
5	Régularisation pour l'assimilation d'images	107
5.1	Introduction	107
5.1.1	Contexte	107
5.1.2	Rappels sur les problèmes inverses et leur résolution	108
5.2	Etat de l'art : régularisation par pénalisation	112
5.2.1	Régularisation des champs de vecteurs	112
5.2.2	Techniques d'amélioration de l'impact de la régularisation	116
5.2.3	Limites de la régularisation par pénalisation	116
5.3	Etude expérimentale de la régularisation par pénalisation	118
5.3.1	Fonctions de diagnostic	118
5.3.2	Erreur angulaire	119
5.3.3	Expérience sans régularisation	120
5.3.4	Evolution des diagnostics en fonction du paramètre de pondération	121
5.3.5	Choix du paramètre de pondération	124

5.3.6	Evolution des fonctions de diagnostic au cours de la minimisation	125
5.3.7	Comparaison des diagnostics pour différents termes de régularisation . .	129
5.3.8	Interprétation mathématique et problèmes liés à l'usage de la régulari- sation par pénalisation	132
5.3.9	Conclusion partielle	133
5.4	Régularisation par diffusion généralisée	133
5.4.1	Définition de la nouvelle approche	133
5.4.2	Des fonctions de régularisation aux opérateurs de projection	134
5.4.3	Opérateurs de projection pour quelques fonctions de régularisation . . .	136
5.4.4	Mise en œuvre	140
5.4.5	De la diffusion généralisée au préconditionnement	143
5.4.6	Extension : vers la restauration du signal	143
5.5	Application à l'estimation du mouvement	143
5.5.1	De la régularisation par pénalisation du gradient à l'opérateur de pro- jection	143
5.5.2	Résultats expérimentaux	144
5.5.3	Etude comparative avec l'approche classique	146
5.5.4	Sensibilité à la fonction de confiance	152
5.5.5	conclusion partielle	156
5.6	Conclusion	158
6	Conclusion générale et perspectives	159
6.1	Conclusion	159
6.2	Perspectives	161
6.2.1	Opérateurs images	161
6.2.2	Vers des problèmes réalistes	161
6.2.3	Distance entre images	162
	Annexes	164
A	Approches séquentielles de l'assimilation de données	165
A.1	Interpolation optimale	165
A.2	Le filtre de Kalman	167
A.3	Problèmes non linéaires : le filtre de Kalman étendu	168
B	Méthodes de descente pour la minimisation	169
B.1	Principe général des méthodes de descente	169
B.2	Méthodes du gradient	170
B.2.1	Méthode de descente de la plus grande pente	170
B.2.2	Méthode du gradient conjugué	171
B.3	Méthodes de Newton	171
B.3.1	Principe des méthodes de Newton	171
B.3.2	Algorithmes de type quasi-Newton	172

C	Modèles d'atmosphère et observations conventionnelles	175
C.1	Variables du système	175
C.1.1	Le vent	175
C.1.2	La température	176
C.1.3	L'humidité	176
C.1.4	La pression atmosphérique	176
C.1.5	Les précipitations	177
C.1.6	Les nuages	177
C.2	Observations conventionnelles	177
C.2.1	Réseau d'observations <i>in situ</i>	177
C.2.2	Télémesures	179
D	Estimation du mouvement	183
D.1	Description de l'estimation du mouvement	183
D.2	Modèle expérimental d'estimation du mouvement	184
D.3	Analyse d'incertitudes	185
D.4	Données expérimentales pour l'estimation du mouvement	185
E	Etude de la régularisation par pénalisation	187
E.1	Expérience sans régularisation	187
E.2	Evolution en fonction du paramètre de pondération	188
E.3	Evolution au cours de la minimisation	191
	Tables	203
	Références bibliographiques	203
	Liste des tableaux	214
	Liste des figures	215
	Index	219

Chapitre 1

Introduction générale

1.1 Introduction

Problématique

Les **fluides géophysiques** tels que l'atmosphère et les océans sont des systèmes qui jouent un rôle capital pour la vie. Cependant, leur évolution peut être à l'origine de phénomènes dont les conséquences sont désastreuses pour l'économie et fatale pour la vie. De tels phénomènes incluent : les ouragans, les tempêtes et la propagation des produits dangereux. Afin d'adapter et d'améliorer sa qualité de vie tout en minimisant l'impact d'éventuelles catastrophes provoquées par l'évolution de ces systèmes, il est capital pour l'homme, d'une part de chercher à mieux les comprendre et d'autre part, d'être en mesure de prévoir leur comportement. La prévision du comportement de ces systèmes a des enjeux multiples parmi lesquels :

- des enjeux environnementaux regroupant la traçabilité des gaz à effet de serre, la limitation de la pollution des nappes phréatiques ainsi que la surveillance des marées noires ;
- des enjeux économiques tel que les prévisions météorologiques et océanographiques pour le transport et la pêche. L'utilisation des prévisions météorologiques pour la gestion des conséquences du volcan islandais de 2010 en est une illustration ;
- des enjeux de sécurité publique qui regroupent entre autres : la mise en garde contre les événements météorologiques et climatiques dont les ouragans, les inondations ou la sécheresse ; la mise en garde contre la propagation de produits dangereux et l'utilité des prévisions météorologiques pour l'organisation de manifestations publiques.

Ces enjeux font de l'étude des fluides géophysiques une véritable gageure de la recherche scientifique. Ces recherches ont permis de mettre en œuvre un ensemble de moyens d'information facilitant la compréhension de ces systèmes. Ces sources d'information incluent :

- les modèles mathématiques qui décrivent la dynamique dans l'espace et dans le temps du système. Ce sont en général des équations de Navier-Stokes sur des domaines continus. On utilise dans la pratique, des modèles numériques qui sont issus de la simplification et de la discrétisation des modèles mathématiques continus. Ils sont entachés d'incertitudes liées à la simplification, la discrétisation et les traitements numériques. De plus, ils dépendent en général de paramètres qui ne sont pas connus avec précision. Le terme **paramètre** regroupe les conditions initiales pour les modèles évolutifs, les conditions aux frontières pour les domaines ouverts et les paramètres physiques du modèle ;
- les observations qui sont des mesures des variables décrivant l'état du système. Ce sont

notamment des mesures par les stations au sol des quantités physiques comme le vent pour les modèles d'atmosphère et les télémessures diverses faites par les satellites et les radars. Elles fournissent des informations sur l'état du système aux points et aux dates de mesure. Elles sont entachées d'erreurs de mesure induites par les limites des instruments de mesure. À elle seules, les observations ne peuvent renseigner ni sur le futur du système, ni sur l'état du système dans les régions non observées ;

- les statistiques qui représentent la connaissance des erreurs de modélisation et d'observation. Elles incluent les erreurs de représentativité dues aux limites des modèles numériques ;
- les informations qualitatives, c'est l'ensemble des connaissances accumulées par les experts et qui ne sont décrites ni par les modèles ni par les statistiques. L'expérience des spécialistes, est également un apport non négligeable pour l'interprétation de certains processus.

Les statistiques et les informations qualitatives sont des outils complémentaires permettant une meilleure utilisation des modèles et des observations.

Ces sources d'information, hétérogènes en nature, en quantités et en qualités sont individuellement insuffisantes pour caractériser le système étudié. Cependant, de part leurs caractéristiques, elles sont complémentaires et doivent donc être utilisées conjointement. On utilise à cet effet des techniques d'assimilation de données.

Méthodologie

Definition 1.1.1 *on désigne par **assimilation de données** l'ensemble des méthodes permettant de combiner, de façon optimale, toutes les sources d'information sur un système donné afin d'en déterminer les paramètres.*

L'assimilation de données permet entre autres de déterminer les paramètres physiques du modèle décrivant un système et d'en reconstituer un état. Ce dernier peut être : l'état actuel, on parle de *nowcast* ; un état passé, on parle de ré-analyse ou un état futur, c'est la prévision. Les applications englobent la plupart des domaines qui nécessitent la modélisation. Nous pouvons citer entre autres la météorologie, l'océanographie, l'hydrologie, la glaciologie, la sismologie, la médecine et l'agriculture.

On distingue deux grandes approches d'assimilation de données qui sont l'approche stochastique basée sur le filtrage et l'approche variationnelle basée sur la théorie du contrôle optimal. Dans l'un ou l'autre cas, l'assimilation de données permet de déterminer l'ensemble des paramètres minimisant l'écart entre les observations et les sorties du modèle en tenant compte des statistiques d'erreurs et des informations qualitatives. Les sources d'observation varient selon les systèmes. Ce sont des mesures plus ou moins directes des variables décrivant l'état du système.

Application aux observations de type image et séquence d'images

Les images constituent une des plus importantes sources d'observation des systèmes physiques. Elles ont connues une évolution importante ces dernières décennies. On a assisté à la multiplicité des satellites d'observation qui permettent d'améliorer les connaissances de l'atmosphère et des océans.

Ces satellites produisent, entre autres, des images illustrant la terre et sa couverture fluide. La suite dans le temps des mesures effectuées par satellites constituent des séquences d'images

qui montrent clairement l'évolution dynamique de certains phénomènes météorologiques et océanographiques tels que les dépressions (cyclones, ouragans, tourbillons marins etc.), les fronts, etc. Ces phénomènes sont identifiables dans les séquences d'images par des structures caractéristiques que la vision humaine peut détecter. Les images et les séquences d'images représentent ainsi une source de donnée dotée d'un potentiel informationnel important. Son utilité justifie l'importance des moyens mobilisés pour la mise en place des systèmes d'acquisition. Ces moyens incluent dans le cas des satellites, leur construction, leur mise en orbite ainsi que leur entretien. De plus, les images constituent la seule source d'observation dans certains domaines. C'est le cas de l'astrophysique où les observations conventionnelles sont presque inexistantes. Un autre exemple, c'est la médecine où on ne peut pas se permettre certaines manipulations. Jusqu'à présent, les images sont en général sous-utilisées en assimilation de données à cause de leur complexité. Leur utilisation se limite dans la plupart des cas à un aspect qualitatif tel que l'exploitation par les experts pour la correction manuelle des sorties des modèles numériques.

L'objectif de cette thèse est d'étudier les méthodes d'assimilation variationnelle de données et leur extension aux observations de type image. Concrètement, il s'agit de proposer un formalisme permettant d'utiliser de façon optimale l'information contenue dans les images conjointement avec les modèles numériques, les autres sources d'observation et les connaissances *a priori*.

Actuellement, on peut distinguer deux approches d'utilisation des séquences d'images en assimilation de données :

- la technique des **pseudo-observations** : elle consiste à estimer d'autres variables d'état du système. Ces estimations sont ensuite utilisées comme observations desdites variables. Cette approche a déjà été expérimentée dans le contexte de la prévision météorologique [Schmetz et al.(1993)] avec les pseudo-observations du vent.
- l'**assimilation directe d'images** : c'est l'utilisation directe des images dans le schéma d'assimilation variationnelle de données. Des travaux récents ont montré la faisabilité et la robustesse de cette approche [Papadakis and Mémin(2008a), Corpetti et al.(2009)]. Dans un cadre similaire, l'assimilation d'images a été utilisée pour la vision par ordinateur [Papadakis and Mémin(2008b)]. Ces travaux se limitent à la modélisation basique de l'image consistant à la considérer comme un tableau de pixels. Nous pensons qu'une représentation adaptée des images doit être utilisée afin d'exploiter au mieux leurs caractéristiques.

L'utilisation directe de séquences d'images en assimilation de données peut présenter plusieurs difficultés liées à la nature de ce type d'information :

- les images par satellites sont des mesures indirectes des variables du modèle. Les quantités observées sont plus ou moins liées aux variables du modèle par des opérateurs complexes qui doivent être pris en compte dans le processus d'assimilation ;
- en vision (humaine ou par ordinateur), l'information pertinente dans les images est localisée dans les discontinuités. Malheureusement, ces dernières sont en général lissées par les modèles numériques ;
- les processus physiques observés ne sont pas toujours représentables par le modèle numérique. Nous pouvons citer comme exemple la convection locale dans les modèles atmosphériques ou la couleur de l'eau dans les modèles océanographiques. Bien que n'étant pas propre aux observations de type image, ce problème est presque systématique avec

- cet type d'observation ;
- les images représentent des informations bidimensionnelles tandis que les processus physiques observés sont tridimensionnels. Cette propriété des images induit des différences significatives selon le système observé. Dans le cas de l'océan, les satellites produisent des observations de la surface tandis que dans le cas de l'atmosphère, ils fournissent une intégration des propriétés radiatives du système sur la verticale de la colonne observée. En outre, les phénomènes observés sont parfois localisés : les structures météorologiques tels que les nuages sont en général situées dans des couches spécifiques de l'atmosphère. Des informations supplémentaires sur l'altitude de ces phénomènes sont nécessaires pour une utilisation optimale. Dans le cas de l'océan, les observations par satellite fournissent des informations sur une fine couche de la surface, de l'ordre de quelques millimètres alors que les modèles numériques divisent l'océan en couches de l'ordre de quelques dizaines de mètres. Le phénomène observé est ainsi loin d'être reproduit par le modèle et peut représenter une surestimation exagérée de la réalité dans la couche supérieure considérée dans les modèles ;

Organisation du mémoire

Le présent document est organisé autour de deux grandes parties : une première partie constituée de deux chapitres qui fait l'état de l'art de l'assimilation variationnelle de données et une deuxième partie qui introduit l'assimilation directe de séquences d'images. La suite de ce chapitre introductif présente les notations générales utilisées dans le reste du document. Le chapitre (2) présente l'approche variationnelle d'assimilation de données. L'approche stochastique quant à elle est présentée en annexe (A) Dans le chapitre (3), nous présentons les modèles et les observations. Nous y décrivons en particulier les observations de type image dont l'utilisation en assimilation de données a motivé ce travail. Pour clore le chapitre, nous présentons les techniques actuelles d'utilisation des observations de type image. Les autres types d'observation sont présentés en annexe (C). Le chapitre (4) présente une extension du formalisme d'assimilation variationnelle de données aux observations de type image. Il marque le début de la deuxième partie du mémoire. Nous y décrivons les particularités des images qui doivent être prises en compte pour l'utilisation de cette source d'observation en assimilation de données. Nous introduisons les notions de niveau d'interprétation des images, d'opérateurs image et d'espaces d'images. A partir de ces notions, nous proposons un formalisme de construction d'opérateurs d'observation et de distances pour le calcul de l'écart aux observations de type image. Dans le chapitre (5), nous abordons le problème de la régularisation. Elle permet de faire face au caractère mal posé des problèmes inverses dont l'assimilation de données est un cas particulier. Nous faisons une revue des approches de régularisation couramment utilisées dans le contexte du traitement d'images. Nous introduisons ensuite une nouvelle approche d'utilisation de l'information de régularisation pour la résolution des problèmes inverses mal posés. Une implémentation élégante basée sur la diffusion généralisée de cette approche est proposée. Cette implémentation permet d'utiliser l'information de régularisation pour le pré-conditionnement du problème. Les résultats expérimentaux présentés démontrent clairement la supériorité de cette approche par rapport aux approches classiques. Le chapitre (6) clôt ce mémoire par une conclusion et des perspectives. Il est suivi par un ensemble d'annexes qui complète le document.

1.2 Vocabulaire

Utilisation générale des symboles

Nous introduisons dans cette section les notations qui seront utilisées pour les équations dans la suite du document. Le tableau (1.1) donne la convention générale d'utilisation des symboles dans les équations et le tableau (1.2) liste les principales notations utilisées. Nous présentons dans les paragraphes suivants, en complément des tableaux (1.1) et (1.2), la liste des symboles les plus utilisés dans ce document ; ces symboles sont définis suivant la convention de [Ide et al.(1997)].

TABLE 1.1 – Convention d'utilisation des symboles

Alphabet	Police	Usage	Exemple
Latin/grec	Gras, minuscule	Vecteurs	\mathbf{x}
Latin	Italique, minuscule	Scalaire	t
Latin	italique, majuscule	Opérateurs discret en espace	M
Latin	Gras, majuscule	Opérateurs linéaires et matrices	\mathbf{B}
Latin	calligraphie, Majuscule	Opérateurs continu et ensembles	\mathcal{M}
Grec	Minuscule	Constantes scalaire	λ

Principaux symboles utilisés

Variable d'état du modèle et ses déclinaisons

On appelle variable d'état ou vecteur d'état l'ensemble des variables décrivant le système étudié. Ces variables incluent : le courant, la température, l'élévation de la surface libre, la salinité, la concentration des espèces chimiques pour l'océan ; le vent, la pression, la température, l'humidité pour l'atmosphère. Ce sont des fonctions de l'espace et du temps. Nous noterons :

- $\mathbf{x}(t)$, le vecteur d'état au temps t . Il représente une discrétisation spatiale (différences finies, volumes finis, éléments finis, \dots) des variables continues. Partout où il n'y aura pas d'ambiguïté nous noterons simplement \mathbf{x} pour $\mathbf{x}(t)$.
- $\mathbf{x}_i = \mathbf{x}(t_i)$ le vecteur d'état au temps t_i dans un cadre discret en temps
- \mathbf{x}^t , (t pour 'true') l'état vrai ;
- \mathbf{x}^b , (b pour 'background') l'ébauche, c'est la valeur de la variable d'état avant assimilation de données ;
- \mathbf{x}^a , (a pour 'analysed') l'état analysé, c'est la valeur de la variable d'état calculée par assimilation de données ;
- $\langle \cdot, \cdot \rangle$ le produit scalaire dans \mathbb{R}^n ;
- $\|\mathbf{x}\|_{\mathbf{B}}^2$ la norme associée au produit scalaire $\langle \mathbf{x}, \mathbf{x} \rangle_{\mathbf{B}} = \mathbf{x}^T \mathbf{B} \mathbf{x}$ avec \mathbf{B} une matrice définie positive.

Modèle

Modèle continu

L'évolution de la variable d'état \mathbf{x} obéit en général à un système d'équations aux dérivées partielles que nous noterons :

TABLE 1.2 – Convention d'utilisation des symboles

Symbole	Définition	Remarques
Principaux vecteurs		
\mathbf{x}	Vecteur d'état	
$\delta\mathbf{x}$	Incrément sur le vecteur d'état	
\mathbf{y}^o	Vecteur d'observation	mesures
η	Erreur modèle	
ε	Erreur d'observation	
Principaux opérateurs discrets		
H	Opérateur d'observation	
J	Fonction coût	Même notation en continu
M	Modèle dynamique	
Principaux opérateurs non-linéaires		
\mathcal{H}	Opérateur d'observation	
\mathcal{M}	Modèle dynamique	
Principaux Espaces		
\mathcal{X}	Espace d'état	
\mathcal{O}	Espace d'observation	
Principaux opérateurs linéaires matriciels		
\mathbf{B}	Covariance d'erreur d'ébauche	
\mathbf{H}	Opérateur d'observation linéarisé	
\mathbf{I}	Identité	
\mathbf{K}	Matrice de gain	
\mathbf{P}	Covariance d'erreur d'analyse	
\mathbf{Q}	Covariance d'erreur modèle	
\mathbf{R}	Covariance erreurs d'observation	
Utilisation des exposants		
$(\cdot)^{-1}$	Inverse	
$(\cdot)^{-I}$	Inverse généralisé	
$(\cdot)^T$	Transposé	
$(\cdot)^a$	Analyse	calculé par le modèle numérique
$(\cdot)^b$	Ébauche	b pour background
$(\cdot)^f$	Prévision	f pour forecast
$(\cdot)^o$	Observé	
$(\cdot)^t$	Vrai	t pour true

$$\begin{cases} \frac{d\mathbf{x}}{dt} = \mathcal{M}(\mathbf{x}), & t \in [0, T], \\ \mathbf{x}(0) = \mathbf{u}. \end{cases} \quad (1.1)$$

avec \mathbf{u} donné.

Modèle discret

Dans le cadre discret, nous noterons \mathcal{M}_i le modèle d'évolution de l'état du système du pas de temps i au pas de temps $i + 1$ avec un schéma explicite s'écrit :

$$\begin{cases} \mathbf{x}_{i+1} &= \mathcal{M}_i(\mathbf{x}_i), i = 1, 2, \dots \\ \mathbf{x}_0 &= \mathbf{u}, \end{cases} \quad (1.2)$$

C'est une écriture simplifiée mais suffisante pour présenter clairement le principe de l'assimilation de données. Dans un cas réaliste, on peut être amené à écrire l'évolution de la variable d'état en fonction des conditions aux frontières pour les systèmes ouverts et du jeu de paramètres d'évolution du système. Ces éléments communément appelés paramètres (au sens général du terme) peuvent être considérés comme inclus dans l'opérateur \mathcal{M} selon qu'ils soient connus ou pas.

Observations et opérateur d'observation

Nous noterons :

- $\mathbf{y}^o \in \mathcal{O}$, le vecteur d'observations ;
- \mathcal{H} , l'opérateur d'observation qui fait le lien entre l'état du système et les observations ;
- $\mathbf{y} = \mathcal{H}(\mathbf{x})$, l'équivalent des observations calculé à partir des sorties du modèle numérique.

Statistiques d'erreurs

Nous noterons :

- $\varepsilon^b = \mathbf{x}^b - \mathbf{x}^t$, l'erreur d'ébauche et \mathbf{B} la matrice de covariance associée ;
- $\varepsilon^o = \mathbf{y}^o - \mathcal{H}(\mathbf{x}^t)$, l'erreur d'observation et \mathbf{R} la matrice de covariance associée ;
- $\nu_i : \mathbf{x}_{i+1} = \mathcal{M}_i(\mathbf{x}_i^t) + \nu_i$, l'erreur modèle covariance et \mathbf{Q}_i la matrice de covariance associée ;
- $\varepsilon^a = \mathbf{x}^a - \mathbf{x}^t$, l'erreur d'analyse et \mathbf{P}^a la matrice de covariance associée.

Première partie

Assimilation de données : concepts
fondamentaux

Chapitre 2

Approche variationnelle de l'assimilation de données

2.1 Introduction à la théorie de l'estimation

Nous introduisons dans cette section quelques notions de la théorie de l'estimation nécessaires à la compréhension de la suite du chapitre. Pour une description détaillée de cette théorie, il existe d'excellents ouvrages parmi lesquels [Scharf(1991)], [Kay(1993)] et [Levy(2008)].

2.1.1 Estimation probabiliste

On s'intéresse au problème d'estimation des valeurs d'un vecteur \mathbf{x} de n paramètres d'un système donné, sur la base de l'observation de p valeurs parmi les n , et stockées dans le vecteur d'observation \mathbf{y}^o : on parle aussi de problème inverse¹. Nous notons \mathbf{x}^a le vecteur estimé de paramètres, \mathbf{x}^t le vecteur recherché et $\varepsilon^a = \mathbf{x}^a - \mathbf{x}^t$ l'erreur d'estimation. On suppose également que le processus d'observation n'est pas parfait, ce qui permet de considérer les observations comme la réalisation d'une variable aléatoire. Parallèlement, on considère \mathbf{x} comme la réalisation d'un vecteur aléatoire.

Approche bayésienne

Dans l'approche bayésienne de l'estimation, on introduit la fonctionnelle J de risque associé à l'estimation obtenue. Cette fonctionnelle est définie comme suit :

$$J(\mathbf{x}^a) = \int_{-\infty}^{+\infty} \mathcal{J}(\varepsilon^a) p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} \quad (2.1)$$

où $p_{\mathbf{x}}(\mathbf{x})$ est la densité de probabilité marginale de \mathbf{x} ; \mathcal{J} une fonction d'évaluation du critère de risque. J est souvent appelée fonction coût. Si on dispose de l'observation \mathbf{y}^o , alors J peut se réécrire en fonction de cette information comme suit :

$$J(\mathbf{x}^a) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathcal{J}(\varepsilon^a) p_{\mathbf{xy}}(\mathbf{x}, \mathbf{y}^o) dy d\mathbf{x} \quad (2.2)$$

où $p_{\mathbf{xy}}(\mathbf{x}, \mathbf{y}^o)$ est la densité de probabilité jointe du couple $(\mathbf{x}, \mathbf{y}^o)$ des paramètres et des observations. La détermination de \mathbf{x}^a se ramène alors à la minimisation du risque. \mathbf{x}^a ainsi

1. le problème direct consiste à déterminer l'état du système à partir des paramètres

déterminé est appelé estimation optimale. Cette estimation dépend en général de la fonction de coût utilisée. Une des principales propriétés souhaitées pour une estimation est qu'elle soit non biaisée ; ou en d'autres termes que l'erreur d'estimation soit de moyenne nulle.

Les fonctions usuelles d'évaluation du critère de risque sont :

$$\mathcal{J}_1(\mathbf{x}) = \mathbf{x}^T \mathbf{E} \mathbf{x}, \quad (2.3)$$

où la matrice carrée \mathbf{E} de taille $n \times n$ est supposée symétrique et définie positive ; et :

$$\mathcal{J}_2(\mathbf{x}) = \begin{cases} 0, & \|\mathbf{x}\| < \epsilon \\ 1/2\epsilon, & \|\mathbf{x}\| \geq \epsilon \end{cases} \quad (2.4)$$

Ces deux fonctions permettent de définir respectivement l'estimation optimale de variance minimale et l'estimation de probabilité *a posteriori* maximale.

Estimateur du minimum de variance

L'estimateur du minimum de variance \mathbf{x}_{MV}^a (Minimum Variance estimator en anglais) minimise le risque défini par la fonction d'évaluation du critère de risque (2.3) comme suit :

$$J_{MV}(\mathbf{x}^a) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\mathbf{x} - \mathbf{x}^a)^T E(\mathbf{x} - \mathbf{x}^a) p_{\mathbf{xy}}(\mathbf{x}, \mathbf{y}^o) d\mathbf{y} d\mathbf{x}. \quad (2.5)$$

En utilisant la loi de Bayes :

$$p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}^o) = \frac{p_{\mathbf{y}}(\mathbf{y}^o|\mathbf{x})p_{\mathbf{x}}(\mathbf{x})}{p_{\mathbf{y}}(\mathbf{y}^o)} \quad (2.6)$$

basée sur les lois de probabilités conditionnelles :

$$p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}) = \frac{p_{\mathbf{xy}}(\mathbf{x}, \mathbf{y})}{p_{\mathbf{y}}(\mathbf{y})}; \quad (2.7)$$

$$p_{\mathbf{y}}(\mathbf{y}|\mathbf{x}) = \frac{p_{\mathbf{xy}}(\mathbf{x}, \mathbf{y})}{p_{\mathbf{x}}(\mathbf{x})}. \quad (2.8)$$

La fonction de risque peut se récrire :

$$J_{MV}(\mathbf{x}^a) = \int_{-\infty}^{+\infty} \left\{ \int_{-\infty}^{+\infty} (\mathbf{x} - \mathbf{x}^a)^T E(\mathbf{x} - \mathbf{x}^a) p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}^o) d\mathbf{x} \right\} p_{\mathbf{y}}(\mathbf{y}^o) d\mathbf{y}. \quad (2.9)$$

$p_{\mathbf{y}}(\mathbf{y}^o)$ étant toujours positif comme une densité de probabilité, et l'intégrale externe ne faisant pas intervenir \mathbf{x}^a , la minimisation de la fonction (2.9) peut se ramener à celle de la fonction équivalente :

$$J_{MV}(\mathbf{x}^a) = \int_{-\infty}^{+\infty} (\mathbf{x} - \mathbf{x}^a)^T E(\mathbf{x} - \mathbf{x}^a) p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}^o) d\mathbf{x}. \quad (2.10)$$

On peut facilement démontrer en utilisant la différentiation d'une fonction scalaire et l'équation d'Euler Lagrange que la solution optimale est donnée par :

$$\mathbf{x}_{MV}^a(\mathbf{y}^o) = \int_{-\infty}^{+\infty} \mathbf{x} p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}^o) d\mathbf{x}. \quad (2.11)$$

Estimateur du maximum de probabilité *a posteriori*

En utilisant la fonction d'évaluation du critère de risque (2.4) dans la fonction de risque, on définit l'estimateur du maximum de probabilité *a posteriori* que nous noterons *MAP* (Maximum A Posteriori probability estimation) comme le minimum de la fonctionnelle :

$$J_{MAP}(\mathbf{x}^a) = \int_{-\infty}^{+\infty} \frac{1}{2\epsilon} \left\{ \int_{-\infty}^{\mathbf{x}^a - \epsilon} p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}^o) d\mathbf{x} + \int_{\mathbf{x}^a + \epsilon}^{+\infty} p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}^o) d\mathbf{x} \right\} p_{\mathbf{y}}(\mathbf{y}^o) d\mathbf{y}, \quad (2.12)$$

$\mathbf{x}^a \pm \epsilon$ représentant le vecteur $(\mathbf{x}_i^a \pm \epsilon)_{i=1 \dots n}$.

En prenant en compte les considérations suivantes :

- $p_{\mathbf{x}}$ est une densité de probabilité; $\int_{-\infty}^{+\infty} p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}^o) d\mathbf{x} = 1$;
- $p_{\mathbf{y}}(\mathbf{y}^o)$ est positif en tant densité de probabilité.

On peut ramener la minimisation de la fonction (2.12) à celle de la fonctionnelle :

$$J_{MAP}(\mathbf{x}^a|\mathbf{y}^o) = -\frac{1}{2\epsilon} \int_{\mathbf{x}^a - \epsilon}^{\mathbf{x}^a + \epsilon} p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}^o) d\mathbf{x}. \quad (2.13)$$

Si on fait tendre ϵ vers 0, le théorème de la valeur moyenne pour les intégrales nous permet de réduire le problème à la minimisation de :

$$J_{MAP}(\mathbf{x}^a|\mathbf{y}^o) = -p_{\mathbf{x}}(\mathbf{x}^a|\mathbf{y}^o). \quad (2.14)$$

La minimisation de cette fonctionnelle est équivalente à la maximisation de la probabilité conditionnelle $p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}^o)$, d'où l'expression probabilité *a posteriori* maximale. La résolution de l'équation d'Euler-Lagrange :

$$\frac{\partial p_{\mathbf{x}}(\mathbf{x}^a|\mathbf{y}^o)}{\partial \mathbf{x}^a} = \frac{\partial p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}^o)}{\partial \mathbf{x}} = 0 \quad (2.15)$$

permet de déterminer le vecteur de paramètre optimal.

On peut également déterminer le vecteur optimal comme celui maximisant le logarithme de la densité de probabilité. En utilisant la loi de Bayes, ce logarithme peut s'écrire :

$$\ln p_{\mathbf{x}}(\mathbf{x}|\mathbf{y}^o) = \ln [p_{\mathbf{y}}(\mathbf{y}^o|\mathbf{x})p_{\mathbf{x}}(\mathbf{x})] - \ln p_{\mathbf{y}}(\mathbf{y}^o). \quad (2.16)$$

$p_{\mathbf{y}}(\mathbf{y}^o)$ ne dépendant pas de \mathbf{x} , on peut laisser le terme $\ln p_{\mathbf{y}}(\mathbf{y}^o)$, \mathbf{x}_{MAP}^a est alors solution de :

$$\frac{\partial \ln [p_{\mathbf{y}}(\mathbf{y}^o|\mathbf{x})p_{\mathbf{x}}(\mathbf{x})]}{\partial \mathbf{x}} = 0, \quad (2.17)$$

ou son équivalent :

$$\frac{\partial p_{\mathbf{y}}(\mathbf{y}^o|\mathbf{x})p_{\mathbf{x}}(\mathbf{x})}{\partial \mathbf{x}} = 0. \quad (2.18)$$

Cet estimateur n'est pas garanti sans biais dans le cas général.

Approche du maximum de vraisemblance

Pour dériver l'estimateur du maximum de probabilité *a posteriori*, on suppose connue, la loi de probabilité de la variable aléatoire \mathbf{x} ; ce qui n'est pas toujours le cas dans la pratique. L'estimateur du maximum de vraisemblance que nous noterons *ML* (Maximum Likelihood estimation en anglais) est une simplification de l'estimateur du maximum de probabilité *a posteriori*; simplification basée sur le manque d'information sur la variable aléatoire à déterminer.

Si on suppose que \mathbf{x} suit une loi gaussienne de moyenne $\mu_{\mathbf{x}}$ et de covariance $\mathbf{P}_{\mathbf{x}}$, alors sa densité de probabilité est donnée par :

$$p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} \|\mathbf{P}_{\mathbf{x}}\|^{1/2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mu_{\mathbf{x}})^T \mathbf{P}_{\mathbf{x}}^{-1} (\mathbf{x} - \mu_{\mathbf{x}}) \right], \quad (2.19)$$

de logarithme :

$$\ln p_{\mathbf{x}}(\mathbf{x}) = -\ln \left[(2\pi)^{n/2} \|\mathbf{P}_{\mathbf{x}}\|^{1/2} \right] - \frac{1}{2} \left[(\mathbf{x} - \mu_{\mathbf{x}})^T \mathbf{P}_{\mathbf{x}}^{-1} (\mathbf{x} - \mu_{\mathbf{x}}) \right]. \quad (2.20)$$

La dérivé logarithmique de $p_{\mathbf{x}}(\mathbf{x})$ peut alors s'écrire :

$$\frac{\partial \ln p_{\mathbf{x}}(\mathbf{x})}{\partial \mathbf{x}} = -\mathbf{P}_{\mathbf{x}}^{-1} (\mathbf{x} - \mu_{\mathbf{x}}). \quad (2.21)$$

Si on manque d'informations sur $p_{\mathbf{x}}(\mathbf{x})$, ce qui se traduit par une variance infinie ($\mathbf{P}_{\mathbf{x}} \rightarrow \infty$) et donc $\mathbf{P}_{\mathbf{x}}^{-1} \rightarrow 0$, (2.17) devient :

$$\frac{\partial \ln p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}{\partial \mathbf{x}} = 0, \quad (2.22)$$

ou son équivalent :

$$\frac{\partial p_{\mathbf{y}}(\mathbf{y}|\mathbf{x})}{\partial \mathbf{x}} = 0. \quad (2.23)$$

La solution de cette équation est appelée estimateur du maximum de vraisemblance. Comme l'estimateur du maximum de probabilité *a posteriori*, cet estimateur n'est pas garanti sans biais. De plus il n'est pas toujours fiable à cause des hypothèses de sa dérivation.

Remarque Si les erreurs sont gaussiennes, l'estimateur du maximum de vraisemblance est équivalent à l'estimateur du minimum de variance. Cela n'est pas vrai dans le cas général.

2.1.2 Estimation par moindre carré

Dans les techniques d'estimation probabilistes, on suppose connues les statistiques sur toutes (ou presque) les variables aléatoires prises en compte. La méthode du maximum de vraisemblance est obtenue en relaxant les contraintes de connaissances statistiques sur la variable à déterminer. En allant plus loin, on peut supposer inconnues toutes les statistiques prises en compte jusqu'ici ; on détermine dans ce cas un estimateur optimal au sens des moindres carrés. Nous le noterons *MS* (Mean Square en anglais).

Considérons à nouveau le système de la section précédente, d'état $\mathbf{x} \in \mathbb{R}^n$. On suppose qu'on a acquis un vecteur d'observations $\mathbf{y}^o \in \mathbb{R}^m$ qui satisfait la relation :

$$\mathbf{y}^o = \mathbf{H}\mathbf{x} + \varepsilon^o, \quad (2.24)$$

où \mathbf{H} est une matrice $m \times n$ (aussi appelé opérateur d'observation) et ε^o l'erreur d'observation. L'estimateur de moindre carré détermine \mathbf{x}^a comme minimum de la fonction :

$$J_{LS}(\mathbf{x}) = \frac{1}{2} \|\mathbf{H}\mathbf{x} - \mathbf{y}^o\|^2. \quad (2.25)$$

J est une fonction déterministe, ce qui permet d'utiliser les techniques d'optimisation usuelles pour déterminer son minimum. La solution \mathbf{x}_{LS}^a à ce problème est donnée par l'équation d'Euler-Lagrange :

$$\nabla J_{LS}() = \mathbf{H}^T(\mathbf{H}\mathbf{x} - \mathbf{y}^o) = 0. \quad (2.26)$$

Ce qui permet moyennant l'existence de $(\mathbf{H}^T\mathbf{H})^{-1}$, de déterminer \mathbf{x}_{LS}^a comme :

$$\mathbf{x}_{LS}^a = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}\mathbf{y}^o. \quad (2.27)$$

Cette méthode peut se généraliser en utilisant une matrice de pondération définie par la confiance relative accordée à chaque mesure. La fonction coût se réécrit alors :

$$J_{LS}(\mathbf{x}) = \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y}^o)^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}^o), \quad (2.28)$$

où \mathbf{R}^{-1} est la matrice de poids, symétrique et définie positive. L'écriture de \mathbf{R}^{-1} sous la forme $\mathbf{R}^{-1} = \mathbf{C}\mathbf{C}^T$ (factorisation de Cholesky) permet de récrire $J_{LS}(\mathbf{x})$ comme :

$$J_{LS}(\mathbf{x}) = \frac{1}{2}(\mathbf{C}(\mathbf{H}\mathbf{x} - \mathbf{y}^o))^T \mathbf{C}(\mathbf{H}\mathbf{x} - \mathbf{y}^o). \quad (2.29)$$

Si on fait le changement de variable $\bar{\mathbf{y}}^o := \mathbf{C}\mathbf{y}^o$, $\bar{\mathbf{H}} := \mathbf{C}\mathbf{H}$, on se ramène au cas précédent et la solution au problème de minimisation est donnée par :

$$\mathbf{x}_{LS}^a = (\bar{\mathbf{H}}^T \mathbf{R}^{-1} \bar{\mathbf{H}})^{-1} \bar{\mathbf{H}} \mathbf{R}^{-1} \bar{\mathbf{y}}^o. \quad (2.30)$$

2.2 De l'estimation probabiliste à l'assimilation variationnelle

2.2.1 Généralisation de l'approche du maximum de vraisemblance

Nous avons présenté au paragraphe (2.1.1) l'approche du maximum de vraisemblance pour la détermination des paramètres ou de l'état d'un système, connaissant des observations du système. Dans un cadre plus général, si l'on dispose en plus des observations d'une approximation \mathbf{x}^b , ébauche du vecteur \mathbf{x} à déterminer, l'estimateur du maximum de variance peut se récrire comme :

$$\mathbf{x}_{ML}^a = \max_{\mathbf{x}} p_{\mathbf{x}}(\mathbf{x} | \mathbf{y}^o \text{ et } \mathbf{x}^b) \quad (2.31)$$

ou de façon équivalente :

$$\mathbf{x}_{ML}^a = \max_{\mathbf{x}} \ln p_{\mathbf{y}}(\mathbf{y}^o \text{ et } \mathbf{x}^b | \mathbf{x}) \quad (2.32)$$

où $p_{\mathbf{x}}(\mathbf{x} | \mathbf{y}^o \text{ et } \mathbf{x}^b) = p_{\mathbf{y}\mathbf{x}}(\mathbf{y}^o \text{ et } \mathbf{x}^b | \mathbf{x})$ désigne la probabilité conditionnelle de \mathbf{x} connaissant \mathbf{y}^o et \mathbf{x}^b . Si nous définissons la fonction coût \mathbf{J} par :

$$J(\mathbf{x}) = -\ln p_{\mathbf{y}}(\mathbf{y}^o \text{ et } \mathbf{x}^b | \mathbf{x}), \quad (2.33)$$

alors l'estimateur du maximum de vraisemblance généralisé s'écrit :

$$\mathbf{x}_{ML}^a = \min_{\mathbf{x}} J(\mathbf{x}). \quad (2.34)$$

C'est l'approche variationnelle de données pour la résolution du problème d'estimation. Sous l'hypothèse de la non corrélation des erreurs d'ébauche et d'observation, auquel cas :

$$p_{\mathbf{y}}(\mathbf{y}^o \text{ et } \mathbf{x}^b | \mathbf{x}) = p_{\mathbf{y}\mathbf{x}}(\mathbf{y}^o | \mathbf{x}) p_{\mathbf{x}}(\mathbf{x}^b | \mathbf{x}), \quad (2.35)$$

la fonction coût devient :

$$J(\mathbf{x}) = -\ln p_{\mathbf{y}}(\mathbf{y}^o|\mathbf{x}) - \ln p_{\mathbf{x}}(\mathbf{x}^b|\mathbf{x}). \quad (2.36)$$

Si en plus les distributions de probabilités sont gaussiennes :

$$p_{\mathbf{x}}(\mathbf{x}^b|\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\|\mathbf{B}\|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b)\right], \quad (2.37)$$

$$p_{\mathbf{y}}(\mathbf{y}^o|\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\|\mathbf{R}\|^{1/2}} \exp\left[-\frac{1}{2}(\mathcal{H}(\mathbf{x}) - \mathbf{y}^o)^T \mathbf{R}^{-1}(\mathcal{H}(\mathbf{x}) - \mathbf{y}^o)\right], \quad (2.38)$$

avec \mathbf{B} la matrice de covariance d'erreur sur l'ébauche, \mathbf{R} la matrice de covariance d'erreur sur les observations et \mathcal{H} l'opérateur d'observation qui calcule la contrepartie observation du modèle. On peut utiliser cette approximation de la fonction coût devient :

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathcal{H}(\mathbf{x}) - \mathbf{y}^o)^T \mathbf{R}^{-1}(\mathcal{H}(\mathbf{x}) - \mathbf{y}^o). \quad (2.39)$$

C'est la fonction coût de l'assimilation variationnelle de données. Les constantes additives qui n'ont pas d'influence sur la minimisation n'y figurent pas. $(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b)$ est appelé terme de **rappel à l'ébauche** et $(\mathcal{H}(\mathbf{x}) - \mathbf{y}^o)^T \mathbf{R}^{-1}(\mathcal{H}(\mathbf{x}) - \mathbf{y}^o)$ est appelé terme d'**écart aux observations**.

Dans l'approche variationnelle, l'assimilation de données est formulée comme un problème d'optimisation. Cette approche a été introduite par Sasaki [Sasaki(1955), Sasaki(1958)] et a été adoptée depuis quelques années par les principaux centres opérationnels de prévision météorologique. Le problème d'optimisation est formulé comme la minimisation de l'écart (au sens des moindres carrés) entre l'état analysé et les observations. Le modèle est ainsi utilisé comme une contrainte du problème d'optimisation. Nous allons présenter le problème d'assimilation variationnelle de données pour le contrôle de la condition initiale, le contrôle des autres paramètres se faisant suivant le même principe. Il s'agit de combiner de façon optimale les différentes sources d'information disponibles (modèles, observation, statistiques d'erreurs, ...) pour déterminer l'état initial du système.

Remarque l'approche du maximum de vraisemblance est une approche générale. La connaissance des distributions de probabilité permet de définir la fonction coût et de se ramener à l'assimilation variationnelle de données. Cependant, la détermination d'un minimum peut se révéler difficile dans le cas de distributions non gaussiennes.

2.2.2 Assimilation variationnelle 4D

La fonction coût telle que formulée à l'équation (2.39) est une fonction générique; la spécification de la variable \mathbf{x} et de l'opérateur d'observation \mathcal{H} permet de définir des schémas particuliers d'assimilation variationnelle de données. Celui qui nous intéresse est le schéma variationnel 4D encore appelé 4D-VAR.

Si \mathbf{x} représente l'état du système sur le domaine spatial 3D dans une fenêtre de temps où les observations sont disponibles, et \mathcal{H} l'opérateur d'observation spatio-temporel, alors la fonction coût se récrit :

$$\begin{aligned}
J(\mathbf{x}) &= \frac{1}{2} \left(\mathbf{x}(0) - \mathbf{x}^b \right)^T \mathbf{B}^{-1} \left(\mathbf{x}(0) - \mathbf{x}^b \right) \\
&+ \frac{1}{2} \int_0^T (\mathcal{H}(\mathbf{x}(t), t) - \mathbf{y}^o(t))^T \mathbf{R}(t)^{-1} (\mathcal{H}(\mathbf{x}(t), t) - \mathbf{y}^o(t)) dt.
\end{aligned} \tag{2.40}$$

Avec $\mathbf{x}(t)$, $\mathcal{H}(\cdot, t)$, $\mathbf{y}^o(t)$ et $\mathbf{R}(t)$ l'état du système, l'opérateur spatial au temps t , l'observation au temps t et la matrice de covariance d'erreur d'observation respectivement. L'ébauche est réduite à sa valeur au temps 0 pour des raisons pratiques qui émergeront par la suite.

De part les propriétés des matrices de covariance d'erreurs, $\mathbf{x}^T \mathbf{B}^{-1} \mathbf{y}$ définit un produit scalaire que nous noterons $\langle \cdot, \cdot \rangle_{\mathbf{B}}^{-1}$ et la norme associée $\| \cdot \|_{\mathbf{B}}^{-1}$; la fonction coût peut ainsi se récrire :

$$\begin{aligned}
J(\mathbf{x}) &= \frac{1}{2} \left\langle \mathbf{x}(0) - \mathbf{x}^b, \mathbf{x}(0) - \mathbf{x}^b \right\rangle_{\mathbf{B}}^{-1} \\
&+ \frac{1}{2} \int_0^T \langle \mathcal{H}(\mathbf{x}(t), t) - \mathbf{y}^o(t) | \mathcal{H}(\mathbf{x}(t), t) - \mathbf{y}^o(t) \rangle_{\mathbf{R}(t)^{-1}} dt,
\end{aligned} \tag{2.41}$$

ou encore :

$$J(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}(0) - \mathbf{x}^b\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \int_0^T \|\mathcal{H}(\mathbf{x}(t), t) - \mathbf{y}^o(t)\|_{\mathbf{R}(t)^{-1}}^2 dt. \tag{2.42}$$

C'est la version continue en temps de la fonction coût de l'assimilation variationnelle de données qui suppose les observations continues en temps. Ce qui n'est pas le cas dans la pratique. On discrétise alors le modèle en temps et le terme d'écart aux observations est la somme des écarts aux observations spatiales des différents pas de temps :

$$J(\mathbf{x}) = \frac{1}{2} \|\mathbf{x}(0) - \mathbf{x}^b\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \sum_{i=0}^T \|\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i^o\|_{\mathbf{R}_i^{-1}}^2 dt. \tag{2.43}$$

Avec \mathbf{x}_i , \mathcal{H}_i , \mathbf{y}_i^o et \mathbf{R}_i l'état du système, l'opérateur d'observation 3D, le vecteur d'observation et la matrice de covariance d'erreurs d'observation au temps t_i respectivement.

2.3 Assimilation variationnelle de données sous contraintes fortes

2.3.1 Le modèle direct

Ici, l'état du système est régi par un modèle d'évolution encore appelé modèle direct dans la communauté d'assimilation de données. Ce modèle définit l'état du système comme la solution du système d'équations (1.1) que nous reprenons :

$$\begin{cases} \frac{d\mathbf{x}}{dt} &= \mathcal{M}(\mathbf{x}), \quad t \in [0, T], \\ \mathbf{x}(0) &= \mathbf{u}. \end{cases} \tag{2.44}$$

Une discrétisation en temps de ce modèle continu avec un schéma d'Euler explicite d'ordre 1 permet d'obtenir le modèle direct discrétisé :

$$\begin{cases} \frac{\mathbf{x}_{i+1} - \mathbf{x}_i}{\Delta t} &= \mathcal{M}_i(\mathbf{x}_i), \quad i \in [0, N], \\ \mathbf{x}_0 &= \mathbf{u}. \end{cases} \tag{2.45}$$

Avec Δt le pas de temps de discrétisation, $N + 1$ le nombre de points de discrétisation, $T = N\Delta t$, $\mathbf{x}_i = x(i\Delta t)$ l'état 3D du système au pas de temps i , $\mathcal{M}_i = \mathcal{M}_{i \rightarrow i+1}$ le modèle d'évolution du système entre le temps $i\Delta t$ et le temps $(i + 1)\Delta t$. Il est courant de faire le changement de variable $\mathcal{M}_i(\mathbf{x}_i) := \mathbf{x}_i + \Delta t \mathcal{M}_i(\mathbf{x}_i)$ pour récrire le modèle discret comme :

$$\begin{cases} \mathbf{x}_{i+1} &= \mathcal{M}_i(\mathbf{x}_i), & i \in [0, N], \\ \mathbf{x}_0 &= \mathbf{u}. \end{cases} \quad (2.46)$$

L'état du système est ainsi défini comme fonction de \mathbf{u} que nous appellerons également variable de contrôle.

2.3.2 La fonction coût

Modèle et observations continus en temps

Avec le modèle et les observations continus en temps, la fonction coût peut s'écrire comme une fonction de l'état initial \mathbf{u} du système :

$$J(\mathbf{u}) = \frac{1}{2} \|\mathbf{u} - \mathbf{u}^b\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \int_0^T \|\mathcal{H}(\mathbf{x}) - \mathbf{y}^o\|_{\mathbf{R}(t)^{-1}}^2 dt, \quad (2.47)$$

ou, en termes de produit scalaire :

$$J(\mathbf{u}) = \frac{1}{2} \langle \mathbf{u} - \mathbf{u}^b, \mathbf{B}^{-1}(\mathbf{u} - \mathbf{u}^b) \rangle + \frac{1}{2} \int_0^T \langle \mathcal{H}(\mathbf{x}) - \mathbf{y}^o, \mathbf{R}^{-1}(\mathcal{H}(\mathbf{x}) - \mathbf{y}^o) \rangle dt. \quad (2.48)$$

C'est la version continue de la fonction coût qui suppose que le modèle et les observations sont continus en temps.

Modèle et observations discrets en temps

Dans la pratique, avec les modèles discrets et les observations non continues en temps, l'intégrale est remplacée par une somme discrète. La fonction coût devient :

$$J(\mathbf{u}) = \frac{1}{2} \|\mathbf{u} - \mathbf{u}^b\|_{\mathbf{B}^{-1}}^2 + \frac{1}{2} \sum_{i=0}^N \|\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i^o\|_{\mathbf{R}_i^{-1}}^2 \Delta t, \quad (2.49)$$

ou, en termes de produit scalaire :

$$J(\mathbf{u}) = \frac{1}{2} \langle \mathbf{u} - \mathbf{u}^b, \mathbf{B}^{-1}(\mathbf{u} - \mathbf{u}^b) \rangle + \frac{1}{2} \sum_{i=0}^N \langle \mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i^o, \mathbf{R}_i^{-1}(\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i^o) \rangle \Delta t. \quad (2.50)$$

Le problème d'assimilation de données se définit alors comme :

$$\begin{cases} \text{Trouver } \mathbf{u}^* \in \mathcal{X} \text{ tel que :} \\ J(\mathbf{u}^*) = \text{Inf}_{\mathbf{u} \in \mathcal{X}} J(\mathbf{u}), \end{cases} \quad (2.51)$$

C'est un problème d'optimisation sous contrainte du modèle. La contrainte forte est le modèle qui est supposé parfait. Il convient de distinguer cette contrainte de celles qui peuvent être imposées aux variables du modèle ; par exemple, la concentration d'un espèce chimique ne peut

être négative. Si J est strictement convexe, semi-continue inférieure et $\lim_{\|\mathbf{u}\| \rightarrow +\infty} J(\mathbf{u}) \rightarrow +\infty$, (2.51) admet une solution unique. Si J est différentiable, une condition nécessaire pour que \mathbf{u}^* soit solution de (2.51) est donnée par l'équation d'Euler-Lagrange :

$$\nabla J(\mathbf{u}^*) = 0. \quad (2.52)$$

Les méthodes d'optimisation déterministes permettent de déterminer la solution de ce problème. Les méthodes de minimisation en constituent un cas particulier.

2.4 Méthodes de minimisation de la fonction coût

Les méthodes de minimisation constituent un cas particulier de méthodes d'optimisation. Elles peuvent être utilisées pour la résolution des problèmes inverses dont l'assimilation de données constitue un cas particulier. Les problèmes inverses sont en général mal posés et/ou mal conditionnés. Ces défauts doivent être pris en compte dans leur résolution.

Dans les situations réalistes, les observations utilisées pour la résolution des problèmes inverses sont en quantité très limitée par rapport à la dimension du problème. A titre d'illustration, les modèles numériques en météorologie sont actuellement décrits par environ 10^9 variables tandis que le nombre d'observations est de l'ordre de 10^6 à 10^7 . On dit que le problème est sous déterminé. Dans certains cas, il peut y avoir plus d'observations que de paramètres à déterminer, on dit que le problème est surdéterminé. On obtient des systèmes surdéterminés notamment quand la résolution des observations est plus fine que la résolution des modèles. Ces défauts font que le problème ne soit pas bien posé au sens d'Hadamard : on dit qu'il est mal posé. C'est un défaut très souvent rencontré avec les problèmes inverses. Pour prendre en considération ce défaut dans la résolution de ce type de problème, on utilise des connaissances *a priori* sur le problème. Les connaissances *a priori* les plus souvent utilisées portent sur la régularité de la solution : on parle de régularisation. Le chapitre (5) y est entièrement consacré.

On parle de mauvais conditionnement² lorsque de légères variations dans les données entraînent de grandes variations dans la solution du problème. Lorsqu'il s'agit d'un problème inverse, les données correspondent au modèle et aux observations et la solution correspond aux paramètres estimés. En cas du mauvais conditionnement, les techniques de préconditionnement sont utilisées. Elles permettent d'améliorer le conditionnement du problème.

Dans la suite de cette section, nous présenterons les méthodes de résolution des problèmes de minimisation ainsi que la notion de préconditionnement.

2.4.1 Théorie de l'optimisation

2.4.2 Rappel sur les notions d'optimisation

Nous définissons dans cette section les principales notations qui seront utilisées dans cette partie.

2. En analyse numérique, le conditionnement est une notion qui permet de mesurer la dépendance de la solution d'un problème numérique aux données du problème, cela permet de contrôler la validité d'une solution calculée.

Définition Soit \mathcal{V} un espace vectoriel, et \mathcal{U} un sous ensemble de \mathcal{V} , un problème d'optimisation sur \mathcal{U} peut être modélisé par une minimisation de fonction; trouver $\mathbf{u} \in \mathcal{U}$ qui minimise f :

$$\mathbf{u} = \text{ArgMin } f(\mathbf{v}), \mathbf{v} \in \mathcal{U}, \quad (2.53)$$

la fonction $f : \mathcal{V} \mapsto \mathbb{R}$ est appelée **critère**, fonction **objectif** ou fonction **coût**; \mathcal{U} est l'espace des solutions admissibles ou encore espace de contrôle. Le problème de maximisation pouvant se ramener à celui de la minimisation par la multiplication de la fonction objectif par -1 , nous nous limiterons à la minimisation.

On parle d'optimisation sans contrainte si l'espace de contrôle est identique à l'espace de définition de la fonction coût ($\mathcal{U} = \mathcal{V}$); dans le cas contraire, on parle d'optimisation sous contraintes.

Notation Nous nous plaçons dans le cas où $\mathcal{V} = \mathbb{R}^n$ et nous notons :

- $c(\mathbf{x}) = (c_i(\mathbf{x}))_{1 \leq i \leq p}$: fonction de contraintes que nous définirons plus bas selon les cas avec $\mathbf{x} \in \mathcal{V}$;
- $H_f(\mathbf{x})$ le Hessien de la fonction f ;
- $A(\mathbf{x})$ le jacobien de $c(\mathbf{x})$.

Conditions d'optimalité

Les conditions d'optimalité résument les conditions d'existence et d'unicité de la solution d'un problème d'optimisation ainsi que les propriétés de cette solution. Nous allons donner dans ce paragraphe les principaux théorèmes qui définissent les conditions d'optimalité d'un problème d'optimisation. Pour les problèmes d'optimisation des chapitres suivants, nous nous restreindrons aux cas où ces conditions sont garanties. Les démonstrations de ces théorèmes peuvent se trouver dans les ouvrages classiques d'optimisation tels que [Gould and Leyffer(2003)].

Optimisation sans contrainte

Un point \mathbf{x}^* est un minimum local s'il existe un voisinage ouvert \mathcal{N} de \mathbf{x}^* tel que :

$$\forall \mathbf{x} \in \mathcal{N}, \mathbf{x} \neq \mathbf{x}^*, f(\mathbf{x}) \geq f(\mathbf{x}^*), \quad (2.54)$$

On parlera de minimum **isolé** si la condition d'infériorité est stricte $f(\mathbf{x}) > f(\mathbf{x}^*)$. La figure 2.1 illustre la notion de minimum local non isolé (weak local minimum en anglais) et de minimum isolé (strong local minimum en anglais).

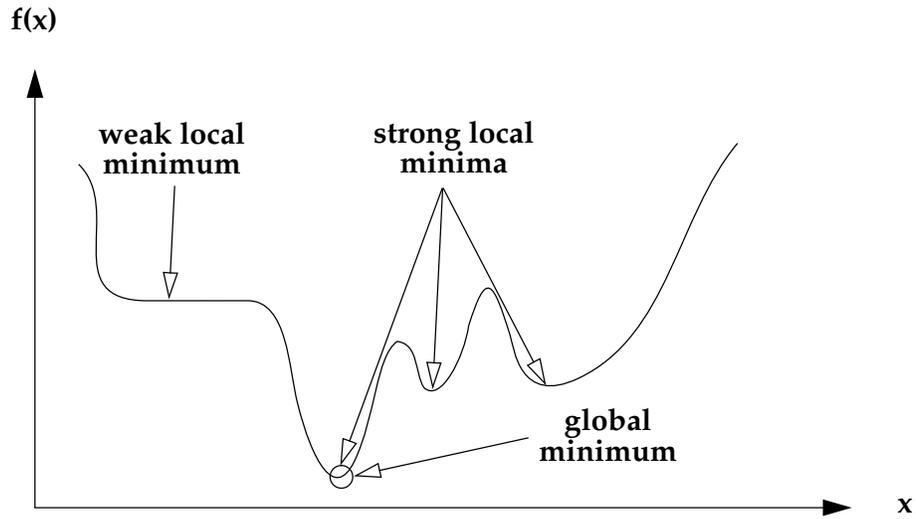


FIGURE 2.1 – Illustration des minima en minimisation sans contrainte; Source : [Gould and Leyffer(2003)].

Conditions d'existence et d'unicité

Théorème 2.1 *Si f est continue et $\lim_{\|\mathbf{x}\| \rightarrow +\infty} f(\mathbf{x}) = +\infty$, alors, f admet au moins un minimum.*

Théorème 2.2 *Si \mathcal{V} est convexe et f est strictement convexe, alors il existe au plus un minimum.*

Les conditions des théorèmes (2.1) et (2.2) garantissent l'existence et l'unicité de la solution du problème de minimisation.

Conditions d'optimalité

Théorème 2.3 *Si f est de classe C^1 et \mathbf{x}^* est un minimum local de f , alors $\nabla f(\mathbf{x}^*) = 0$.*

Théorème 2.4 *Si f est de classe C^2 et \mathbf{x}^* est un minimum local de f , alors $\nabla f(\mathbf{x}^*) = 0$ et $H_f(\mathbf{x}^*)$ est défini semi positif :*

$$\langle \mathbf{y}, H_f(\mathbf{x}^*)\mathbf{y} \rangle \geq 0, \forall \mathbf{x} \in \mathcal{V} \quad (2.55)$$

Théorème 2.5 *Si f est de classe C^2 et \mathbf{x}^* est un minimum local de f , alors \mathbf{x}^* est isolé si $H_f(\mathbf{x}^*)$ défini positif :*

$$\langle \mathbf{y}, H_f(\mathbf{x}^*)\mathbf{y} \rangle > 0, \forall \mathbf{x} \in \mathcal{V}. \quad (2.56)$$

Optimisation sous contraintes

Les conditions d'optimalité sous contraintes sont un peu plus compliquées. La figure 2.2 illustre les minima d'un problème d'optimisation sous contrainte.

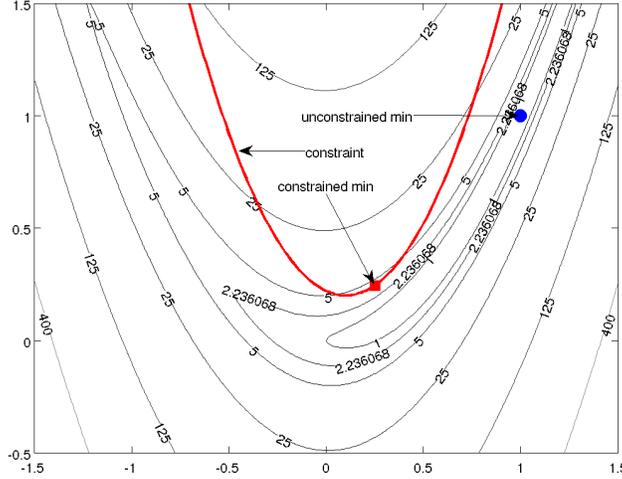


FIGURE 2.2 – Illustration graphique du minimum sous contraintes

On distingue les contraintes d'égalité et les contraintes d'inégalité. On parlera de contraintes d'égalité quand le sous espace des contraintes \mathcal{U} se définit comme :

$$\mathcal{U} = \{\mathbf{x} \in \mathcal{V} / c(\mathbf{x}) = 0\}. \quad (2.57)$$

Dans le cas où le sous espace de contrainte se définit comme :

$$\mathcal{U} = \{\mathbf{x} \in \mathcal{V} / c(\mathbf{x}) \geq 0\}, \quad (2.58)$$

on parlera de contraintes d'inégalité. Les contraintes $c(\mathbf{x}) \leq 0$ pouvant s'y ramener en multipliant par -1 . Dans les problèmes réels, on peut retrouver les deux types de contraintes.

Optimisation sous contraintes d'égalité

Théorème 2.6 *Si f et c sont de classe C^1 , \mathbf{x}^* un minimum local de f sous contraintes $c(\mathbf{x}) = 0$ et les $\nabla c_i(\mathbf{x})_{1 \leq i \leq p}$ sont linéairement indépendants, alors il existe $\mathbf{y} = (\lambda_i)_{1 \leq i \leq p} \in \mathbb{R}$ tel que :*

$$c(\mathbf{x}) = 0 \text{ et} \quad (2.59)$$

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^p \lambda_i \nabla c_i(\mathbf{x}^*) = 0. \quad (2.60)$$

Les coefficients λ_i sont appelés les **multiplicateurs de Lagrange**. Si on pose :

$$\mathcal{L}(\mathbf{x}, \mu_1, \mu_2, \dots, \mu_p) = \nabla f(\mathbf{x}^*) + \sum_{i=1}^p \mu_i \nabla c_i(\mathbf{x}^*), \quad (2.61)$$

alors optimiser f sous contraintes $c(\mathbf{x}) = 0$ revient à chercher le point-selle de \mathcal{L} encore appelé **Lagrangien**.

Théorème 2.7 Si f et c sont de classe C^2 , \mathbf{x}^* un minimum local de f sous contraintes $c(\mathbf{x}) = 0$ et les $\nabla c_i(\mathbf{x})_{1 \leq i \leq p}$ sont linéairement indépendants, alors il existe $\mathbf{y} = (\lambda_i)_{1 \leq i \leq p} \in \mathbb{R}$ tel que :

$$c(\mathbf{x}) = 0 \text{ et} \quad (2.62)$$

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^p \lambda_i \nabla c_i(\mathbf{x}^*) = 0, \quad (2.63)$$

où les λ_i sont les **multiplicateurs de Lagrange**. Si on pose :

$$\mathcal{L}(\mathbf{x}, \mu_1, \mu_2, \dots, \mu_p) = \nabla f(\mathbf{x}^*) + \sum_{i=1}^p \mu_i \nabla c_i(\mathbf{x}^*), \quad (2.64)$$

avec en plus :

$$\langle \mathbf{z}, H_{\mathcal{L}}(\mathbf{x}^*, \mathbf{y}) \mathbf{z} \rangle \geq 0, \forall \mathbf{z} \text{ satisfaisant } A(\mathbf{x}^*) \mathbf{z} = 0, \quad (2.65)$$

où A est le jacobien de c

alors optimiser f sous contraintes $c(\mathbf{x}) = 0$ revient à chercher le point-selle de la fonction \mathcal{L} appelée lagrangien. Le problème d'optimisation sous contrainte d'égalité est ainsi ramené à un problème d'optimisation sans contrainte.

2.4.3 Principe général des méthodes de descente pour la minimisation

Tout problème d'optimisation étant ramené à un problème d'optimisation sans contrainte, nous allons à présent regarder les méthodes de résolution de ce dernier type de problèmes. La présentation sera restreinte aux méthodes itératives de minimisation dites déterministes qui sont les plus adaptées aux problèmes de grande dimension tels que ceux rencontrés en assimilation de données.

Le principe général des méthodes que nous allons décrire dans cette section peut se résumer par l'algorithme suivant.

Algorithm 1: Principe général des algorithmes déterministes de minimisation

Entrées: approximation initiale \mathbf{x}_0 pour $k = 0$

répéter

 Calculer la direction de recherche \mathbf{d}_k ;

 Déterminer le pas de progression dans la direction de recherche $\alpha_k > 0$;

 Calculer l'approximation suivante $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$

jusqu'à convergence ;

Les critères de convergence sont définis selon le problème et les ressources temps et matériels disponibles. Il est courant de se baser sur le résidu ou la progression de l'approximation au cours des itérations. La direction de recherche est le vecteur indiquant la direction dans laquelle il faut poursuivre la recherche et α_k indique la distance à parcourir dans cette direction.

Direction de recherche

La valeur de la fonction objectif doit diminuer au fur et à mesure des itérations : $f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) < f(\mathbf{x}_k)$. On dit aussi que la direction de recherche est une direction de descente. De l'approximation tronquée de Taylor :

$$f(\mathbf{x} + \alpha \mathbf{d}) \approx f(\mathbf{x}) + \alpha (\nabla f(\mathbf{x}))^T \mathbf{d}, \quad (2.66)$$

on déduit que les directions suivant lesquelles la dérivée directionnelle est négative ($\langle \nabla f(\mathbf{x}), \mathbf{d} \rangle < 0$) sont de bons candidats. Les méthodes itératives déterministes se distinguent par le choix de la direction de descente. On les divise en trois grandes classes :

- les méthodes de relaxation qui parcourent toutes les directions de l'espace pour rechercher la bonne ; ces méthodes sont très simples à mettre en œuvre mais très coûteuses en calcul quand la dimension du problème est grand ;
- les méthodes du gradient (basée le gradient) ;
- les méthodes de Newton qui utilisent le Hessien.

L'annexe A.3 présente les deux dernière approches. Les algorithmes utilisés pour les travaux présentés dans ce mémoire sont basés sur les méthodes de Newton.

Pas de recherche

Quand la direction de recherche est trouvée, il faut déterminer le pas de progression encore appelé pas de descente. L'approche utilisée pour la détermination du pas de descente permet de classer les méthodes de minimisation en deux catégories :

Les méthodes à pas constant On choisit :

$$\alpha_k = \alpha, \forall k \quad (2.67)$$

Les méthodes à pas optimal l'objectif est de choisir le pas optimal suivant la direction de descente :

$$\alpha_k = \text{Arg} \min_{\alpha \in \mathbb{R}} f(\mathbf{x}_k + \alpha \mathbf{d}_k) \quad (2.68)$$

On distingue deux principales méthodes pour y arriver :

- recherche linéaire ;
- région de confiance.

Le principe détaillé de ces méthodes peut se trouver dans les ouvrages classiques d'optimisation, et particulièrement dans [Gould and Leyffer(2003)], [Dennis and Schnabel(1983)], [Luenberger(1984)], [Celis et al.(1985)] ou encore [Byrd et al.(1987)].

2.4.4 Préconditionnement pour les problèmes inverses

Préconditionnement des systèmes linéaires

En mathématiques appliquées, les méthodes les plus robustes pour la résolution d'une grande classe de problèmes sont les méthodes itératives. Elles constituent en général le seul choix lorsqu'il s'agit de problèmes non linéaires. C'est aussi un choix de prédilection pour les problèmes linéaires de grandes dimensions ; dans ces cas en effet, les méthodes directes bien que connues comme déterministes sont en général très coûteuses surtout en espace mémoire

et dans les cas extrêmes s'avèrent non praticables même avec les gros ordinateurs. Cependant, la convergence de ces méthodes itératives se dégrade rapidement lorsque le problème n'est pas bien conditionné. Ainsi, en algèbre linéaire, pour la résolution de systèmes linéaire de la forme :

$$\mathbf{Ax} = \mathbf{y}^o \quad (2.69)$$

il est courant de résoudre non pas le système d'origine mais un système modifié équivalent avec un meilleur conditionnement. On fait notamment usage du préconditionnement à gauche pour obtenir le système :

$$\mathbf{P}^{-1}\mathbf{Ax} = \mathbf{P}^{-1}\mathbf{y}^o, \quad (2.70)$$

ou du préconditionnement à droite :

$$\mathbf{AP}^{-1}(\mathbf{Px}) = \mathbf{y}^o. \quad (2.71)$$

L'objectif du préconditionnement en résolution numérique de système linéaire est d'obtenir une matrice dont les valeurs propres sont mieux réparties ; le conditionnement du problème modifié se trouve ainsi amélioré. Le préconditionnement idéal consisterait à choisir convenablement l'opérateur \mathbf{P} de telle sorte que la matrice du système préconditionné $\mathbf{P}^{-1}\mathbf{A}$ (pour le préconditionnement à gauche) ou \mathbf{AP}^{-1} (pour le préconditionnement à droite) ait un faible conditionnement et/ou des valeurs propres groupées autour de l'unité. Avec le conditionnement optimal égal à l'unité, une itération unique est requise pour converger. Cependant, l'obtention d'un tel conditionnement peut être aussi difficile que la résolution du système initial ; on recherche alors l'opérateur de préconditionnement qui permettra d'obtenir le meilleur compromis entre le coût de calcul et le taux de convergence. Pour les systèmes linéaires, l'opérateur de préconditionnement \mathbf{P} est en général obtenu par factorisation incomplète de la matrice \mathbf{A} du système. Des détails sur ces techniques peuvent se trouver dans le livre de Golub [Golub and Loan(1996)] ou encore celui de Saad [Saad(2003)]. Quand il s'agit des systèmes symétriques et définis positifs, le préconditionnement à droite est en général préféré ; dans ce cas, l'expression de l'opérateur de préconditionnement idéal par factorisation de Cholesky³ conduit au système préconditionné suivant :

$$\mathbf{AP}^{-\frac{T}{2}}\mathbf{z} = \mathbf{y}^o. \quad (2.72)$$

Avec le changement de variable $\mathbf{z} = \mathbf{P}^{\frac{T}{2}}\mathbf{x}$. D'un point de vue complexité de calcul, l'application de l'opérateur $\mathbf{P}^{-\frac{T}{2}}$ est plus avantageux que l'opérateur \mathbf{P}^{-1} grâce à sa structure triangulaire.

Préconditionnement pour les problèmes inverses généraux

Considérons le cas général où le problème inverse est défini par la minimisation de la fonction coût :

$$J(\delta\mathbf{v}) = \frac{1}{2}\|\mathcal{H}(\mathcal{M}(\mathbf{v}^b + \delta\mathbf{v})) - \mathbf{y}^o\|_{\mathcal{O}}^2 + \frac{1}{2}\alpha_b\|\delta\mathbf{v}\|_{\mathcal{V}}^2 \quad (2.73)$$

avec $\mathbf{v} \in \mathcal{V}$ le vecteur de paramètre (c'est la variable de contrôle en assimilation de données) à déterminer, \mathcal{V} l'espace de contrôle ou l'espace des paramètres, \mathbf{v}^b l'ébauche sur \mathbf{v} , $\delta\mathbf{v} = \mathbf{v} - \mathbf{v}^b$, \mathcal{M} le modèle qui lie la variable de contrôle à l'état \mathbf{x} du système, \mathbf{y}^o l'état observé (vecteur

3. Si \mathbf{A} est une matrice symétrique et définie positive, alors sa factorisation de Cholesky est définie par la matrice \mathbf{L} telle que $\mathbf{A} = \mathbf{LL}^T$

des observations), \mathcal{H} l'opérateur d'observation, $\|\cdot\|_{\mathcal{V}}$, $\|\cdot\|_{\mathcal{O}}$, la norme appropriée dans l'espace de contrôle, l'espace des observations respectivement. Le modèle \mathcal{M} est défini par :

$$\begin{aligned} \mathcal{M} : \mathcal{V} &\rightarrow \mathcal{X} \\ \mathbf{v} &\mapsto \mathbf{x} = \mathcal{M}(\mathbf{v}); \end{aligned} \quad (2.74)$$

et l'opérateur d'observation par :

$$\begin{aligned} \mathcal{H} : \mathcal{X} &\rightarrow \mathcal{O} \\ \mathbf{x} &\mapsto \mathbf{y} = \mathcal{H}(\mathbf{x}) \end{aligned} \quad (2.75)$$

On se ramène facilement au cas de l'assimilation variationnelle de données en réécrivant le modèle et l'opérateur d'observation sur toute la trajectoire. Nous allons nous intéresser uniquement au cas linéaire⁴, l'adjectif linéaire se rapportant aux opérateurs modèle \mathcal{M} et d'observations \mathcal{H} et veut dire qu'ils peuvent s'exprimer en terme de matrices \mathbf{M} et \mathbf{H} pour le problème discrétisé. Sous ces conditions, la fonction coût se réécrit :

$$J(\delta\mathbf{v}) = \frac{1}{2} \|HM(\mathbf{v}^b + \delta\mathbf{v}) - \mathbf{y}^o\|_{\mathbf{R}^{-1}}^2 + \frac{1}{2} \|\delta\mathbf{v}\|_{\mathbf{B}^{-1}}^2 \quad (2.76)$$

où les opérateurs \mathbf{B} et \mathbf{R} sont utilisés pour définir les normes appropriées dans l'espace de contrôle ($\|\cdot\|_{\mathbf{B}^{-1}} = \|\cdot\|_{\mathcal{V}}$) et dans l'espace d'observation ($\|\cdot\|_{\mathbf{R}^{-1}} = \|\cdot\|_{\mathcal{O}}$). Ces opérateurs sont en général définis par la matrice de covariance d'erreurs d'ébauche pour \mathbf{B} et par la matrice de covariance d'erreurs d'observation pour \mathbf{R} . Avec la formulation de l'équation (2.76), le gradient de la fonction coût s'exprime comme :

$$\nabla J(\delta\mathbf{v}) = \mathbf{B}^{-1}\delta\mathbf{v} + \mathbf{M}^T\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{M}(\mathbf{v}^b + \delta\mathbf{v}) - \mathbf{y}^o). \quad (2.77)$$

Le développement de Taylor nous permet d'écrire :

$$\nabla J(\delta\mathbf{v}) = \nabla J(0) + \mathbf{G}\delta\mathbf{v}, \quad (2.78)$$

où $\mathbf{G} = \mathbf{B}^{-1} + \mathbf{M}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{M}$ est la matrice hessienne de la fonction coût.

A l'optimum (\mathbf{v}^*), on a :

$$0 = \nabla J(0) + \mathbf{G}\delta\mathbf{v}^*, \text{ étant donné que } \nabla J(\delta\mathbf{v}^*) = 0, \quad (2.79)$$

et $\delta\mathbf{v}^*$ s'écrit :

$$\delta\mathbf{v}^* = -\mathbf{G}^{-1}\nabla J(0). \quad (2.80)$$

Dans le cas linéaire, la résolution d'un problème inverse peut se ramener à la résolution d'un système linéaire avec comme matrice, l'opérateur hessien. Comme nous l'avons dit plus haut, la convergence des méthodes itératives pour la résolution de ce type de système est fortement dépendante du conditionnement de la matrice du système; d'où la nécessité de résoudre un système équivalent préconditionné. Le préconditionnement à droite est plus approprié dans le cas où la matrice \mathbf{G} est symétrique et définie positive.

4. Dans la pratique, il est courant de se ramener à ce cas par linéarisation des opérateurs non linéaires

Simplification du préconditionnement en présence de modèles complexes

Selon la dimension du problème à résoudre, il peut être difficile et même impossible de manipuler la matrice hessienne \mathbf{G} ; c'est notamment le cas en assimilation de données avec des modèles qui peuvent être très complexes et la dimension du problème très grande. Dans ces cas, l'opérateur $\mathbf{M}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{M}$ peut être très difficile à estimer ou à manipuler (voir la thèse de Vidard [Vidard(2001)] pour plus de détails). Il est courant de définir une approximation de l'opérateur \mathbf{G} par $\tilde{\mathbf{G}} \approx \mathbf{B}^{-1}$. Cette simplification permet de définir un opérateur de préconditionnement qui conduit à un système avec des valeurs propres augmentées de 1 et un conditionnement substantiellement réduit [Courtier(1997)].

Schéma simplifié de construction et d'utilisation des opérateurs de covariance d'erreurs

La description faite dans ce paragraphe permettra de mieux comprendre l'utilisation de l'opérateur de diffusion généralisée pour le préconditionnement. Cet opérateur sera introduit au chapitre (5). Dans certaines applications, notamment dans la prévision des écoulements géophysiques, la matrice de covariance d'erreur d'ébauche peut être découpée en sous matrices conduisant à une matrice bloc de taille $n \times n$. La décomposition en blocs peut être basée sur les propriétés physiques du système, mais aussi peut correspondre uniquement à une décomposition logique ou les deux. La décomposition basée sur les propriétés physiques correspond en général à une décomposition du domaine physique en sous-domaines complémentaires [Courtier et al.(1998)] selon les propriétés du modèle dans chaque sous-domaine. Un exemple courant est la séparation correspondant aux différences de comportement des modèles météorologiques à différentes latitudes [Derber and Bouttier(1999)]. La décomposition logique quant à elle correspond en général à une séparation du traitement des variables indépendantes (la référence est faite ici aux variables du modèle continu et l'indépendance correspond à un point donné du repère temps).

Illustration sur un modèle d'écoulement en eaux peu profondes

Le système d'équations décrivant les écoulements en eaux peu profondes (ou shallow water en anglais) aussi connu comme le modèle de Saint-Venant, donne l'évolution sous la surface(libre ou non) d'un fluide d'épaisseur négligeable par rapport à la largeur. Les variables du modèle dans ce cas sont la vitesse et la pression (qui peut être réduite à la hauteur de la colonne du fluide sous la surface). Dans un modèle de Saint-Venant en deux dimensions, la variable vitesse est constituée de deux composantes, on obtient ainsi un modèle à trois variables, les deux composantes de la vitesse (\mathbf{u} et \mathbf{v}) et la hauteur du fluide (\mathbf{h}). Une décomposition logique permet d'écrire la matrice de covariance d'erreur comme une matrice 3×3 constituée de sous matrices exprimée sous la forme :

$$\mathbf{B} = \begin{pmatrix} \mathbf{B}_{uu} & \mathbf{B}_{uv} & \mathbf{B}_{uh} \\ \mathbf{B}_{vu} & \mathbf{B}_{vv} & \mathbf{B}_{vh} \\ \mathbf{B}_{hu} & \mathbf{B}_{hv} & \mathbf{B}_{hh} \end{pmatrix}. \quad (2.81)$$

La matrice en elle même n'est pas d'une grande importance, seule son application à un vecteur est requise dans une application ($\mathbf{x} \mapsto \mathbf{B}\mathbf{x}$) ; pour les blocs de la diagonale, Weaver et Courtier [Weaver and Courtier(2001)] proposent l'utilisation d'un opérateur de diffusion dont une approximation peut être donnée par une convolution gaussienne. Une autre approche consiste

à utiliser les filtres numériques [Purser et al.(2003)]. La partie diagonale que nous notons $\hat{\mathbf{B}}$ et qui constitue la partie univariée⁵ peut s'écrire :

$$\hat{\mathbf{B}} = (\boldsymbol{\Sigma})diag(\mathbf{C}_i)(\boldsymbol{\Sigma}), \quad (2.82)$$

où $\boldsymbol{\Sigma}$ représente la matrice diagonale des écarts types (erreurs), $\mathbf{C}_i = \boldsymbol{\Lambda}_i \mathbf{L}_i \boldsymbol{\Lambda}_i$, \mathbf{L}_i étant l'opérateur de diffusion de Weaver et Courtier ; $\boldsymbol{\Lambda}_i$ un opérateur de normalisation permettant d'obtenir des valeurs unitaires sur la diagonale des matrices \mathbf{C}_i . De façon simplifiée, les relations croisées qui définissent la partie non diagonale ou multivariée (par opposition à univariée) de la matrice de covariance \mathbf{B} s'obtiennent par application à $\hat{\mathbf{B}}$ d'opérateurs qui définissent l'équilibre entre les différentes variables :

$$\mathbf{B} = \mathbf{K} \hat{\mathbf{B}} \mathbf{K}^T. \quad (2.83)$$

L'opérateur \mathbf{K} représente un tel opérateur d'équilibre. Dans un modèle de Saint-Venant, il peut être remplacé par l'équilibre géostrophique. Cette construction facilite la factorisation par la méthode de Cholesky de la matrice de covariance d'erreur ; suivant le schéma de construction que nous venons de présenter, cette factorisation est donnée par :

$$\mathbf{B}^{\frac{1}{2}} = \mathbf{K}(\boldsymbol{\Sigma})\mathbf{C}^{\frac{1}{2}} \text{ avec } \mathbf{C}^{\frac{1}{2}} = \boldsymbol{\Lambda}\mathbf{L}^{\frac{1}{2}} \quad (2.84)$$

Rôle des opérateurs \mathbf{L} et \mathbf{K} L'opérateur \mathbf{L} comme opérateur de diffusion, permet de faire un lissage local. Il permet ainsi d'atténuer l'importance des incohérences locales ; il a un effet régularisant. L'opérateur \mathbf{K} permet de transférer l'information d'une variable à une autre (cas de la décomposition logique) ou bien d'une région du domaine physique à une autre (cas de la décomposition physique) ; il permet de palier au problème de manque d'informations.

2.5 Calcul du gradient : le modèle adjoint

Pour faire usage des méthodes d'optimisation décrites dans la section (2.4) dans la résolution du problème posé à la section (2.51), il est nécessaire de disposer d'un moyen de calcul du gradient de la fonction coût ∇J en \mathbf{u} . L'idée du calcul explicite est à éliminer lorsqu'il s'agit de problèmes réels d'assimilation de données en météorologie ou océanographie. En effet, l'état du modèle ne peut pas être explicité sauf dans quelques cas particuliers avec beaucoup de simplifications. L'objectif est donc de calculer le gradient numérique. La solution la plus évidente est d'utiliser l'approximation par différences finies pour calculer la dérivée de gâteau dans toutes les directions. Ceci requiert autant d'exécution du modèle que de direction (dimension de la variable de contrôle). Pour illustration, le temps d'exécution du modèle global du centre européen, résolution *T799* est de 6 minutes et la dimension de sa variable d'état est de 10^8 . Si on procède par différences finies pour le calcul du gradient du gradient de la fonction coût avec ce modèle, le temps de calcul sera de 6×10^8 minutes, soit un peu plus de 19 années de calcul avec un calculateur équivalent. Une méthode ingénieuse pour y arriver est celle de l'adjoint. L'usage de l'adjoint en assimilation de données a été introduit par [Le Dimet(1982)] et [Le Dimet and Talagrand(1986)]. C'est une technique basée sur l'approche de Lions de résolution du problème de contrôle optimal [Lions(1971)]

5. qui décrit les statistiques d'erreurs sur une seule variable

2.5.1 Adjoint continu

Soit $\tilde{\mathbf{u}} \in \mathcal{X}$, une perturbation sur \mathbf{u} .

Modèle linéaire tangent

La dérivée directionnelle du modèle direct (1.1) dans la direction $\tilde{\mathbf{u}}$ donne le modèle linéaire tangent suivant :

$$\begin{cases} \frac{d\hat{\mathbf{x}}}{dt} &= \left[\frac{\partial \mathcal{M}}{\partial \mathbf{x}} \right] \hat{\mathbf{x}} \\ \hat{\mathbf{x}}(0) &= \tilde{\mathbf{u}} \end{cases} \quad (2.85)$$

où $\left[\frac{\partial \mathcal{M}}{\partial \mathbf{x}} \right]$ est le Jacobien de \mathcal{M} par rapport à la variable \mathbf{x} et $(\hat{\cdot})$ dénote la dérivée directionnelle.

La dérivée directionnelle de la fonction coût dans la direction $\tilde{\mathbf{u}}$ est donnée par :

$$\hat{J}(\mathbf{u}, \tilde{\mathbf{u}}) = \left\langle \mathbf{u} - \mathbf{u}^b, \mathbf{B}^{-1} \tilde{\mathbf{u}} \right\rangle + \int_0^T \left\langle \mathcal{H}(\mathbf{x}) - \mathbf{y}^o, \mathbf{R}^{-1} \left[\frac{\partial \mathcal{H}}{\partial \mathbf{x}} \right] \hat{\mathbf{x}} \right\rangle dt. \quad (2.86)$$

Si la fonction J est Frechet-différentiable au point \mathbf{u} , alors $\hat{J}(\mathbf{u}, \tilde{\mathbf{u}}) = \langle \nabla J(\mathbf{u}), \tilde{\mathbf{u}} \rangle$. Ainsi, la mise en évidence d'une relation de dépendance linéaire de $\hat{J}(\mathbf{u}, \tilde{\mathbf{u}})$ par rapport à $\tilde{\mathbf{u}}$ permet d'obtenir le gradient de J . Pour y arriver, on introduit la variable adjointe.

Modèle adjoint

Soit \mathbf{p} (qui sera défini plus loin) la variable adjointe de même dimension que \mathbf{x} , la linéarité de $\hat{J}(\mathbf{u}, \tilde{\mathbf{u}})$ par rapport à $\tilde{\mathbf{u}}$ peut être mise en évidence en faisant le produit scalaire convenablement choisi de la première équation du modèle linéaire tangent (2.85) par \mathbf{p} , puis en intégrant le résultat sur l'intervalle $[0, T]$:

$$\int_0^T \left\langle \frac{d\hat{\mathbf{x}}}{dt}, \mathbf{p} \right\rangle dt = \int_0^T \left\langle \left[\frac{\partial \mathcal{M}}{\partial \mathbf{x}} \right] \hat{\mathbf{x}}, \mathbf{p} \right\rangle dt. \quad (2.87)$$

L'intégration par partie du premier membre nous donne :

$$\langle \hat{\mathbf{x}}(T), \mathbf{p}(T) \rangle - \langle \hat{\mathbf{x}}(0), \mathbf{p}(0) \rangle - \int_0^T \left\langle \hat{\mathbf{x}}, \frac{d\mathbf{p}}{dt} \right\rangle dt = \int_0^T \left\langle \left[\frac{\partial \mathcal{M}}{\partial \mathbf{x}} \right] \hat{\mathbf{x}}, \mathbf{p} \right\rangle dt. \quad (2.88)$$

La transposition du second membre et le regroupement des termes en $\hat{\mathbf{x}}$ nous conduit à :

$$\langle \hat{\mathbf{x}}(T), \mathbf{p}(T) \rangle - \langle \hat{\mathbf{x}}(0), \mathbf{p}(0) \rangle = \int_0^T \left\langle \frac{d\mathbf{p}}{dt} + \left[\frac{\partial \mathcal{M}}{\partial \mathbf{x}} \right]^T \mathbf{p}, \hat{\mathbf{x}} \right\rangle dt. \quad (2.89)$$

Si l'on défini \mathbf{p} comme solution du système (2.90) encore appelé **modèle adjoint** :

$$\begin{cases} \frac{d\mathbf{p}}{dt} + \left[\frac{\partial \mathcal{M}}{\partial \mathbf{x}} \right]^T \mathbf{p} = \left[\frac{\partial \mathcal{H}}{\partial \mathbf{x}} \right]^T \mathbf{R}^{-T} (\mathcal{H}(\mathbf{x}) - \mathbf{y}^o), t \in [0, T], \\ \mathbf{p}(T) = 0, \end{cases} \quad (2.90)$$

où $\left[\frac{\partial \mathcal{H}}{\partial \mathbf{x}}\right]^T$ est l'adjoint de l'opérateur d'observation linéarisé, alors, on obtient :

$$\begin{aligned} -\langle \hat{\mathbf{x}}(0), \mathbf{p}(0) \rangle &= \int_0^T \left\langle \left[\frac{\partial \mathcal{H}}{\partial \mathbf{x}}\right]^T \mathbf{R}^{-T} (\mathcal{H}(\mathbf{x}) - \mathbf{y}^o), \hat{\mathbf{x}} \right\rangle dt \\ &= \hat{J}(\mathbf{u}, \tilde{\mathbf{u}}) - \langle \mathbf{B}^{-T}(\mathbf{u} - \mathbf{u}^b), \tilde{\mathbf{u}} \rangle. \end{aligned} \quad (2.91)$$

En remplaçant $\hat{\mathbf{x}}(0)$ par sa valeur $\tilde{\mathbf{u}}$, la dérivée directionnelle de la fonction coût devient :

$$\begin{aligned} \hat{J}(\mathbf{u}, \tilde{\mathbf{u}}) &= -\langle \tilde{\mathbf{u}}, \mathbf{p}(0) \rangle + \langle \mathbf{B}^{-T}(\mathbf{u} - \mathbf{u}^b), \tilde{\mathbf{u}} \rangle \\ &= \langle \tilde{\mathbf{u}}, -\mathbf{p}(0) + \mathbf{B}^{-T}(\mathbf{u} - \mathbf{u}^b) \rangle. \end{aligned} \quad (2.92)$$

On en déduit le gradient de la fonction coût :

$$\nabla J = -\mathbf{p}(0) + \mathbf{B}^{-T}(\mathbf{u} - \mathbf{u}^b) \quad (2.93)$$

Le contrôle optimal \mathbf{u}^* satisfait la relation :

$$\nabla J(\mathbf{u}^*) = -\mathbf{p}(0) + \mathbf{B}^{-T}(\mathbf{u}^* - \mathbf{u}^b) = 0. \quad (2.94)$$

Pour obtenir le gradient de la fonction coût, il suffit ainsi de résoudre le modèle direct, puis le modèle adjoint de façon rétrograde.

2.5.2 Adjoint discret

Dans un problème concret, le modèle est discrétisé en temps et les observations ne sont pas continues en temps. Il est conseillé de faire l'adjoint du modèle discret et non de discrétiser le modèle adjoint continu. Dans la pratique, on est amené à dériver l'adjoint à partir du code informatique représentant le modèle direct, toutes les simplifications et approximations utilisées pour le code direct sont ainsi prises en compte dans le code adjoint et permet d'avoir des résultats conformes aux traitements. Nous allons reprendre ici le processus de dérivation de modèle adjoint avec un modèle discret (2.45) et la fonction coût discrète (2.49).

Soit $\tilde{\mathbf{u}} \in \mathcal{X}$, une perturbation sur \mathbf{u} .

Modèle linéaire tangent, version discrète

La dérivée de Gâteaux dans la direction $\tilde{\mathbf{u}}$ conduit au modèle linéaire tangent :

$$\begin{cases} \frac{\hat{\mathbf{x}}_{i+1} - \hat{\mathbf{x}}_i}{\Delta t} = \left[\frac{\partial \mathcal{M}_i}{\partial \mathbf{x}_i}\right] \hat{\mathbf{x}}_i, \\ \hat{\mathbf{x}}_0 = \tilde{\mathbf{u}}. \end{cases} \quad (2.95)$$

La dérivée directionnelle de la fonction coût dans la direction $\tilde{\mathbf{u}}$ est donnée par :

$$\hat{J}(\mathbf{u}, \tilde{\mathbf{u}}) = \langle \mathbf{u} - \mathbf{u}^b, \mathbf{B}^{-1} \tilde{\mathbf{u}} \rangle + \sum_0^N \left\langle \mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i^o, \mathbf{R}_i^{-1} \left[\frac{\partial \mathcal{H}_i}{\partial \mathbf{x}_i}\right] \hat{\mathbf{x}}_i \right\rangle \Delta t. \quad (2.96)$$

Modèle adjoint, version discrète

On fait le produit scalaire de la première équation du modèle linéaire tangent (2.95) par la variable adjointe \mathbf{p}_i (qui sera défini plus loin,) puis on fait l'intégrale discrète (méthode des rectangles) sur la fenêtre temporelle :

$$\sum_{i=0}^N \left\langle \frac{\hat{\mathbf{x}}_{i+1} - \hat{\mathbf{x}}_i}{\Delta t}, \mathbf{p}_i \right\rangle \Delta t = \sum_{i=0}^N \left\langle \left[\frac{\partial \mathcal{M}_i}{\partial \mathbf{x}_i} \right] \hat{\mathbf{x}}_i, \mathbf{p}_i \right\rangle \Delta t. \quad (2.97)$$

En utilisant la règle de dérivation du produit de fonction :

$$(fg)' = f'g + fg' \Leftrightarrow f'g = (fg)' - fg', \quad (2.98)$$

où f et g sont des fonctions de la variable temps et $(.)'$ la dérivée première par rapport au temps, ou son approximation discrète :

$$\frac{f_{i+1} - f_i}{\Delta t} g_i = \frac{f_{i+1} g_{i+1} - f_i g_i}{\Delta t} - f_i \frac{g_{i+1} - g_i}{\Delta t}, \quad (2.99)$$

qu'on peut obtenir par la formule de Taylor en négligeant les termes d'ordre supérieur, le premier membre de l'équation (2.97) devient :

$$\sum_{i=0}^N \frac{\langle \hat{\mathbf{x}}_{i+1}, \mathbf{p}_{i+1} \rangle}{\Delta t} \Delta t - \sum_{i=0}^N \frac{\langle \hat{\mathbf{x}}_i, \mathbf{p}_i \rangle}{\Delta t} \Delta t - \sum_{i=0}^N \left\langle \hat{\mathbf{x}}_i, \frac{\mathbf{p}_{i+1} - \mathbf{p}_i}{\Delta t} \right\rangle \Delta t. \quad (2.100)$$

Après simplification, on obtient :

$$\langle \hat{\mathbf{x}}_{N+1}, \mathbf{p}_{N+1} \rangle - \langle \hat{\mathbf{x}}_0, \mathbf{p}_0 \rangle - \sum_{i=0}^N \left\langle \hat{\mathbf{x}}_i, \frac{\mathbf{p}_{i+1} - \mathbf{p}_i}{\Delta t} \right\rangle \Delta t. \quad (2.101)$$

La substitution dans l'équation (2.97) nous ramène à :

$$\begin{aligned} \langle \hat{\mathbf{x}}_{N+1}, \mathbf{p}_{N+1} \rangle - \langle \hat{\mathbf{x}}_0, \mathbf{p}_0 \rangle &= \sum_{i=0}^N \left\langle \hat{\mathbf{x}}_i, \frac{\mathbf{p}_{i+1} - \mathbf{p}_i}{\Delta t} \right\rangle \Delta t + \sum_{i=0}^N \left\langle \left[\frac{\partial \mathcal{M}_i}{\partial \mathbf{x}_i} \right] \hat{\mathbf{x}}_i, \mathbf{p}_i \right\rangle \Delta t, \\ &= \sum_{i=0}^N \left\langle \hat{\mathbf{x}}_i, \frac{\mathbf{p}_{i+1} - \mathbf{p}_i}{\Delta t} + \left[\frac{\partial \mathcal{M}_i}{\partial \mathbf{x}_i} \right]^T \mathbf{p}_i \right\rangle \Delta t. \end{aligned} \quad (2.102)$$

De même que dans le cadre continu, on peut choisir \mathbf{p}_i comme solution de :

$$\begin{cases} \frac{\mathbf{p}_{i+1} - \mathbf{p}_i}{\Delta t} + \left[\frac{\partial \mathcal{M}_i}{\partial \mathbf{x}_i} \right]^T \mathbf{p}_i = \left[\frac{\partial \mathcal{H}_i}{\partial \mathbf{x}_i} \right]^T \mathbf{R}_i^{-T} (\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i^o), 0 \leq i \leq N \\ \mathbf{p}_{N+1} = 0 \end{cases} \quad (2.103)$$

qui est la version discrète du modèle adjoint.

L'équation (2.97) devient :

$$\begin{aligned} - \langle \hat{\mathbf{x}}_0, \mathbf{p}_0 \rangle &= \sum_{i=0}^N \left\langle \hat{\mathbf{x}}_i, \left[\frac{\partial \mathcal{H}_i}{\partial \mathbf{x}_i} \right]^T \mathbf{R}_i^{-T} (\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i^o) \right\rangle \Delta t \\ &= \hat{J}(\mathbf{u}, \tilde{\mathbf{u}}) - \left\langle \mathbf{u} - \mathbf{u}^b, \mathbf{B}^{-1} \tilde{\mathbf{u}} \right\rangle \end{aligned} \quad (2.104)$$

$$= \hat{J}(\mathbf{u}, \tilde{\mathbf{u}}) - \left\langle \mathbf{B}^{-T} (\mathbf{u} - \mathbf{u}^b), \tilde{\mathbf{u}} \right\rangle. \quad (2.105)$$

Sachant que $\hat{\mathbf{x}}_0 = \tilde{\mathbf{u}}$, on en déduit que :

$$\hat{J}(\mathbf{u}, \tilde{\mathbf{u}}) = \left\langle \tilde{\mathbf{u}}, -\mathbf{p}_0 + \mathbf{B}^{-T}(\mathbf{u} - \mathbf{u}^b) \right\rangle \quad (2.106)$$

qui exprime la linéarité de $\hat{J}(\mathbf{u}, \tilde{\mathbf{u}})$ par rapport à $\tilde{\mathbf{u}}$. On en déduit le gradient de la fonction coût :

$$\nabla J(\mathbf{u}) = -\mathbf{p}_0 + \mathbf{B}^{-T}(\mathbf{u} - \mathbf{u}^b). \quad (2.107)$$

Le contrôle optimal \mathbf{u}^* se déduit de la relation :

$$\nabla J = -\mathbf{p}_0 + \mathbf{B}^{-T}(\mathbf{u} - \mathbf{u}^b) = 0 \quad (2.108)$$

qu'on pouvait deviner à partir de l'équation (2.94)

Nous avons ainsi traité le problème d'assimilation variationnelle de données pour le contrôle de la condition initiale. La démarche reste la même si l'on veut contrôler d'autres paramètres du système notamment les conditions aux limites.

Dans la pratique, d'autres contraintes sont ajoutées à la fonction coût sous forme de termes de régularisation. Les méthodes variationnelles d'assimilation de données sont également appelées méthodes adjointes du fait de l'usage des opérateurs adjoints.

2.5.3 Le système d'optimalité

Pour obtenir le gradient de la fonction coût J par rapport au contrôle, on doit résoudre le modèle direct (1.1), ensuite résoudre le modèle adjoint (2.90) de façon rétrograde et appliquer la formule (2.93). Une fois le gradient obtenu, on l'utilise dans un processus de minimisation qui permet de déterminer les paramètres optimaux ; ce processus consiste globalement en la résolution du système d'optimalité suivant :

$$\left\{ \begin{array}{l} \frac{d\mathbf{x}}{dt} = \mathcal{M}(\mathbf{x}), \quad t \in [0, T], \\ \mathbf{x}(0) = \mathbf{u}, \\ \\ \frac{d\mathbf{p}}{dt} + \left[\frac{\partial \mathcal{M}}{\partial \mathbf{x}} \right]^T \mathbf{p} = \left[\frac{\partial \mathcal{H}}{\partial \mathbf{x}} \right]^T (\mathcal{H}(\mathbf{x}) - \mathbf{y}^o), \quad t \in [0, T], \\ \mathbf{p}(T) = 0, \\ \\ \nabla J = -\mathbf{p}(0) + \mathbf{B}^{-T}(\mathbf{u} - \mathbf{u}^b) = 0. \end{array} \right. \quad (2.109)$$

2.5.4 Utilité de l'adjoint

L'usage de l'adjoint est loin d'être limité à la détermination du gradient de la fonction coût. L'adjoint au second ordre [Wang et al.(1992), Le Dimet et al.(2002)] permet notamment de déterminer le Hessien et d'étudier la sensibilité aux observations. Le principe de dérivation reste le même sauf qu'il faut travailler sur le système d'optimalité et non sur le modèle direct. Rappelons que le Hessien est utilisé dans les méthodes de minimisation de type Newton. L'adjoint au second ordre offre un moyen de multiplication du Hessien par un vecteur. A défaut de disposer de la hessienne inverse, cet algorithme peut être utilisé dans les méthodes

itératives de résolution de systèmes linéaires (gradient conjugué, GMRES) pour calculer le produit de la hessienne inverse par un vecteur. Ainsi pour des problèmes de taille raisonnable on peut faire usage de la méthode de Newton sans simplification.

2.6 Pratique des méthodes basées sur l'adjoint

Dans la pratique, les codes informatiques sont les principaux outils qui permettent d'utiliser les modèles mathématiques. Le code informatique associé à chaque modèle doit être écrit afin de le rendre utilisable. Le but de cette section est de donner une idée de ce qui est fait dans la pratique pour les modèles adjoints. En théorie, il existe trois principales méthodes pour obtenir le code informatique associé au modèle adjoint :

- la discrétisation des équations du modèle adjoint suivi de l'écriture du code informatique associé ;
- l'écriture du code informatique associé au modèle adjoint discret dérivé de la discrétisation du modèle direct ;
- l'écriture du code informatique associé au modèle adjoint à partir du code informatique du modèle direct : c'est la différentiation algorithmique.

Pour les problèmes de grande dimension, seule la troisième méthode est conseillée. Elle présente plusieurs avantages. Le premier étant que le code adjoint ainsi produit calcule le gradient exact d'un point de vue arithmétique aux erreurs de précision de la machine près [Elizondo et al.(2002)]. Un autre avantage non négligeable est qu'un tel code peut être généré de façon automatique par des outils de différentiation automatique.

2.6.1 Principe de différentiation algorithmique

La différentiation algorithmique exploite le fait qu'un programme informatique calculant la fonction différentiable F définie par :

$$\begin{aligned} F &: \mathbb{R}^n \rightarrow \mathbb{R}^m, \\ \mathbf{x} &\mapsto \mathbf{y} = F(\mathbf{x}) \end{aligned} \quad (2.110)$$

peut être décomposée en une séquence d'appels de fonctions élémentaires. Chacune de ces fonctions élémentaires pouvant être différenciée de façon triviale. De telles fonctions mathématiques élémentaires sont : les additions, les multiplications, les fonctions exponentielles, les fonctions trigonométriques (sin, cos, tan), etc. En suivant une telle décomposition, la fonction F peut se réécrire :

$$F = F^K \circ \dots \circ F^1 := \bigcirc_{l=1}^K F^l. \quad (2.111)$$

Les fonctions F^l , ($l = 1, \dots, K$) sont les fonctions élémentaires qui forment la fonction composite F et sont définies par :

$$\begin{aligned} F^l &: \mathbb{R}^{n_{l-1}} \rightarrow \mathbb{R}^{n_l}, \\ \mathbf{z}^{l-1} &\mapsto \mathbf{z}^l = F^l(\mathbf{z}^{l-1}). \end{aligned} \quad (2.112)$$

La sortie \mathbf{z}^{l-1} de la fonction élémentaire F^{l-1} constitue l'entrée de la fonction élémentaire F^l avec $\mathbf{z}^0 = \mathbf{x}$ et $\mathbf{z}^K = \mathbf{y}$. Dans cette représentation, les composantes du vecteur intermédiaire \mathbf{z}^l ne correspondent pas exactement aux variables du programme informatique. En effet, les

variables du programme peuvent voir leur valeur changer d'une étape à une autre du calcul de F alors que le vecteur \mathbf{z}^l lui contient les n_l résultats intermédiaires valides après l'étape l du programme.

En appliquant la règle de différentiation de la composée de fonctions à la formulation séquentielle de la fonction F (voir équation 2.111), on peut calculer la dérivée (gradient) de chaque fonction mathématique différentiable et calculable par un programme informatique. L'application de la règle de différentiation des fonctions composées permet en effet d'écrire la matrice jacobienne de la fonction F comme :

$$A(\mathbf{x}_0) = \frac{\partial F^K}{\partial \mathbf{z}^{K-1}} \Big|_{\mathbf{z}^{K-1} = \bigcirc_{i=1}^{K-1} F^i(\mathbf{x}_0)} \cdots \frac{\partial F^1}{\partial \mathbf{z}^0} \Big|_{\mathbf{z}^0 = \mathbf{x}_0}. \quad (2.113)$$

Rappelons que la matrice jacobienne d'une fonction vectorielle est la matrice de ses dérivées partielles du premier ordre. Pour la fonction F de l'équation (2.111), elle est définie par :

$$A_{i,j}(\mathbf{x}_0) = \frac{\partial F_i}{\partial z_j} \Big|_{\mathbf{x} = \mathbf{x}_0} \quad (i = 1, \dots, m, j = 1, \dots, n). \quad (2.114)$$

La règle définie par l'équation (2.113) permet ainsi de calculer toutes les dérivées partielles du premier ordre d'une fonction vectorielle différentiable et calculable par un programme informatique. Le produit matriciel étant associatif, la stratégie de calcul de la matrice jacobienne définie par (2.113) dépend des valeurs relatives de m et n . On distingue deux principaux modes :

- le mode direct ou les multiplications (équation 2.113) sont effectuées dans le même ordre que la composition (équation 2.111). C'est à dire qu'on calcule dans un premier temps $\frac{\partial F^2}{\partial \mathbf{z}^1} \cdot \frac{\partial F^1}{\partial \mathbf{z}^0}$, puis on multiplie $\frac{\partial F^3}{\partial \mathbf{z}^2}$ par le résultat et ainsi de suite ;
- le mode rétrograde ou les multiplications (équation 2.113) sont effectuées dans l'ordre inverse de la composition. Dans ce mode, on commence par $\frac{\partial F^K}{\partial \mathbf{z}^{K-1}} \cdot \frac{\partial F^{K-1}}{\partial \mathbf{z}^{K-2}}$.

Dans le premier cas, les résultats intermédiaires ont chacun n colonnes et dans le second cas, ils ont chacun m lignes. Si la matrice jacobienne est pleine, alors d'un point de vue volume de calcul, le mode direct est plus approprié si $n < m$ tandis que le mode rétrograde est plus approprié si $n > m$. La fonction coût en assimilation variationnelle de données correspond en général au deuxième cas avec $m = 1$. Si la règle définie par l'équation (2.113) est appliquée de façon stricte, alors la différentiation d'un algorithme peut être faite de façon automatique. C'est le principe des outils de différentiation automatique. Pour plus de détails sur le principe de la différentiation algorithmique, les travaux de Giering et Kaminski [Giering and Kaminski(1998)] constituent une bonne référence.

2.6.2 Outils de différentiation automatique

Les outils de différentiation automatique existent depuis les années 1990. Depuis lors, les performances se sont accrues et les possibilités offertes se sont multipliées. Parmi ces outils, on note principalement la série ADI de Bischof et ses collaborateurs dont ADIFOR [Bischof et al.(1992), Bischof et al.(1996)] pour le fortran77, ADIC [Bischof et al.(1997)] pour le C ANSI et ADIMat [Bischof et al.(2003)] pour Matlab. La différentiation algorithmique constitue un domaine de recherche très actif (voir le site internet www.autodiff.com). Les logiciels de différentiation automatique sont toujours plus perfectionnés et s'adaptent aux

nouveaux langages de programmation qui offre un confort dans l'écriture des programmes. C'est le cas par exemple de Tapenade [Hascoët and Pascual(2004)] qui prend en compte le Fortran 95.

Chapitre 3

Modèles et données : cas de l'atmosphère

La qualité de la prévision numérique dépend en grande partie de l'état initial ayant servi à la produire. Les techniques d'assimilation de données abordées dans le chapitre précédent permettent de déterminer l'état initial optimal encore appelé analyse qui est utilisé pour la prévision. Les observations de l'état du système constituent un des principaux ingrédients de l'assimilation de données. Dans le cas des prévisions météorologiques, le système est défini de l'atmosphère dans sa globalité.

Le but de ce chapitre est de faire une brève présentation des principales variables caractéristiques de l'état de l'atmosphère ainsi que des principaux types de mesures qui sont effectuées pour la prévision météorologique. L'annexe (C) présente une descriptions plus détaillés de ces concepts. Nous nous attarderons particulièrement sur les observations de type image produites par les satellites météorologiques.

3.1 Grandeurs physiques caractéristiques du système

Les principales variables d'un modèle mathématique de l'atmosphère sont le vent (vitesse et direction), la température, l'humidité et la pression. Ce sont les grandeurs physiques nécessaires à la description mathématique de l'état de l'atmosphère. Leur choix est guidé par les lois de la physique qui régissent le comportement du système.

3.2 Observations conventionnelles

En assimilation de données pour la météorologie, une observation désigne l'estimation quantitative d'une grandeur météorologique mesurable ou repérable telle que le vent, la pression atmosphérique, la température, etc. Cette opération fournit pour un instant (ou un intervalle de temps), et pour un lieu (ou un espace) déterminé, un nombre en unité connue (ou sans dimension, éventuellement) exprimant la valeur de la grandeur physique mesurée ou repérée. Selon l'étendu du volume de l'atmosphère prise en compte dans l'échantillonnage et la position relative de l'instrument de mesure, on distingue les observations directes (mesures *in situ*) et les observations distantes (télémesures).

3.2.1 Mesures *in situ*

On désigne par mesures *in situ* les observations réalisées par des instruments géographiquement situés dans la région où la mesure est effectuée. Ces instruments permettent d'échantillonner un volume d'air très petit (comparé à celui des phénomènes météorologiques décrits). Les observations conventionnelles sont en général des mesures directes des variables du modèle et peuvent donc être utilisées telles quelles dans un système d'assimilation de données (après éventuellement une interpolation). La présence de l'instrument dans la région échantillonnée permet d'obtenir des observations bien localisées avec une grande précision.

Moyens de mesure

Les mesures *in situ* s'effectuent aussi bien par les instruments conventionnels situés en surface que par des sondes traversant toute l'épaisseur de l'atmosphère. Les observations de surface au niveau des continents sont réalisées dans les stations des services météorologiques de chaque pays. En mer, les systèmes de mesure sont installés sur des bouées (dérivantes ou fixes) et sur certains navires commerciaux. En altitude, les mesures sont effectuées par radiosondage ou par des avions de ligne. Des ballons-sondes transportant des capteurs électroniques traversent l'atmosphère en effectuant des mesures des grandeurs météorologiques à une fréquence donnée. Le suivi de la trajectoire du ballon permet de déterminer la vitesse et la direction du vent. C'est le moyen le plus courant de mesure du vent en altitude. Ces données sont transmises périodiquement par radio vers les stations au sol qui les traitent en temps réel. Certains avions de ligne sont aussi équipés d'instruments météorologiques qui leur permettent de faire des mesures tout au long des vols. Toutes ces mesures doivent répondre aux normes de l'Organisation Météorologique Mondiale (OMM) en termes d'heures d'observation, de qualité et de mise en forme afin d'être diffusées sur le réseau international pour être utilisées par les centres de prévision et les météorologistes.

Réseaux d'observation

Les postes d'observation utilisés pour la mesure des paramètres de l'atmosphère sont organisés en réseaux. Chaque réseau étant conçu pour une application ou un usage particulier. La densité des observations dans l'espace et dans le temps définit la résolution du réseau. Cette information est d'une importance capitale. Elle permet d'estimer l'erreur de représentativité affectant les usages postérieurs des mesures. En effet, lorsque l'étendu géographique d'un phénomène est très petite par rapport à l'écart entre deux stations de mesure, le phénomène peut ne pas être détecté (s'il est localisé dans une région à mi-chemin entre les deux stations) ou être éventuellement surestimé (si son étendu géographique inclut une des stations). **L'erreur de représentativité** qualifie l'estimation des erreurs que l'on commet en moyenne du fait de la mauvaise représentation des échelles plus fines que la résolution du réseau de mesure. Les problèmes posés par les erreurs de représentativité sont loin d'être résolus. La solution immédiate qui consiste à construire un réseau plus dense se heurte aux coûts exorbitants d'installation et de maintenance des réseaux, surtout en mer et dans les déserts. Il est nécessaire de faire un compromis entre les coûts et la densité des réseaux tout en évitant la redondance de l'information.

Exemple : le réseau synoptique

Ainsi désigne-t-on le réseau mondial soutenu par l'OMM. Il permet de décrire assez fidèlement

les phénomènes météorologiques à l'échelle synoptique.

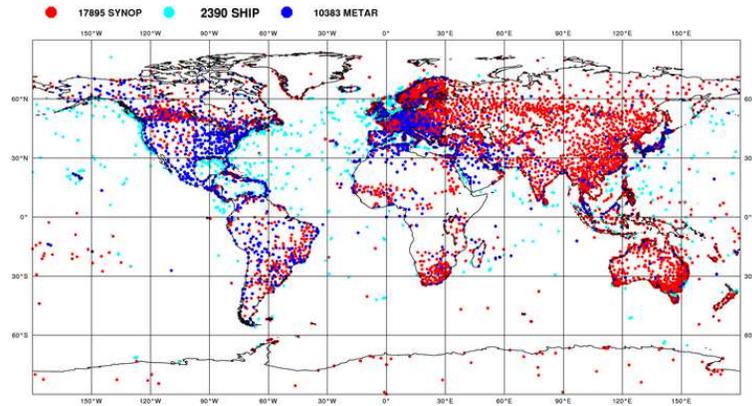


FIGURE 3.1 – Observations du réseau synoptique et bateau de ligne. Répartition/couverture des données utilisées au centre européen pour les prévisions à moyen terme (données du 24 février 2010 à 00 UTC). Le nombre total d'observations est de 30668 Source : <http://www.ecmwf.int/>.

La figure (3.1) montre la carte des mesures opérationnelles du réseau synoptique et des observations par bateau le 24 février 2010 à 00 TU. Cette carte montre l'inégale répartition des mesures sur le globe terrestre. La couverture est très faible dans de nombreuses régions à l'intérieur des tropiques, les zones désertiques, les océans et les régions polaires. Dans ces régions, d'autres mesures doivent être utilisées pour obtenir des prévisions de qualité équivalente à celle des régions les mieux couvertes du globe où ces observations décrivent correctement les phénomènes d'échelle synoptique voir sous-synoptique [Malardel(2005)]. Même dans les régions qui semblent totalement couverte sur la carte, il y a des disparités sous-régionales que l'on ne peut pas observer à cette échelle. Le nombre total d'observations, 30668, comparé à la surface couverte nous en dit plus. Ce nombre correspond à peine à une observation sur une superficie de $600km^2$ si on regroupe toutes ces données sur un territoire de la taille de la Russie. Ces observations du réseau synoptiques sont complétées par d'autres observations telles que ceux des réseaux de bouées et les réseaux de mesure en altitude qui sont présentés en annexe (C).

Remarque 3.2.1 *Les observations du réseau synoptique (mesures conventionnelles par les stations météorologiques terrestres) sont utilisées par les centres de prévision opérationnels. Ce sont notamment des observations de pression et d'humidité. De même, les observations de pression et de vent des bateaux et des bouées marines sont utilisées par les centres de prévision opérationnels.*

3.2.2 Télémessures

Les télémessures désignent les observations indirectes réalisées par des instruments situés à une certaine distance du lieu de mesure. Les moyens de mesure dans cette catégorie incluent : les radars, les satellites à orbite polaire¹ et les satellites géostationnaires². Les ob-

1. Les satellites à orbite polaire sont situés sur des orbite basses (400 à 800 kilomètres). A cette altitude, la région couverte à un instant donné est assez réduite mais la précision est plus forte (voir l'annexe C pour une description plus détaillée).

2. Les satellites géostationnaires sont situés sur une orbite à environ 36000 kilomètres de la Terre. A cette altitude, la région couverte est très grande mais la précision est plus faible (voir l'annexe C pour une description

servations dans ce cas sont des mesures indirectes des variables d'état des modèles. Les radiances constituent un exemple de mesures indirectes dans le cas des satellites. De ce fait, des post-traitements peuvent être nécessaires avant utilisation dans un système classique d'assimilation de données. La télé-observation permet d'échantillonner des régions géographiques considérables par rapport aux instruments conventionnels. Les mesures dans ce cas peuvent être actives ou passives. On parlera de mesure passive lorsque l'instrument mesure les radiations émises par la Terre ou l'atmosphère. La plupart des observations satellitaires rentre dans cette classe de mesure. On retrouve dans cette catégorie les satellites à orbite polaire NOAA 15, 16 et 17, et les satellites géostationnaires européens Meteosat et américains GOES. Quand le ciel est dégagé, les radiances fournissent des informations sur la température et l'humidité. Lorsque la couverture nuageuse est importante, le taux d'erreur devient prohibitif et les informations de température et d'humidité ne sont plus utilisées. Dans une telle situation, les radiances sont utilisées pour dériver les champs de vecteurs atmosphériques (en abrégé AMV pour Atmospheric Motion Vector en anglais) utilisés comme pseudo observations de vent dans les couches supérieures de l'atmosphère. La figure (3.2) montre les pseudo-observations (en date du 24 février 2010) du champ de vitesse (AMV) dérivé des mesures de vapeurs d'eau par la constellation de satellites géostationnaires opérationnels.

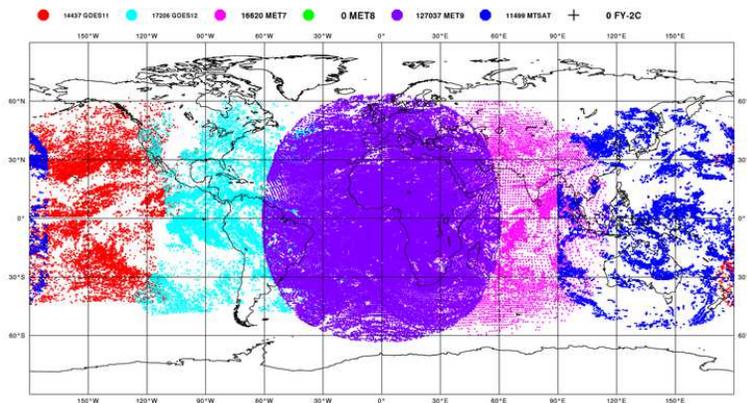


FIGURE 3.2 – Pseudo-Observations de champs de vitesse extraits des images vapeur d'eau des satellites géostationnaires. Répartition des données AMV utilisées au centre européen pour les prévisions à moyen terme (données du 24 février 2010 à 00 UTC). Le nombre total d'observations est de 186799. Source : <http://www.ecmwf.int/>.

Les variables des modèles atmosphériques peuvent être également mesurées par des technologies qui émettent des radiations vers la Terre ou l'atmosphère et mesurent la réflexion des radiations émises : on parle alors de mesures actives. Cette technologie est utilisée dans les radars. C'est le cas des scatteromètres qui fournissent les champs de vent dérivés de l'estimation de l'ondulation de surface des océans sur le signal réfléchi. On retrouve cette technologie à bord du satellite QuickScat pour la mesure du vent à la surface de la mer. Ces signaux peuvent être affectés par la présence de la pluie en fonction de la longueur d'onde du signal. Les télémessures sont en général moins précises que les mesures *in situ*, mais ont une couverture géographique importante dans le cas des satellites.

3.2.3 Ingrédients pour l'assimilation des télémessures

Les télémessures telles que décrites dans les paragraphes précédents, sont des observations indirectes des variables caractéristiques des modèles atmosphériques. Ce sont des mesures des radiations émises ou réfléchies par l'atmosphère. Contrairement aux mesures conventionnelles, ces mesures sont en général une intégration sur la couche de l'atmosphère traversée par la radiation. Elles cumulent les non linéarités de l'intégration le long de cette couche aux incertitudes liées à l'état de l'atmosphère entre le lieu sondé et l'instrument de mesure. Elles fournissent cependant des informations qui permettent de mesurer la présence ou la variation (gradient) d'un élément particulier dans l'atmosphère. Ces mesures permettent d'une certaine manière de quantifier les variables d'état des modèles atmosphériques. Elles peuvent donc être considérées comme des observations. L'utilisation des télémessures dans les schémas variationnels d'assimilation de données nécessite leur prise en compte dans la fonction coût, notamment dans le terme d'écart aux observations qui s'exprime par :

$$J_o = (\mathcal{H}(\mathbf{x}) - \mathbf{y}^o)^T R^{-1} (\mathcal{H}(\mathbf{x}) - \mathbf{y}^o). \quad (3.1)$$

Pour rappel, l'opérateur de covariance d'erreur d'observation \mathbf{R} combine l'erreur de mesure (liée aux instruments) et l'erreur de représentativité. L'opérateur d'observation \mathcal{H} permet de calculer l'équivalent observé de la variable d'état du système de l'espace d'état dans l'espace des observations. Deux alternatives d'utilisation des télémessures s'offrent :

- extraction de l'équivalent variables d'état des télémessures et leur utilisation comme des observations conventionnelles. On parle alors de pseudo-observations ;
- assimilation directe des télémessures.

Dans l'un ou l'autre cas, la connaissance approfondie du phénomène observé et du processus d'acquisition sont incontournables. Nous allons considérer dans cette partie que le terme d'écart aux observations de la fonction coût est limité aux télémessures. Nous noterons par \mathcal{H} l'opérateur modélisant le phénomène observé et le processus d'acquisition. Nous l'avons volontairement confondu à l'opérateur d'observation (noté par le même symbole) pour des raisons qui vont se préciser par la suite. Nous ferons des précisions partout où il pourrait avoir ambiguïté. Comme nous l'avons déjà mentionné, les télémessures diffèrent des mesures conventionnelles entre autres par le fait qu'elles sont des mesures indirectes des variables d'état du système. Soit \mathbf{x} la variable ou l'ensemble des variables d'état nécessaires à la modélisation du phénomène observé et du processus d'acquisition, \mathbf{y}^o l'observation, alors l'équivalent observé de l'état du modèle numérique s'exprime par :

$$\mathbf{y} = \mathcal{H}(\mathbf{x}, \mathbf{z}), \quad (3.2)$$

où \mathbf{z} représente d'autres paramètres dont dépendent les observations. L'opérateur \mathcal{H} modélise le phénomène observé et le processus d'observation, il décrit les relations physiques entre \mathbf{x} , \mathbf{y} et \mathbf{z} . Par exemple, si on s'intéresse aux radiances, \mathbf{x} représenterait la pression à la surface, la température à la surface, les profils de température et d'humidité. \mathcal{H} représenterait le modèle de transfert radiatif et \mathbf{z} les paramètres tels que l'émissivité de la surface (dépendant de type de sol, de l'humidité du sol, etc.), les profils de dioxyde de carbone (CO₂), de méthane et d'ozone, etc. A propos de transfert radiatif et des radiations atmosphériques, il existe des ouvrages parmi lesquels [Andrews et al.(1987), Goody and Yung(1995)] qui en donnent les fondements. Les articles [Strow et al.(2003), Strow et al.(2006), Hannon et al.(1996)] détaillent la méthodologie du transfert radiatif dans les sondeurs de type AIRS. Les phénomènes

observés par les instruments de télémessures sont en général le résultat de processus très complexes et/ou faisant intervenir plusieurs variables du modèle : ils sont fortement non-linéaires. C'est le cas de la microphysique atmosphérique à l'origine des nuages, qui en plus des non-linéarités est caractérisée par un ensemble de processus à seuil. Ce sont les principales difficultés caractéristiques des modèles de description (opérateur \mathcal{H}) des phénomènes observés et du processus d'observation par télémessure. Des problèmes similaires sont rencontrés avec le transfert radiatif.

Télémessures comme pseudo observations

Des télémessures on peut envisager l'extraction des variables d'état influençant le phénomène observé et/ou le processus d'acquisition. De façon simple, l'extraction se fait par inversion de l'opérateur d'observation. C'est une tâche difficile en présence d'opérateurs non linéaires et presque impossible en présence de processus à seuil. Dans la pratique, cette extraction se fait soit par des approximations statistiques, soit par résolution de problèmes inverses (assimilation de données sur le modèle d'observation simplifié) en utilisant l'ébauche sur les différents champs. L'illustration de ce principe peut se trouver dans [Storto and Tveter(2009)] qui présente une approche d'utilisation des observations binaires de couverture nuageuse par satellite dans le schéma variationnel 3D du modèle ALADIN de météo France par approche pseudo observation. Une extraction alternative consiste à considérer l'évolution de la télémessure dans l'espace et dans le temps. On peut ainsi estimer les paramètres/variables qui n'influencent pas (ou presque) l'opérateur \mathcal{H} . Cette approche permet d'estimer les sources/puits de concentration de certaines quantités ou encore les paramètres de la dynamique de la mesure. Cette technique est exploitée en estimation du mouvement que nous présentons en annexe (D). Elle permet de déterminer le champ de vitesse à partir d'une séquence d'images. Une application est le calcul des champs de vecteur atmosphérique. La figures (3.2) montre un exemple de répartition des données AMV. Dans le même registre, [Maksyutov et al.(2003)] présente la modélisation inverse pour l'estimation des sources de dioxyde de carbone en surface à partir des observations satellitaires. Il montre dans quelles conditions ces mesures devraient être utilisées pour des résultats optimaux.

Remarque On voudrait que l'utilisation des pseudo-observations en assimilation de données ait le même effet que l'utilisation des observations conventionnelles. Cependant l'usage des télémessures comme pseudo-observations peut être sujet à deux problèmes. D'une part, l'inversion nécessite des modèles d'observations linéaires (ce qui n'est pas le cas dans la réalité), ou alors leur linéarisation. D'autre part, les pseudo-observations sont en général construites à partir de l'ébauche de l'état du système comme nous l'avons mentionné précédemment. L'erreur associée aux pseudo-observations est donc corrélée à l'erreur d'ébauche, ce qui viole une des hypothèses de base de l'assimilation variationnelle de données. Si les pseudo-observations sont donc utilisées conjointement aux observations conventionnelles, elles peuvent contribuer à donner un poids relatif plus élevé et non contrôlé à l'ébauche. Pour éviter ce problème, on peut utiliser les deux sources d'information séparément l'une avant l'autre, le dernier permettant d'améliorer le résultat obtenu par le premier. Logiquement, l'usage des pseudo-observations ne devrait être envisagé que si le modèle d'observation est linéaire mais avec beaucoup de précaution. Dans la plupart des cas, l'alternative pseudo-observation n'est pas optimale.

Utilisation directe de télémessures

L'utilisation directe des télémessures dans la fonction coût requiert un opérateur d'observation approprié. L'opérateur d'observation étant l'opérateur qui permet de projeter la variable d'état du système dans l'espace des observations, il est équivalent dans ce cas au modèle d'observation que nous avons défini à l'équation (3.2), d'où son appellation \mathcal{H} .

L'assimilation directe des télémessures repose sur une bonne caractérisation de l'opérateur d'observation $\mathcal{H}(\mathbf{x}, \mathbf{z})$. Si $\mathcal{H}(\mathbf{x}, \mathbf{z})$ n'est pas une bonne approximation de la réalité, soit parce que l'opérateur est inconnu ou parce que l'approximation dans \mathbf{z} (paramètres de l'opérateur) est très grossière, alors les observations ne peuvent pas être utilisées de façon robuste. Ces situations conduisent en général à des erreurs corrélées. S'il arrive des situations où les mesures ne peuvent pas être utilisées avec robustesse, elles doivent être éliminées. C'est ce que doit effectuer le contrôle de qualité sur les observations. Par exemple, en cas de couverture nuageuse, les données de radiances sont peu utilisables telles quelles.

3.3 Observation de type image

Les images constituent une source importante d'observation de phénomènes physiques. Les techniques d'acquisition d'images sont nombreuses et varient selon la nature de l'information recherchée, la nature et l'étendu du milieu concerné. On peut citer :

- la photographie pour le grand public ;
- l'imagerie médicale pour la médecine dont l'IRM (imagerie par résonance magnétique, la radiographie par rayon X, etc.) ;
- l'imagerie par RADAR décrite à la section (3.2.2) ;
- l'imagerie par satellite décrite à la section (3.2.2).

Les instruments d'imagerie permettent en général de mesurer les radiations émises ou réfléchies par un milieu. Ces mesures qui constituent l'image permettent de caractériser la nature du milieu en question.

Dans cette section, nous nous intéresseront particulièrement à la description des observations par satellites géostationnaires pour lesquels les mesures peuvent être considérées comme des photographies. Leur position fixe par rapport aux régions visées permet d'avoir des séquences d'images (vidéo) qui peuvent être exploitées en complément aux utilisations actuelles et aux autres sources d'information. Toutes les descriptions techniques des systèmes d'acquisition seront basées sur les satellites européens METEOSAT de deuxième génération.

3.3.1 Images brutes

Les instruments d'acquisition d'images transforment le signal reçu en signal électrique qui est ensuite numérisé pour former l'image brute (raw image en anglais). Ces instruments sont dotés d'une matrice de capteurs mesurant la quantité de rayonnement reçu. Une image brute est une matrice de pixels³, la valeur de chaque pixel quantifiant le rayonnement reçu par le capteur pendant le processus d'acquisition. Les images brutes peuvent donc être considérées comme des mesures de radiances (canaux infra rouges des satellites) ou de la lumière réfléchie (canaux visible des satellites, photographie traditionnelle). La surface physique couverte par un pixel représente la résolution spatiale de l'instrument de mesure.

3. pixel, abréviation de picture element, qui signifie «élément d'image» est l'unité de surface permettant de mesurer une image numérique.

Les images se caractérisent par leur profondeur : c'est le nombre de valeurs possibles que peut prendre un pixel. Elle est en général définie en nombre de bits. La profondeur définit également le nombre de valeurs distinctes que le capteur peut mesurer. Cette profondeur est de 10 bits pour les images brutes de niveau 1.0 des satellites METEOSAT de seconde génération. Il faut noter que les nombres flottants en simple précision sont représentés sur 32 bits (64 pour la précision double), ce qui représente la précision des calculs. Si un pixel est considéré comme une mesure, alors la précision de sa représentation est plus faible que celle des sorties des modèles numériques qui correspondent à la précision des ordinateurs. Cette information doit être prise en compte dans l'utilisation des images en assimilation de données.

D'un point de vue mathématique, une image en deux dimensions est couramment considérée comme une fonction scalaire de deux variables. L'image numérique représente ainsi la discrétisation de la fonction mathématique associée.

3.3.2 Présentation des images satellitaires

Depuis la fin des années 70, une surveillance globale de la couverture fluide de la Terre est assurée par une constellation de satellites géostationnaires au dessus de l'équateur, et un ensemble de satellites en orbite quasi-polaire. Chaque satellite météorologique géostationnaire fournit des images dites plein-disque couvrant pratiquement un quart de la surface terrestre. La qualité et la quantité de ces images en font une donnée importante pour le futur des modèles numériques de prévision. Contrairement aux autres systèmes d'observation, les observations satellitaires couvrent tout le globe terrestre comme le montre la figure (3.3). Ces observations ont une résolution inégalée par les modèles actuels de prévision globale ainsi que les réseaux d'observation conventionnelle. La table (3.1) présente les caractéristiques des images METEOSAT de deuxième génération : ces images ont une résolution au point sous satellite de $1km$ dans le canal visible haute résolution (High Resolution Visible en anglais) en abrégé, HRV et $3km$ dans les canaux Infrarouge (IR) et Vapeur d'eau (WV). Ces images sont produites à la fréquence d'une image chaque quart d'heure. A titre d'exemple, une image du satellite METEOSAT 9 couvre une région de $12500km \times 12500km$.

TABLE 3.1 – Résolutions horizontales et dimensions des images METEOSAT

Canal d'acquisition	résolution au pixel sous satellite (km^2)	taille en pixels de l'image plein disque
Visible haute résolution	1×1	11136×5568
Infrarouge	3×3	3712×3712

La couverture globale de la Terre par les images satellitaires en fait une source précieuse d'observations complémentaires aux observations conventionnelles. Leur résolution spatiale et temporelle permet de résoudre certains problèmes liés à la faible résolution des grilles de calcul et à la limitation des observations conventionnelles. La figure (3.4) montre une image plein-disque du canal vapeur d'eau du satellite METEOSAT et un zoom sur la région Europe de l'Ouest.

3.3.3 Principe d'acquisition des images satellite

Les satellites européens METEOSAT de seconde génération, MSG (Meteosat Second Generation) sont équipés du radiomètre SEVERI (Spinning Enhanced Visible and Infrared Imager).

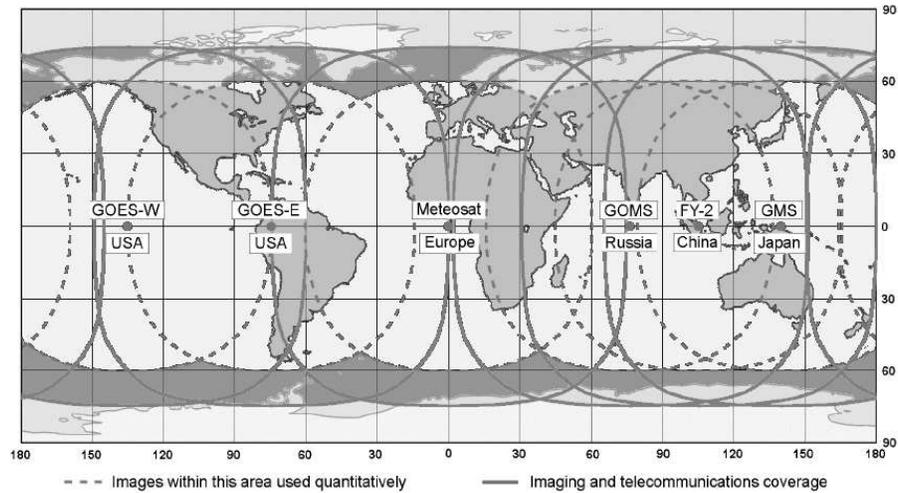


FIGURE 3.3 – Couverture nominale par les satellites météorologiques géostationnaires opérationnels.
Source : <http://medhycos.mpl.ird.fr/en/data/tec/dcp/Meteosat/TD05-part3.pdf>

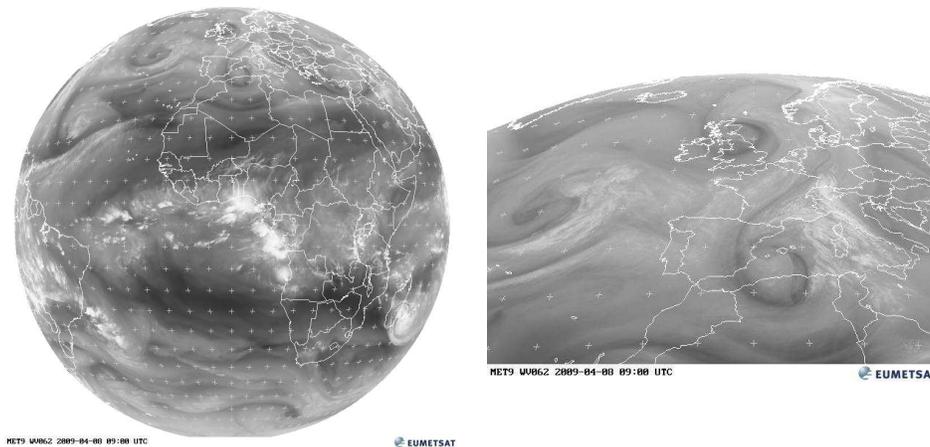


FIGURE 3.4 – Images METEOSAT du canal vapeur d'eau : à gauche image plein-disque ; à droite zoom sur la région Europe de l'Ouest.

C'est un radiomètre de 50cm de diamètre d'ouverture, à balayage ligne par ligne doté d'une capacité de sondage dans douze canaux différents : quatre canaux dans le spectre du visible et 8 canaux dans le spectre infrarouge. Le radiomètre SEVERI produit une image plein-disque toutes les quinze minutes avec une résolution au pixel sous-satellite d'un kilomètre pour le canal visible haute résolution, HRV et trois kilomètres pour les autres canaux.

Une image complète est obtenue par un balayage bidimensionnel (voir la figure 3.5) de la surface observée en combinant la rotation du satellite sur lui même et celle du système optique du radiomètre :

- le balayage très rapide d'une ligne (Est-Ouest) est obtenu par la rotation (à 100 tours par minutes) du satellite autour de son axe qui est perpendiculaire au plan orbital et orienté suivant l'axe Sud-Nord de la Terre ;
- le balayage de toutes les lignes (Sud-Nord) est rendu possible par la rotation du système optique avec un pas constant de 125.8 micro radian.

Le radiomètre SEVERI effectue un balayage entier de la surface observée en 12.5 min, le temps d'étalonnage, positionnement et stabilisation du système optique complète le cycle à 15 minutes. Les détails techniques du radiomètre SEVERI sont présentés dans [Aminou et al.(1997)]

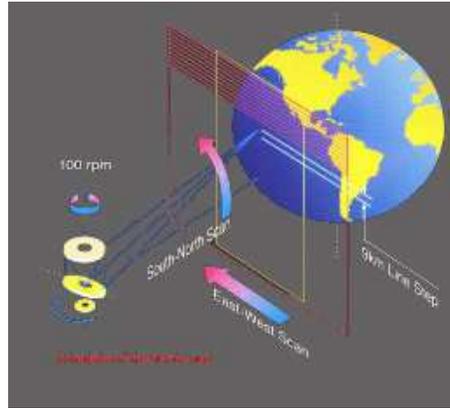


FIGURE 3.5 – Principe de l'imagerie par le radiomètre SEVERI équipant les satellites européens METEOSAT de seconde génération.

3.3.4 Couverture géographique et résolution des images par satellite

Couverture terrestre et taille des images

Le plein-disque terrestre dans la zone observée par le satellite MSG est entièrement couvert par tous les canaux sauf le canal HRV. Pour ce dernier, seule la moitié dans la direction Est-Ouest de l'image est fournie, (voir la figure 3.6).

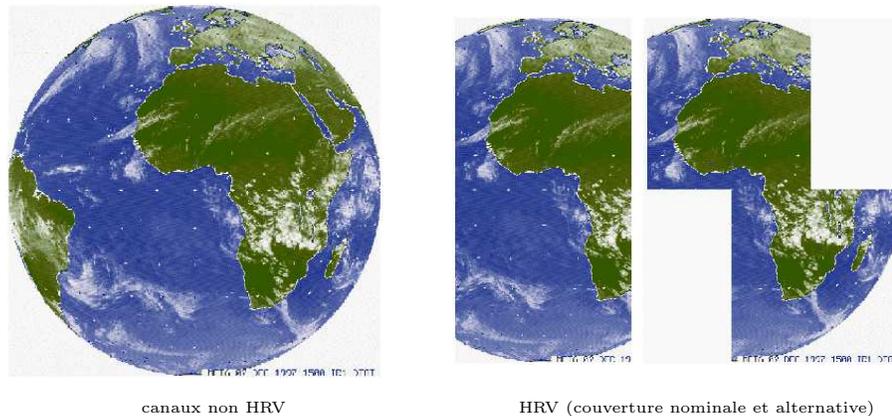


FIGURE 3.6 – Couverture nominale de la Terre par les images MSG.

Le système optique du radiomètre SEVERI est équipé de 42 capteurs dont 9 pour le canal HRV et 3 dans chacun des autres canaux, ceci permet au radiomètre de balayer 9 lignes HRV et 3 lignes de données dans les autres canaux en un tour du satellite. Ce système permet d'acquérir en un cycle des images dites de niveau 1.0 (Level 1.0) de taille 3834 pixels dans la direction Est-Ouest par 3745 dans la direction Sud-Nord pour les canaux non HRV ; 5751 pixels dans la direction Est-Ouest par 11235 dans la direction Sud-Nord pour le canal HRV.

La valeur des pixels d'une image de niveau 1.0 est celle acquise directement par le radiomètre. Une image de niveau 1.0 est donc affectée par toutes les perturbations dues au processus d'acquisition, notamment :

- les défauts géométriques induits par les incertitudes d'orbite et d'altitude ;
- les non-linéarités des mesures par radiomètre ;
- les variations entre les détecteurs d'un même canal.

Le domaine d'une image de niveau 1.0 est défini comme le carré circonscrivant le plein-disque terrestre observé par le satellite exception faite du canal HRV où seule la moitié (Est-Ouest) est fournie. Un pixel d'une image de niveau 1.0 est représenté sur 10 bit indépendamment du canal d'acquisition.

Les images de niveau 1.0 sont traitées en temps réel par la station au sol pour obtenir des images de niveau 1.5. Ce prétraitement permet de corriger les défauts susmentionnés des images 1.0. Pour tous les canaux sauf le HRV, les images de niveau 1.5 sont de taille 3712 lignes par 3712 colonnes (N-S \times E-O) avec une distance d'échantillonnage de $3km \times 3km$ au point sous satellite. Pour le canal HRV, ce sont des images de taille 11136 lignes par 5568 colonnes (N-S \times E-O) et une distance d'échantillonnage de $1km \times 1km$ au point sous satellite. Les dimensions rectangulaires dans le canal HRV résultent du fait que le radiomètre fournit uniquement la moitié dans la direction Est-Ouest de l'image, voir figure (3.6). Une description détaillée des caractéristiques des images MSG peut être trouvée dans le cahier des charges de l'utilisateur final des données MSG [EUMETSAT(2004)].

Résolution au sol

La figure (3.7) présente la carte de la résolution au sol des images MSG de niveau 1.5, exception faite du canal HRV. De l'intérieur vers l'extérieur, les dégradés de gris correspondent à $3.1km$, $4km$, $5km$, $6km$, $8km$ et $11km$ respectivement. La dépendance des coordonnées géographiques s'expliquent par le pas angulaire fixe de la rotation du système optique.

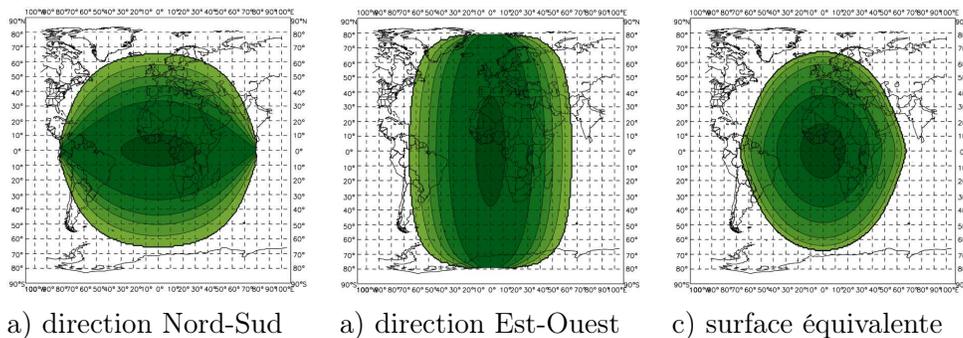


FIGURE 3.7 – Carte de la résolution au sol des images MSG 1.5.

3.3.5 Images et vision par ordinateur

La vision par ordinateur et le traitement d'images sont des techniques permettant d'une part de simuler la vision humaine et d'autre part de la faciliter en améliorant la qualité des images. Les éléments de base identifiables dans les images sont les formes dont la reconnaissance est un aspect fondamental en vision. Chaque forme étant caractérisée par la région

qu'elle circonscrit et les contours définissant ses limites (frontières). Les contours ou frontières des objets constituent des éléments essentiels en vision. Une des activités principale en traitement d'images et plus particulièrement en vision par ordinateur consiste à caractériser les formes dans les images. Cette opération appelée segmentation a pour but de rassembler les pixels entre eux suivant des critères prédéfinis. Le résultat de la segmentation est un ensemble de régions formant une partition de l'image ou un ensemble de contours délimitant les différentes régions de l'image. L'objectif de la segmentation est de simplifier ou de changer la représentation d'une image dans le but de faciliter son analyse. Il peut s'agir par exemple de séparer les objets de l'arrière plan. Les domaines d'applications sont nombreux, on peut citer entre autres :

- la médecine (localisation des tumeurs, mesure du volume des tissus, etc.);
- la cartographie (localisation des routes et forêts sur des images satellites, etc.);
- la météorologie (identification des nuages et des fronts sur des images météorologiques par satellite).

Il existe plusieurs méthodes de segmentation parmi lesquelles : le seuillage dont une description peut être trouvée dans [Gonzalez and Woods(2001)] ou [Sezgin and Sankur(2004)]; la détection des crêtes qui est décrite par plusieurs travaux dont [Ziou and Tabbone(1998)], [Torre and Poggio(1986)], [Williams and Shah(1993)] ou encore [Marr and Hildreth(1980)]. Une version évoluée de la détection des crêtes constitue la détection des contours.

Remarque 3.3.1 *Les crêtes d'intensité et les contours sont d'une importance capitale en traitement d'images et en vision par ordinateur dès que l'on s'intéresse au mouvement entre les images d'une séquence. Dans ce cas, l'information pertinente reste localisée au niveau des crêtes d'intensité ou encore des contours : c'est le problème d'ouverture⁴ bien connu en traitement d'images. Les travaux de Marr [Marr and Ullman(1981)] constituent une bonne référence pour la description de ce problème.*

Seuillage

Le seuillage est une opération de traitement du signal qui permet d'éliminer une partie du signal, plus précisément celle qui n'est pas intéressante pour l'application considérée. On distingue plusieurs variantes selon les applications et les propriétés du signal à traiter.

Binarisation

C'est une méthode de seuillage élémentaire qui permet de séparer les composantes d'un signal en deux classes. En traitement d'images, elle consiste à séparer l'arrière de l'avant plan ou encore le fond de la forme. L'opération de binarisation met à la valeur minimale tous les pixels correspondants à l'arrière plan (partie non significative du signal) de l'image et à la valeur maximale les autres pixels (partie significative) ou inversement. Ce qui revient à considérer une plage d'intensité comme l'arrière plan et l'autre comme l'avant plan. Le problème qui se pose en général est celui de la définition de la limite entre les deux groupes : le **seuil**. La méthode la plus élémentaire consiste à utiliser l'**histogramme de représentativité** des niveaux de gris dans l'image [Otsu(1979)].

Au lieu de se restreindre à deux plages de valeurs, on peut en choisir plusieurs, ce qui définit une stratification. Si au lieu de définir une valeur unique pour la partie significative, on décide

4. Seule le mouvement suivant la normale au contour peut être déterminé avec précision

plutôt de garder leur valeur d'origine, l'opération définit alors le seuillage. Ainsi définit, le seuillage peut servir à supprimer le bruit d'un signal corrompu.

Seuillage brut

L'objectif du seuillage est d'éliminer les coefficients *peu utiles* et de ne garder que les *coefficients utiles*. Les coefficients inférieurs au seuil sont mis à zéro tandis que les autres gardent leur valeurs : c'est le seuillage brut (ou hard thresholding). Si $\sigma > 0$ est la valeur du seuil, alors le seuillage brut se définit comme suit

$$\tau(x) = \begin{cases} x & \text{if } |x| \geq \sigma, \\ 0 & \text{if } |x| < \sigma, \end{cases} \quad (3.3)$$

C'est la définition la plus simple du seuillage. Le seuillage brut est marqué par une discontinuité aux valeurs $-\sigma$ et σ

Seuillage doux

On retrouve par opposition au seuillage brut le seuillage doux (ou soft thresholding). Ici, on met toujours à zéro les coefficients inférieurs à un seuil σ . Par contre, pour ceux supérieurs à σ , on atténue l'amplitude des coefficients par la valeur du seuil afin de s'assurer d'avoir enlevé l'effet du bruit même pour les forts coefficients. La fonction de seuillage se définit alors comme suit :

$$\tau(x) = \begin{cases} x - \text{sign}(x)\sigma & \text{if } |x| \geq \sigma, \\ 0 & \text{if } |x| < \sigma, \end{cases} \quad (3.4)$$

en supposant que la fonction *sign* retourne le signe de son argument avec la convention suivante : -1 pour un argument négatif, 0 pour un argument nul et $+1$ pour un argument positif. On obtient une fonction de seuillage σ continue mais l'amplitude du signal est atténuée.

Seuillage adaptatif

Les opérations de seuillage que nous venons de définir sont des opérations de seuillage global : le signal est traité de la même façon dans toute les régions de son domaine de définition. Ainsi définit, on ne peut obtenir des résultats de bonne qualité que si le signal est assez uniforme sur son domaine de définition. Notamment, les images doivent être bien contrastées et l'arrière plan uniforme. Pour pallier le défaut d'uniformité du signal sur tout le domaine de définition, on peut décomposer le domaine du signal en sous-domaines et associer un seuil spécifique à chaque sous-domaine : c'est le seuillage adaptatif [Chan et al.(1998), Liu et al.(2002)]. On doit définir la bonne taille pour les sous-domaines de sorte que le signal soit uniformément réparti dans chacun d'eux. Avec des sous-domaines trop grands, on se ramène au problème du seuillage global alors que des blocs trop petits conduisent à un temps de calcul prohibitif.

Remarque 3.3.2 *Dans l'un ou l'autre cas, on est confronté à la définition du seuil. Une condition nécessaire pour une détermination déterministe du seuil est que le signal soit séparable. Dans le cas du traitement d'images, cette condition se traduit par un histogramme bimodal avec deux modes bien séparés. Un mode correspondant à l'arrière plan et l'autre à l'avant plan. Le rehaussement de contraste est parfois nécessaire pour y arriver. Une méthode de choix du seuil basée sur le principe du maximum d'entropie est proposée par [Wong and Sahoo(1989), Brink(1994), Tao et al.(2003)]. [Bazi et al.(2007)] propose un seuillage basé sur l'algorithme d'expectation et de minimisation.*

Détection des crêtes

Une crête (crête d'intensité ou crête de luminosité) est un ensemble de points voisins qui délimitent deux régions voisines de l'image. Elle se matérialise dans l'image par un changement plus ou moins brusque de la fonction de luminosité et correspond souvent aux limites géométriques des objets constituant la scène représentée. Etant définies par la régularité de la fonction image, les crêtes peuvent être caractérisées par les dérivées de cette fonction. Les opérateurs de crêtes sont regroupés en deux principales catégories selon l'ordre des dérivées utilisées :

- les opérateurs du premier ordre basés sur les dérivées premières. Dans cette catégorie, on retrouve l'opérateur de Robert, l'opérateur de Sobel [Sobel and Feldman(1968)] et l'algorithme de Canny [Canny(1986), Deriche(1987)] ;
- les opérateurs du second ordre basés sur les dérivées secondes. On retrouve dans cette catégorie l'opérateur dit Zero-Crossing [Hummel and Gidas(1984), Anh et al.(1996)] et l'algorithme de Marr-Hildreth [Marr and Hildreth(1980)].

Les crêtes obtenues par les opérateurs ci-dessus sont le plus souvent discontinues, incomplètes, ou encore, de nombreux crêtes parasites peuvent être détectées, provenant du bruit, ou des textures macroscopiques. Différentes méthodes ont été proposées pour y remédier parmi lesquelles la méthode des contours actifs.

Détection des contours : les contours actifs

Un contour actif est une courbe définie dans le domaine de l'image et se déformant pour minimiser une fonction d'énergie dont le minimum est défini sur les contours de l'image [Kass et al.(1988)]. On distingue deux formulations :

- la formulation paramétrique [Kass et al.(1988), Cohen(1991), Xu and Prince(1998a)], qui définit le contour actif comme une courbe paramétrique minimisant une fonctionnelle d'énergie ;
- la formulation géométrique [Caselles et al.(1993), Malladi et al.(1994), Caselles(1995)], qui est basée sur la théorie de l'évolution des courbes [Malladi et al.(1994)] et des ensembles de niveau [Osher and Sethian(1988), Osher and Fedkiw(2003)]. La courbe définissant le contour actif évolue en utilisant uniquement des mesures géométriques.

3.3.6 Utilisation actuelle des images en assimilation de données : Pseudo-observations

Les images satellites étant des télémessures, la description faite au paragraphe (3.2.3) pour l'assimilation de cette classe de mesures reste valable. Comme toute source d'observation indirecte, on peut envisager leur utilisation en assimilation de données de deux façons :

- pseudo observations ;
- assimilation directe.

Dans cette section, nous parlerons essentiellement de l'approche basée sur les pseudo-observations, l'assimilation directe faisant l'objet du chapitre (4). Dans l'approche pseudo-observation, l'information pertinente est extraite des images et utilisée ensuite comme observation dans un schéma classique d'assimilation de données. Nous nous limiterons aux pseudo-champs de vent qui peuvent être extraits des séquences d'images. L'extraction de telles informations a été abordée dans le cadre de plusieurs travaux et nous allons donner ici un aperçu

des principales méthodes utilisées. Dans la seconde approche, les images sont directement utilisées comme observations dans un schéma d'assimilation de données (voir le chapitre 4).

Estimation de mouvement

L'estimation du mouvement désigne l'ensemble des techniques qui permettent de déterminer les vecteurs de mouvement décrivant la transformation d'une image à une autre. [Maurizot(1997), Corpetti(2002)] passent en revue l'essentiel des méthodes utilisées à cet effet. L'hypothèse de base est la conservation de la luminance [Horn and Schunck(1981)]. On suppose que l'intensité de gris (pour les images en niveau de gris) ou la couleur (pour les images en couleur) d'un point d'une image ne varie pas au cours du déplacement. Il existe plusieurs techniques d'estimation du mouvement. Nous donnons dans ce paragraphe une description sommaire de ces méthodes. Pour plus de détails, l'annexe (D) décrit la méthode du **flot optique**.

Estimation du mouvement par PIV

Une méthode d'estimation de mouvement couramment utilisée en physique expérimentale est la méthode PIV (Particle Image Velocities). Elle est basée sur des statistiques et est particulièrement utilisée sur des images de particules : un ensemble de particules est placé dans un champ de vitesse généré par le phénomène à étudier. La détermination des vecteurs de déplacement des particules permet d'estimer la dynamique du phénomène. La mise en œuvre de cette méthode est assez simple mais la complexité croît rapidement avec le nombre de particules. Les travaux d'Adrian [Adrian(1991)] constituent une bonne référence pour en savoir plus sur cette méthode.

Estimation du mouvement par corrélation

[Anandan(1989)] propose de corrélérer par translation une zone de la première image avec une zone de la deuxième image. Le vecteur de translation associé au maximum de corrélation est utilisé comme descripteur du mouvement du centre de la zone initiale.

Estimation du mouvement par les méthodes paramétriques

Ces méthodes font une étude globale sur l'ensemble du support de la fonction image. Ainsi, la paramétrisation bidimensionnelle exprime le mouvement d'un point à l'aide de polynômes de degré variable (généralement 2). $V(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}$ en degré 1, avec \mathbf{A} , une matrice carrée 2×2 et \mathbf{b} un vecteur sont les paramètres à estimer. On retrouve dans cette catégorie de méthodes la paramétrisation tridimensionnelle [Longuet-Higgins(1981), Fermüller and Aloimonos(2000), Hellier et al.(2001)].

Les méthodes de flot optique

Proposées par Horn et Schunck en 1981 [Horn and Schunck(1981)], ces méthodes sont fondées sur 2 hypothèses : la conservation de la luminance et la cohérence spatiale du champ de vitesse estimé. Cette cohérence spatiale définit la **régularisation** qui dépend de l'application. Les méthodes de cette classe sont décrites avec un peu plus de détails en annexe (D)

Modèles images

Les pseudo-champs de vent extraits par les techniques d'estimation du mouvement sont loin d'être parfaits : l'estimation est basée sur deux images et la physique du phénomène associé n'est pas prise en compte. On obtient en général des champs de vitesse qui n'ont aucune cohérence dans le temps. Afin de résoudre ces problèmes, [Herlin et al.(2006), Huot et al.(2006), Korotaev et al.(2007)] proposent une approche plus avancée prenant partiellement en compte la physique du problème : c'est l'estimation du mouvement par modèle image. Il est basé sur l'approche variationnelle d'assimilation de données. Il permet de prendre en compte plus de deux images tout en garantissant la cohérence dans le temps par l'usage d'un modèle physique décrivant la dynamique du problème en plus de l'hypothèse de la conservation de la luminance. La complexité est réduite en utilisant un modèle dynamique simplifié pour le champ de vitesse \mathbf{v} et non la physique complète du problème. On obtient ainsi un champ de vitesse *continu* dans le temps et cohérent pour une séquence d'images observées $I_{obs}(t)$, $t \in [0, T]$. Le problème d'estimation du mouvement par modèle image se définit par la minimisation de la fonction coût

$$F(\mathbf{v}_0) = \frac{1}{2} \int_0^T \|I - I_{obs}\|^2 dt + \frac{\lambda}{2} \|S(\mathbf{v}_0)\|^2. \quad (3.5)$$

où \mathbf{v}_0 est le champ de vitesse au temps initial : c'est la variable de contrôle. La solution à ce problème de minimisation est soumise aux contraintes

$$\begin{cases} \frac{\partial I}{\partial t} + \nabla I \cdot \mathbf{v} = 0, \\ \frac{\partial \mathbf{v}}{\partial t} = \mathcal{N}(\mathbf{v}), \quad I(0) = I_0, \quad \mathbf{v}(0) = \mathbf{v}_0. \end{cases} \quad (3.6)$$

La première équation traduit la contrainte de la conservation de la luminance et \mathcal{N} définit la dynamique du phénomène ; c'est le **modèle image**.

L'un des grands avantages de cette approche est la prise en compte de toutes les images de la séquence, ce qui permet de résoudre, entre autres, le problème de données manquantes.

Application : les champs de vecteurs atmosphériques

Une application de la technique des pseudo-observations est l'utilisation des champs de vecteurs atmosphériques. Les champs de vecteurs atmosphériques AMV sont des champs de vitesse dérivés des images satellite et utilisés comme pseudo observations de champs de vitesse dans les modèles météorologiques. Le principe d'extraction des champs de vecteurs atmosphériques consiste à localiser des cibles dans les séquences d'images puis à estimer leur déplacement par des techniques de corrélation. Le principe détaillé et ses améliorations peuvent se trouver dans les travaux de Holmlund [Bühler and Holmlund(1993), Dew and Holmlund(2000)] de Deb [Deb et al.(2008)] et de Kishtawal [Kishtawal et al.(2009)].

De récentes études se sont intéressées à la comparaison entre les données de vent AMV et les observations de vent fournies par d'autres techniques. On peut citer les travaux de Bedka [Bedka et al.(2009)] qui font une étude comparative des données AMV avec les observations obtenues par les radiosondes et les profileurs de vent. Les données AMV utilisées dans ces expériences sont extraites de séquences denses avec une image toutes les six minutes. Les

grandes différences observées avec les données conventionnelles à certaines couches de l’atmosphère sont compensés par la stabilité dans le temps et dans l’espace de ces différences et la résolution des données AMV qui est inégale par les autres types d’observation. D’autres travaux ont étudié l’impact des données AMV dans la prévision [Rohn et al.(2001)].

Cependant, les données AMV ne respectent pas les hypothèses de non corrélation spatiale et/ou temporelle des erreurs d’observations [Bormann et al.(2003b), Bormann et al.(2003a), Rohn et al.(2001)]. Ce qui explique la nécessité de la déflation opérée sur ce type d’information avant leur usage dans la plupart des centres opérationnels de prévision [Tsuyuki(2000)]. La difficulté d’obtenir l’impact des corrélations d’erreurs sur ce type d’observation dans un schéma d’assimilation de données suggère la nécessité de revoir les hypothèses sur les erreurs [Bormann et al.(2003b)].

L’utilité des champs de vecteurs atmosphériques ne se limite pas à l’utilisation comme pseudo-observations. Elles permettent, entre autres, l’estimation et la caractérisation de la dynamique méso-échelle [Bedka and Mecikalski(2005)] grâce à la couverture considérable des images par satellites. Leurs propriétés statistiques peuvent servir également d’indicateurs de qualité pour les images par satellite [Holmlund(1998)].

3.3.7 Ingrédients pour la prise en compte directe des observations de type image dans un schéma d’assimilation variationnelle de données

L’information visible sur les images étant présente (directement ou par calcul) dans les sorties des modèles aux erreurs de modélisation et de paramétrisation près, la meilleure utilisation de l’image serait de l’utiliser directement pour corriger les entrées du modèle : on recherche alors les paramètres (conditions initiales en l’occurrence) qui permettraient au modèle de produire le cliché le plus proche de celui que les images rendent. C’est l’assimilation directe d’images. Ainsi définie, l’assimilation d’images ne se fera pas de manière isolée mais en complément à l’assimilation d’autres types d’observation. Cela consiste à compléter la fonction coût par un terme d’image mesurant l’écart entre la variable d’état et les images. Ce terme prend la forme :

$$J_2(\mathbf{x}) = \int_0^T \int_{\Omega} \| \mathcal{H}(\mathbf{x}) - f \|_{\mathcal{F}}^2 d\omega dt, \quad (3.7)$$

où \mathcal{F} représentant l’espace des images, et \mathcal{H} l’opérateur d’observation des images défini de l’espace \mathcal{X} des variables d’état vers l’espace des images. Les principales difficultés du problème consistent à :

- définir \mathcal{F} qui ait une structure d’espace vectoriel doté de la norme $\| \cdot \|_{\mathcal{F}}$, dans lequel la différentiation de J_2 est possible, et qui permet de stocker les images dans des espaces mémoire réduits ;
- définir l’opérateur \mathcal{H} qui permettra à partir des sorties du modèle de définir un cliché image de la situation associée afin de pouvoir calculer l’écart aux observations images.

3.4 Conclusion

Les différentes sources d’observation sont complémentaires :

les observations conventionnelles sont très précises dans les régions habitées et permettent de déclencher des alertes fiables en cas de catastrophe météorologique. Le principal défaut est la couverture géographique limitée et l’inégale répartition sur le globe terrestre. Cette source

d'information est plus dense dans les pays riches, quasi inexistante dans les déserts et peu dense sur les océans.

Les observations par satellites géostationnaires sont peu précises mais ont une couverture géographique quasi-complète du globe. Elles permettent d'obtenir de l'information à une très grande fréquence même dans les zones non couvertes par les observations conventionnelles. Cette source est particulièrement importante dans les déserts et sur les océans. Sa grande résolution au niveau de l'équateur compense d'une certaine façon la rareté des mesures *in situ* dans les régions inter tropicales. La résolution se dégrade au fur et à mesure que l'on s'éloigne de l'équateur, les observations par satellites géostationnaires sont inexistantes au niveau des pôles.

Les observations par satellites à orbite (quasi polaires) font le compromis d'une part entre la couverture géographique des satellites géostationnaires (grâce à l'éloignement de la cible et au défilement) et d'autre part la résolution et la précisions des instruments conventionnels (grâce à leur orbite basse) pour fournir une information de bonne qualité à une bonne fréquence dans les régions polaires qui sont presque privées des autres sources d'observation.

Les images telles que nous les avons présentées représentent une importante source d'observations qui est cependant sous-utilisée. Les outils de traitement d'images permettent d'y extraire des données qui sont ensuite utilisées comme pseudo-observations. Cette approche viole certaines hypothèses fondamentales de l'assimilation de données et sont limitées dans la prise en compte de la physique des phénomènes observés. Le chapitre suivant présente un formalisme d'assimilation directe de séquences d'images qui permet de résoudre les principaux problèmes posés par l'approche pseudo-observation.

Deuxième partie

Vers une prise en compte des
observations de type image en
assimilation de données

Chapitre 4

Assimilation directe de séquences d'images

Les travaux exposés dans le présent chapitre ont fait l'objet des publications suivantes : [Titau et al.(2009)]; [Le Dimet et al.(2009)]; [Vidard et al.(2008)] et [Titau et al.(2008)]. Pour leur réalisation, nous avons développé une bibliothèque numérique pour la prise en compte des informations de type image et séquence d'images en assimilation de données. Bibliothèque baptisée BALAISE (Bibliothèque d'Assimilation LAgrangienne d'Images et de séquences d'images pour les Systèmes Environnementaux), elle est écrite en Fortran avec le formalisme Fortran 90/95 avec une partie en C/C++ pour la manipulation des images en curvelets. Sa structuration facilite sa lisibilité et les commentaires au standard f90doc permettent de générer automatiquement la documentation pour les développeurs. L'objectif à terme est d'avoir une bibliothèque pouvant s'utiliser aisément avec les systèmes numériques pour l'assimilation de données.

4.1 Observations de type image et leurs caractéristiques

Les images constituent une source importante d'observation de phénomènes physiques. Comme nous l'avons mentionné au chapitre (3), les images en particulier et les télémessures en général constituent une source incontournable pour le futur des modèles numériques de prévision.

4.1.1 Interprétation des images

L'interprétation des données de type image dépend de l'application. Selon les cas, seule une ou quelques unes de caractéristiques de l'image sont nécessaires. L'isolation de ces caractéristiques définit un **niveau d'abstraction** de l'image que nous appellerons également **niveau d'interprétation**. Nous en distinguerons deux : le niveau élémentaire et le niveau structuré.

Niveau élémentaire

Si on considère une image sous la forme basique d'un tableau de pixels, alors elle représente une expression instantanée de l'état du système étudié. Nous qualifierons ce niveau d'interprétation de **niveau élémentaire**. Chaque pixel donne alors l'état du système dans la région

correspondante. C'est une capture d'un état apparent du système, indiquant des quantités physiques qui peuvent être simulées par le modèle du système. Les exemples les plus simples sont les traceurs (aérosols en atmosphère ou chlorophylle en océanographie). En général, les quantités physiques mesurées par les satellites ne sont pas des variables du modèle mathématique. Il existe dans la plupart des cas une relation entre les deux quantités, mais elle est en général très complexe comme nous l'avons déjà mentionné à la section (3.2.3). Par exemple, la couverture nuageuse mesurée par les satellites dans le canal visible est une fonction très complexe des variables d'état des modèles météorologiques. Elle résulte d'un processus incluant la micro physique à l'origine des nuages. Les radiances sont dépendantes des processus thermodynamiques et sont donc influencées par la température et l'humidité. Par conséquent, les images satellite dans certains canaux donnent des informations sur la température du système observé (voir la figure (4.1b)). Au niveau pixel, ces images contiennent une grande quantité d'informations non structurées. Comme nous l'avons mentionné au chapitre (3), les images satellite ont généralement une résolution horizontale plus fine que celle des modèles de prévision au niveau global. Pour rappel, une image METEOSAT de seconde génération du canal visible haute résolution est de 6000×11000 . Si l'on multiplie cette quantité par le nombre d'images à prendre en considération, cela représente une quantité importante d'informations nécessitant un espace mémoire important pour les traitements informatiques. Pour cette raison, le niveau élémentaire sous forme de pixels doit être évité dans certaines applications. Il est cependant important de noter que les algorithmes basiques de traitement d'images (débruitage, rehaussement de contraste, inpainting, segmentation, ...) sont généralement définis et appliqués au niveau pixel.

Niveau structuré

Dans certaines applications, l'image est explorée en terme de la structuration globale de son contenu, notamment en terme d'objets composants la scène observée ainsi que leur caractérisation géométrique. Nous qualifierons ce niveau d'interprétation de niveau **structuré**. A ce niveau d'interprétation, l'information est moins dépendante du processus d'acquisition et de la quantité physique mesurée. Cette dépendance faible est illustrée par la figure (4.1) qui montre deux images satellite du Gulf Stream dont une dans le canal visible et l'autre dans le canal infrarouge. Malgré la différence dans les canaux d'acquisition, les deux images présentent, au niveau global, les mêmes structures visibles. Ce niveau d'interprétation est utilisé pour des problématiques de vision par ordinateur tels que l'estimation du mouvement, la détection de contours, la reconnaissance de forme, etc. Nous utiliserons dans la suite du texte, le terme *structure* pour désigner les éléments qui composent une image au niveau global : ce sont les objets ou leurs contours dans le cas des solides et les formes dans le cas des fluides. Dans l'un ou l'autre cas, l'information importante se trouve dans ou la structuration des pixels qui forment les limites des éléments d'intérêt, d'où le terme structure. Si le niveau élémentaire (pixels) permet de s'intéresser aux quantités physiques, le niveau structuré permet de regarder l'organisation ou la structuration globale de l'information. Moyennant des outils de caractérisation appropriés, l'espace mémoire requis pour le stockage peut être substantiellement réduit dans ce dernier cas.

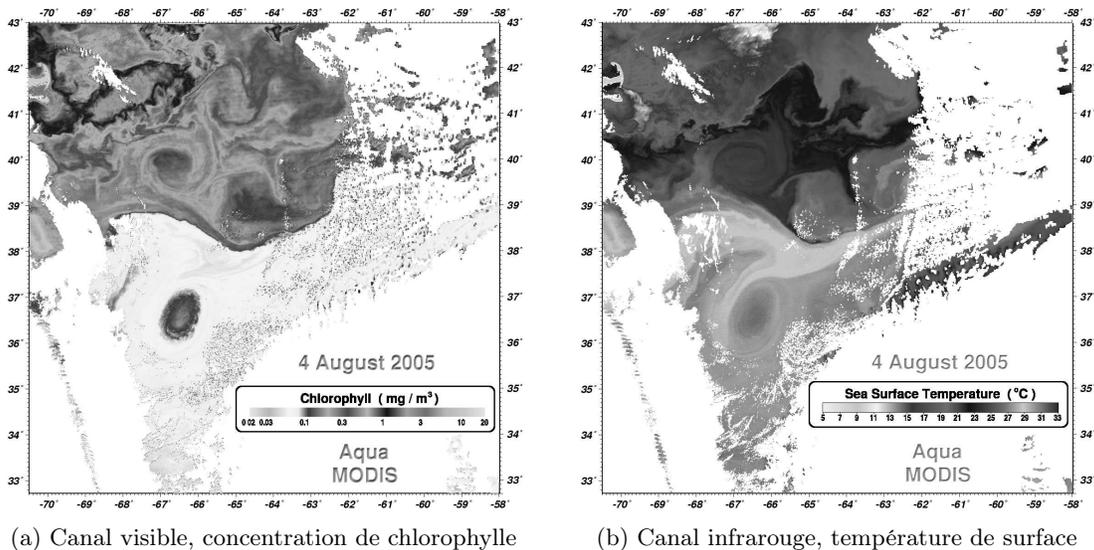


FIGURE 4.1 – Vue du Gulf Stream en la date du 4 Août 2005 par le satellite Aqua (doté du radiomètre MODIS). Les structures observées sur les deux images sont presque identiques malgré la différence de canal d’acquisition (quantités physiques observées). Source : NASA - <http://oceancolor.gsfc.nasa.gov>.

4.1.2 Représentation mathématique des images

Considérée comme une fonction mathématique, une image peut être représentée par ses coefficients dans une base appropriée d’un espace fonctionnel. Dans cette section, nous allons présenter, sans rentrer dans les détails, quelques exemples de représentation de l’image. Nous parlerons de la représentation sous forme de pixels, la représentation en ondelettes et la représentation en curvelets. Ces deux dernières représentations sont plus adaptées pour le traitement du signal en général et pour le traitement d’images en particulier.

Représentation sous forme de pixels

Dans cette représentation, l’image est une matrice de pixels. Cette représentation correspond à une discrétisation de la fonction image sur la grille définie par le quadrillage de l’instrument de capture. Cette représentation se prête par défaut à l’interprétation au niveau élémentaire. L’espace mémoire requis est fonction de la définition (nombre de points composant l’image) et la qualité dépend de la résolution (nombre de pixels par unité de longueur ou de surface). Cette représentation est moins adaptée aux problématiques de la vision par ordinateur. Rappelons qu’en vision par ordinateur, l’information utile dans les images est localisée dans les discontinuités. Les algorithmes de compression peuvent être utilisés pour réduire l’espace de stockage si l’image contient peu de discontinuité. Cependant la compression nécessite la réécriture des algorithmes de traitements pour les adapter au format compressé.

Décomposition dans des bases appropriées

Une image 2D peut être considérée comme une fonction définie sur un sous-ensemble de \mathbb{R}^2 dont la valeur représente l’intensité de la couleur ou de la luminosité en chaque point

(pixel). Vue comme une fonction, une image peut se décomposer dans une base de fonctions mathématiques (Fourier, ondelettes, curvelets, etc.). Ce sont des bases très adaptées pour le traitement du signal (l'image est un cas particulier de signal). Ainsi représentée dans une base, il est courant de ne stocker que les coefficients essentiels à la reconstruction du signal : c'est l'opération de seuillage. Son efficacité dépend de la proportion de discontinuité dans le signal et du niveau de détail recherché. Cette famille de représentation est mieux adaptée dans un contexte mathématique où pourront être exploités les outils de l'analyse fonctionnelle.

Décomposition dans une base de Fourier

L'analyse en série de Fourier peut se résumer en deux formules :

– la formule de décomposition :

$$c_n = \int_{-\infty}^{+\infty} f(x) e^{-inx} \frac{dx}{2\pi} = \langle f(x), e^{inx} \rangle; \quad (4.1)$$

– la formule de reconstruction :

$$f(x) = \sum_0^{+\infty} c_n e^{-inx}. \quad (4.2)$$

Les c_i sont les coefficients des fonctions de la base de Fourier, e^{-inx} les fonctions de la base de Fourier et $f : \mathbb{R} \rightarrow \mathbb{R}$ la fonction considérée. L'algorithme de transformation en série de Fourier et son inverse ont une complexité en $\mathcal{O}(n \log n)$ où n représente le nombre de points constituant le signal étudié. Avec cette représentation, les opérations de traitement d'images telles que le seuillage se définissent aisément par élimination des coefficients associés aux fonctions de base moins représentatifs. Le nombre de coefficients stockés dépend du niveau de détail souhaité pour la reconstruction. En l'occurrence, si la décomposition dans une base de Fourier est utilisée pour une fonction f définie sur l'intervalle $[0, 1]^2$ présentant une discontinuité suivant une courbe C^2 et régulière partout ailleurs, alors, une reconstruction avec les m meilleurs coefficients issus de la décomposition induit une erreur de reconstruction $\|f - f_m\|$ en $\mathcal{O}(m^{-1/2})$. Rappelons qu'une telle reconstruction n'est pas possible avec la représentation en pixels.

Décomposition dans une base d'ondelettes L'analyse en série de Fourier est particulièrement adaptée aux signaux stationnaires et est limitée par le principe d'incertitude de Heisenberg. D'autre part, l'analyse en série de Fourier n'est pas très adaptée aux fonctions présentant des discontinuités. Bien que la transformée de Fourier fenêtrée résout partiellement ces problèmes, l'analyse en ondelettes est mieux adaptée dans ces situations. La pratique des ondelettes remonte aux travaux de Grossman et Morlet [Grossmann and Morlet(1984)]. Elle est rendue populaire par la théorie de l'analyse multi-résolution dont on peut trouver les fondements dans les travaux de Mallat [Mallat(1989a)] et de Meyer [Meyer(1992)]. La littérature est abondante sur la théorie et la pratique des ondelettes. Les travaux de Mallat [Mallat(1989b)], de Coifman [Coifman(1990)] et de Cohen [Cohen(1992)] constituent de bonnes références pour l'étude des ondelettes. L'analyse en ondelettes utilise comme fonction d'analyse les translations, dilatations, et compressions d'une seule et même fonction ψ localisée en temps et en fréquence appelée ondelette mère. ψ étant connu, on définit les ondelettes par :

$$\psi_{a,b} = \frac{1}{\sqrt{a}} \psi \left(\frac{x-b}{a} \right) \text{ avec } a > 0; b \in \mathbb{R}, \quad (4.3)$$

où a (coefficient de dilatation) représente l'échelle et b (coefficient de translation) représente la position spatiale.

La localisation en temps et en fréquence de ψ suppose qu'il existe une constant scalaire κ telle que :

$$\|\psi(t)\| \leq \kappa(1+|t|)^{-1-\epsilon}, \|\hat{\psi}(w)\| \leq \kappa(1+|w|)^{-1-\epsilon}, \quad (4.4)$$

pour tout $\epsilon > 0$. $\hat{\psi}$ représente la transformée de Fourier. Le facteur $1/\sqrt{|a|}$ assure que la fonction $\psi_{a,b}$ ait une norme \mathcal{L}^2 constante. Une fonction se décompose alors suivant la formule :

$$C(a, b) = \langle f, \psi_{(a,b)} \rangle, \quad (4.5)$$

où C représente le coefficient associé à l'ondelette $\psi_{a,b}$.

La reconstruction est donnée par :

$$f = \frac{1}{C_\psi} \iint C(a, b) \psi_{a,b} \frac{da}{a^2} db, \quad (4.6)$$

avec :

$$C_\psi = \int_0^{+\infty} |\hat{\psi}(\xi)|^2 \frac{d\xi}{\xi}. \quad (4.7)$$

Dans la pratique, les coefficients de dilatation et de translation a et b sont choisis dans un espace discret. En général on choisi $a = 2^{-j}$ et $b = k.2^j$ avec $k, j \in \mathbb{Z}$. La reconstruction est ainsi approchée par une série. Si la décomposition dans une base d'ondelettes est utilisée pour une fonction f définie sur l'intervalle $[0, 1]^2$ présentant une discontinuité suivant une courbe C^2 et régulière partout ailleurs, alors, une reconstruction avec les m meilleurs¹ coefficients issus de la décomposition permet d'avoir une erreur de reconstruction $\|f - f_m\|$ en $\mathcal{O}(m^{-1})$. C'est une reconstruction plus précise qu'avec une base de Fourier. La complexité de la décomposition et de la reconstruction dans une base d'ondelettes dyadique est la même que celle de la décomposition ou de la reconstruction dans une base de Fourier. Un des grands avantages de la décomposition en ondelettes est la simplicité de la formulation, la multiplicité des bases d'ondelettes et l'abondance de la littérature dans ce domaine. Ce qui en fait un outil de choix pour le traitement du signal.

Décomposition en curvelets

La transformation en curvelets est une décomposition multi-échelle pyramidale, multi-directionnelles à chaque échelle, avec des éléments filiformes à l'échelle la plus fine. Cette décomposition permet une représentation optimale (avec peu de coefficients) des objets présentant une discontinuité suivant une courbe C^2 et régulière partout ailleurs. Avec cette décomposition, une reconstruction partielle de l'objet f avec m coefficients entraîne une erreur de reconstruction :

$$\|f - f_m\| \leq C.(\log(m))^3.m^{-2} \quad (4.8)$$

qui est optimal dans le sens où on ne peut pas faire mieux [Candes and Donoho(2005)]. Ceci permet une représentation optimale des discontinuités et donc des contours avec très peu de coefficients.

1. meilleurs dans le sens où ils permettent de faire la reconstruction la plus fidèle possible dans la base d'ondelettes choisie

Représentation en curvelets. Soit une fonction $f \in \mathcal{L}^2(\mathbb{R}^2)$, sa décomposition en curvelets s'écrit :

$$f = \sum_{j,l,\mathbf{k}} c(j, l, \mathbf{k}) \phi_{j,l,\mathbf{k}} \quad (4.9)$$

avec :

$$c(j, l, \mathbf{k}) = \langle f, \phi_{j,l,\mathbf{k}} \rangle = \int_{\mathbb{R}^2} f(\mathbf{x}) \overline{\phi_{j,l,\mathbf{k}}(\mathbf{x})}, \quad (4.10)$$

où $c(j, l, \mathbf{k})$ représente le coefficient associé à la curvelet $\phi_{j,l,\mathbf{k}}$ et $\overline{\phi_{j,l,\mathbf{k}}(\mathbf{x})}$, le conjugué complexe de $\phi_{j,l,\mathbf{k}}(\mathbf{x})$. $j, l, \mathbf{k} = (k_1, k_2)$ représentent respectivement l'indice d'échelle, l'indice d'orientation et l'indice de position. Cette décomposition obéit à la relation de Parseval. Si une fonction $f \in \mathcal{L}^2(\mathbb{R}^2)$ est décomposée en (série de) curvelets :

$$f = \sum_{j,l,\mathbf{k}} \langle f, \phi_{j,l,\mathbf{k}} \rangle \phi_{j,l,\mathbf{k}}, \quad (4.11)$$

alors ses coefficients obéissent à :

$$\sum_{j,l,\mathbf{k}} |\langle f, \phi_{j,l,\mathbf{k}} \rangle|^2 = \|f\|_{\mathcal{L}^2(\mathbb{R}^2)}^2. \quad (4.12)$$

Contrairement aux ondelettes où il existe des bases, les curvelets ne constituent pas une base, mais elles ont de bonnes propriétés dont le fait d'être une frame de Parseval. L'implémentation des auteurs [Candes et al.(2006)] est disponible² en open source pour des besoins d'ordre académiques et de recherche.

Remarque 4.1.1 *Que ce soit avec un base de type Fourier, une base de type ondelette ou une frame de type curvelets, la relation de Parseval est vérifiée. Elle permet de calculer la norme \mathcal{L}^2 de la fonction décomposée comme une somme des coefficients issus de la décomposition. Ces décompositions définissent donc des espaces normés avec des propriétés de stockage très intéressantes.*

Remarque 4.1.2 *Il existe également une représentation géométrique sous forme vectorielle des images. L'image est considérée comme un ensemble d'éléments géométriques (ligne, polygones, cercles, ...). Cette représentation est spécialement adaptée aux opérations géométriques (homothétie, mise à l'échelle, etc.). L'espace mémoire requis ne dépend que sa régularité géométrique et la qualité est indépendante des propriétés de l'image.*

Remarque 4.1.3 *Les travaux de Chan et Shen [Chan and Shen(2005)] constituent une bonne référence pour les problèmes de traitement d'images. Ces travaux détaillent les principales méthodologies de modélisation, de représentation et de traitement d'images et leurs interconnexions. Le lecteur y trouvera, en plus des approches de représentation ci-dessus présentées, les approches basées sur les équations aux dérivées partielles ainsi que les approches stochastiques.*

2. <http://www.curvelet.org/software.html>

4.1.3 Séquences d'images

Avec une image unique, on n'a accès qu'à l'information de type Eulérien. Quand on s'intéresse à une succession d'image dans le temps, ce qui constitue une séquence d'images, on peut aussi obtenir de l'information dynamique. Les images peuvent ainsi être interprétées comme des observations lagrangiennes. L'utilité et l'utilisation de ce type d'information constituent un champ de recherche important pour les problèmes d'assimilation de données. Des travaux tels que ceux de Honorat [Honorat(2007)] ou encore de Nodet [Nodet(2005)] représentent des exemples d'utilisation de l'information Lagrangienne en assimilation de données. La figure (4.2) représente une séquence d'images satellite prise au dessus de l'Europe et de l'Atlantique Nord. Cette séquence montre un front sur l'Europe de l'Ouest et un écoulement cyclonique au dessus de l'océan ainsi que leur évolution dans le temps. Ceci illustre le potentiel informatif que contiennent les images et ce que la vision humaine peut en tirer (reconnaissance et suivi de l'évolution de certaines structures caractéristiques des écoulements atmosphériques). A partir d'une telle séquence, on est ainsi en mesure d'estimer grossièrement l'évolution de l'état de l'atmosphère à partir des caractéristiques reconnues. C'est une opération qui est faite par les experts et constitue une source de données supplémentaire pour les modèles numériques de prévision dans les centres opérationnels.

On peut donc naturellement se poser la question de savoir s'il n'est pas possible d'automatiser un tel processus en utilisant les observations de type séquence d'images directement dans un schéma d'assimilation de données. C'est ce que nous analyserons dans la suite de ce chapitre. Un schéma d'assimilation variationnelle de données semble le schéma le plus adapté pour une telle étude. En effet, un tel schéma prend en compte toutes les observations dans une fenêtre d'assimilation de données en une seule fois. Toutes les images d'une séquence peuvent ainsi être traitées ensemble avec toute la cohérence physique du modèle qui est incluse dans le système d'optimalité de l'assimilation variationnelle de données. La prise en compte d'une séquence d'images peut être étudiée de deux façons : premièrement la séquence peut être considérée comme un ensemble d'images $2D$ chacune traitée et analysée séparément en utilisant les concepts de traitement d'images listés plus haut. Une deuxième approche consistant à considérer la séquence entière comme une donnée en trois dimensions (les deux dimensions spatiales et la dimension temps). L'extension des concepts du traitement d'images $2D$ (voir les sections (3.3.5), (4.1.1) et (4.1.2) pourrait permettre de mieux exploiter la consistance en temps de la séquence d'images.

4.1.4 Utilisation classique des images en assimilation de données

L'approche classique d'utilisation des images en particulier et des observations indirectes en général consiste à y extraire l'information pertinente. Cette information est ensuite utilisée comme observations dans un schéma classique d'assimilation de données [Herlin et al.(2004)] : c'est l'approche pseudo-observation. C'est le cas des champs de mouvement atmosphérique AMV en prévision météorologique. L'extraction d'informations (notamment des champs de vitesse) à partir des images est abordée dans le chapitre (3) et dans l'annexe (D). L'utilisation des pseudo-observations pose cependant quelques problèmes parmi lesquels le non respect des hypothèses de non corrélation des erreurs d'observations dans le temps et dans l'espace ainsi que leur non corrélation avec les erreurs d'ébauche [Derber and Bouttier(1999), Daley(1993), Lorenc et al.(2000)].

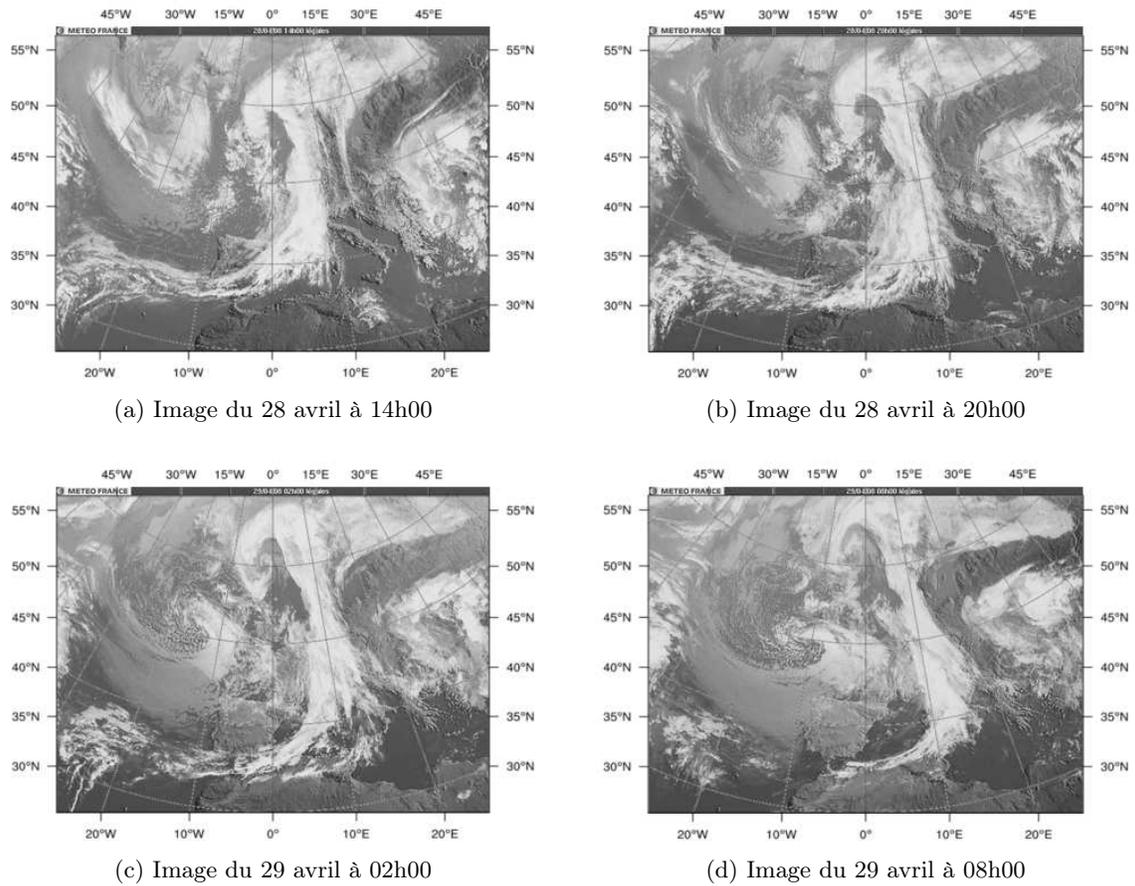


FIGURE 4.2 – Séquence d'images METEOSAT de l'évolution de la couverture nuageuse au dessus de l'Europe et de l'Atlantique Nord ; période du 28 avril 2008 au 29 avril 2008. (images Météo-France)

Les erreurs sur les pseudo-observations sont souvent corrélées en espace et en temps. C'est le cas des données AMV [Bormann et al.(2003b), Bormann et al.(2003a), Rohn et al.(2001)]. Ce qui explique la nécessité de la déflation opérée sur ce type d'information avant leur usage dans la plupart des centres opérationnels de prévision [Tsuyuki(2000)]. De plus, la difficulté d'obtenir l'impact des corrélations d'erreurs sur ce type d'observation dans un schéma d'assimilation de données suggère la nécessité de revoir les hypothèses sur les erreurs [Bormann et al.(2003b)].

Les pseudo-observations sont en général calculées à partir des données d'ébauche, les erreurs associées sont donc corrélées aux erreurs d'ébauche, ce qui est en contradiction avec une autre hypothèse de base de l'assimilation variationnelle de données. En effet, les erreurs d'observation sont supposées indépendantes de l'erreur d'ébauche. Si les pseudo-observations sont donc utilisées conjointement aux observations conventionnelles, elles peuvent contribuer à donner un poids relatif plus élevé et non contrôlé au terme d'ébauche. Pour éviter ce problème, on peut utiliser les deux sources d'information séparément l'une avant l'autre, le dernier permettant d'améliorer le résultat obtenu par le premier.

4.2 Cadre général de l'assimilation directe de séquences d'images

4.2.1 Contexte

Objectifs visés

L'approche d'utilisation des observations de type image proposée dans cette section vise deux principaux objectifs.

D'une part, on voudrait supprimer l'étape intermédiaire d'extraction de pseudo-observations. On résout ainsi trois des principaux problèmes rencontrés avec cette approche. En particulier :

- l'information image est utilisée de façon cohérente avec la physique du problème ;
- l'erreur sur l'ébauche reste non corrélée à l'erreur sur les observations ;
- le terme d'ébauche n'est plus utilisé deux fois et donc son poids relatif dans la fonction coût reste contrôlable.

D'autre part, on ne voudrait modifier ni le schéma d'assimilation de données ni le modèle associé à un problème donné. Pour être opérationnel dans une application réaliste, un modèle nécessite quelques années de développement. C'est notamment le cas pour les modèles de prévision météorologique. Il serait donc irréaliste de proposer une méthodologie de prise en compte d'une nouvelle classe d'observations qui nécessite de développer un nouveau modèle. Un changement du schéma d'assimilation de données est aussi à éviter sauf s'il reste compatible avec les autres classes d'observations. Dans tous les cas, à moins que l'impact de l'utilisation de la nouvelle classe d'observations ne le justifie, le changement de modèle ou de schéma d'assimilation de données devrait être évité.

Motivations et pré requis pour l'assimilation directe de séquences d'images

L'information visible sur les images est présente (directement ou par calcul) dans les sorties des modèles aux erreurs de modélisation, de représentativité et de paramétrisation près. La meilleure exploitation de l'image serait de l'utiliser directement pour corriger les entrées du modèle. On recherche alors les paramètres (conditions initiales en l'occurrence) qui permettent au modèle de produire le cliché le plus proche possible de celui que les images rendent. C'est

l'assimilation directe d'images. Ainsi définie, l'assimilation d'images est un complément à l'assimilation des autres classes d'observations. L'assimilation directe d'images nécessite de compléter la fonction coût par un terme d'images mesurant l'écart entre la variable d'état et les observations de type image. Tel que nous voulons l'introduire ici, la prise en compte des informations de type image en assimilation variationnelle de données nécessite de relever deux principaux défis :

- La définition d'une norme appropriée pour la prise en compte de cette classe d'informations dans la fonction coût ;
- la définition de termes de régularisation appropriés pour résoudre les problèmes inhérents aux informations de type image. Le chapitre (5) y est entièrement consacré.

Le premier défis ne se limite pas à la définition d'une norme mais tient aussi compte de la représentation de l'information de type image. C'est à dire l'espace mathématique approprié pour cette classe d'information. La norme appropriée se résume alors à la meilleure norme dans cet espace.

Le terme d'écart aux observations images prendrait la forme :

$$J_2(\mathbf{x}) = \frac{1}{2} \int_0^T \int_{\Omega} \| \mathcal{H}_{\mathcal{X} \rightarrow \mathcal{F}}(\mathbf{x}) - f \|_{\mathcal{F}}^2 d\omega dt, \quad (4.13)$$

avec Ω , l'espace physique, \mathcal{F} , l'espace des images, f , la fonction image et $\mathcal{H}_{\mathcal{X} \rightarrow \mathcal{F}}$ l'opérateur d'observation associé aux images. Cet opérateur est défini de l'espace \mathcal{X} des variables d'état vers l'espace des images \mathcal{F} . Les principales difficultés consistent à :

- définir l'espace \mathcal{F} qui ait une structure d'espace vectoriel doté de la norme $\| \cdot \|_{\mathcal{F}}$, dans lequel la différentiation de J_2 est facile, et qui permet de stocker les images dans des espaces mémoire réduits ;
- définir l'opérateur $\mathcal{H}_{\mathcal{X} \rightarrow \mathcal{F}}$ qui permet à partir des sorties du modèle de définir un cliché image de la situation associée afin de pouvoir calculer l'écart aux observations images.

Cette approche d'utilisation de l'information de type image, qui la prend en compte directement dans le système d'optimalité est plus cohérente avec la philosophie de l'assimilation variationnelle de données. Les images sont en effet utilisées de façon cohérente avec la physique du modèle. Certains problèmes rencontrés avec les approches par pseudo-observations sont automatiquement résolus. En l'occurrence le problème de corrélation entre les erreurs d'ébauche et les erreurs d'observation. L'apport de chaque source d'information (ébauche et observations) peut être quantifié et contrôlé.

4.2.2 Prise en compte des images dans la fonction coût

Dans l'optique d'assimiler directement les images, la fonction coût (équation 2.42) doit être redéfinie afin de prendre en compte le terme d'écart aux observations de type image (équation 4.13). Elle peut se réécrire sous la forme :

$$J(\mathbf{x}) = \underbrace{\frac{1}{2} \int_0^T \| \mathcal{H}(\mathbf{x}(t), t) - \mathbf{y}^o(t) \|_{\mathbf{R}(t)^{-1}}^2 dt}_{\text{écart aux observations conventionnelles}} + \underbrace{\frac{1}{2} \int_0^T \| \mathcal{H}_{\mathcal{X} \rightarrow \mathcal{F}}(\mathbf{x}(t), t) - f(t) \|_{\mathcal{F}}^2 dt}_{\text{écart aux observations images}}, \quad (4.14)$$

avec $f(t)$ la fonction image au temps t et $\|\cdot\|_{\mathcal{F}}$ la norme appropriée dans l'espace des images \mathcal{F} . Dans cette équation, nous avons omis les termes de régularisation dont le terme d'écart à l'ébauche.

L'opérateur d'image $\mathcal{H}_{\mathcal{X} \rightarrow \mathcal{F}}$ permet de calculer l'équivalent image de l'état du système au temps t à partir des valeurs des variables d'état à cette date. Par la suite, nous appellerons $\mathcal{H}_{\mathcal{X} \rightarrow \mathcal{F}}$, opérateur d'image synthétique (les raisons seront précisées par la suite).

L'opérateur d'image synthétique ainsi défini permet de construire des images associées au niveau d'interprétation élémentaire, c'est à dire au niveau pixel. Les travaux de Papadakis [Papadakis and Mémin(2008b)] montrent que l'assimilation directe d'images au niveau pixel produit des résultats de meilleure qualité que l'utilisation des pseudo-observations de champs de vitesse calculés par flot optique. Un formalisme similaire est utilisé dans [Corpetti et al.(2009)] pour l'assimilation directe d'images de gradient de pression avec un modèle atmosphérique simplifié à trois couches. Dans l'un ou l'autre cas, les résultats montrent clairement qu'il est préférable de faire une assimilation directe d'images que d'utiliser les pseudo-observations de champs de vitesse. L'assimilation directe d'images au niveau pixel peut cependant se heurter à deux limites si on l'applique à des problèmes géophysiques réalistes :

- La taille des images : une image brute en pixels représente une quantité importante de données. Par exemple, une image satellite MOP/MTP³ du canal visible (couvrant à peu près un quart du globe) représente 5000×5000 pixels. Soit 2.5×10^7 données. C'est du même ordre sinon plus que la quantité de données assimilées quotidiennement par les centres météorologiques opérationnels ($10^6 - 10^7$). Avec les nouvelles générations de satellites, une image du canal visible haute résolution (MSG HRV⁴) contient environ 6000×12000 pixels. Si on prend en compte le nombre d'images dans une séquence et le nombre de canaux à exploiter, on retrouve des quantités de données plus importante. Pour illustration, un satellite METEOSAT de seconde génération est doté de 12 canaux de capture et délivre une image chaque quart d'heure ;
- caractéristiques à l'échelle fine : la résolution des images satellite est plus fine que celle des modèles de prévision globale. Les images MSG HRV ont une résolution de $1km$ ($3km$ pour les autres canaux) au point sous satellite (résolution qui va se raffiner avec les futurs satellites $0.5km$ pour les prochains satellites américains GOES⁵). Le radiomètre MODIS⁶ qui équipe les satellites à orbite polaire Terra et Aqua produit des images avec une résolution variant entre $250m$ et $1000m$ selon le canal de capture. Les modèles numériques de prévision globale ont, quant à eux, des résolutions plus grossières. C'est le cas du modèle ARPEGE global de Météo-France dont la résolution varie entre 15 et $90km$. C'est aussi le cas du système de prévision océanographique global MERCATOR-PSY3V2⁷ qui utilise un modèle numérique avec une résolution horizontale au quart de degré ($6 - 26km$). Cette finesse de la résolution des images permet de mieux comprendre certains phénomènes non reproductibles avec la résolution actuelle des modèles. Il est difficile au niveau pixel de conserver encore cet avantage de la finesse de la résolution des images. De même au niveau pixel, l'erreur de représentativité pourrait s'accroître. En

3. Meteosat Operational/Transitional Program

4. Meteosat Second Generation - High Resolution Visible

5. Geostationary Operational Environmental Satellite

6. Moderate Resolution Imaging Spectroradiometer

7. Actuellement, MERCATOR permet de faire des prévisions au douzième de degré, voir le site web du projet http://www.mercator-ocean.fr/html/mercator/index_en.html

effet les structures de petite échelle observées sur les images peuvent être la conséquence de la dynamique à grande échelle. Les filaments en spirales observés sur la dynamique de certains traceurs en océanographie résultent de la variabilité du champ de vitesse de la méso échelle [Lehahn et al.(2007)]. L'utilisation d'une représentation locale telle que la représentation en pixel n'est pas appropriée dans ces situations. De plus l'échelle locale des pixels ne permet pas de représenter convenablement les structures cohérentes telles que les filaments. D'où la nécessité d'étudier d'autres outils de représentation des images, notamment ceux qui facilitent l'interprétation au niveau structuré et permettent de filtrer automatiquement et proprement certaines caractéristiques non souhaitées du signal (bruit et processus non représentés à l'échelle étudiée).

Interprétée au niveau structuré tel que présenté au paragraphe (4.1.1), une image est caractérisée par les structures cohérentes qui la constituent. Ces structures sont par exemple des fronts pour l'atmosphère et des filaments pour l'océan. Ce niveau d'interprétation s'adapte logiquement à l'étude de divers phénomènes aussi bien à l'échelle locale qu'à l'échelle globale. Elle se prête mieux à l'utilisation des observations de type image (échelle fine) de façon conjointe avec les modèles et les autres sources d'observation (échelle plus grossière). Si on pouvait trouver des approximations réalistes des structures contenues dans les images dans un espace fonctionnel \mathcal{S} doté d'une norme appropriée $\|\cdot\|_{\mathcal{S}}$, alors le calcul de l'écart aux observations de type image dans un tel espace et s'écrirait :

$$J(\mathbf{x}) = \frac{1}{2} \int_0^T \|\mathcal{H}(\mathbf{x}(t), t) - \mathbf{y}^o(t)\|_{\mathbf{R}(t)^{-1}}^2 + \frac{1}{2} \int_0^T \|\mathcal{H}_{\mathcal{X} \rightarrow \mathcal{S}}(\mathbf{x}(t), t) - \mathcal{H}_{\mathcal{F} \rightarrow \mathcal{S}}(f(t))\|_{\mathcal{S}}^2, \quad (4.15)$$

où $\mathcal{H}_{\mathcal{F} \rightarrow \mathcal{S}}$ est l'opérateur qui permet de passer de l'espace des images représentées en pixels \mathcal{F} à l'espace des images représentées sous forme structurée \mathcal{S} . Il permet de passer du niveau d'interprétation élémentaire au niveau d'interprétation structuré. L'opérateur $\mathcal{H}_{\mathcal{X} \rightarrow \mathcal{S}}$ est défini de l'espace d'état du système \mathcal{X} vers l'espace des images représentées sous forme structurée \mathcal{S} . La norme $\|\cdot\|_{\mathcal{S}}$ dépend de la définition mathématique de l'image au niveau structuré. Nous donnerons des exemples de tels opérateurs et espaces dans les sections suivantes. Si on choisit $\mathcal{S} = \mathcal{F}$ et $\mathcal{H}_{\mathcal{F} \rightarrow \mathcal{S}}[f] = f$, alors la formulation de la fonction coût donnée par l'équation (4.15) se réduit à celle de l'équation (4.14).

Remarque 4.2.1 *L'opérateur $\mathcal{H}_{\mathcal{F} \rightarrow \mathcal{S}}$ peut être entièrement intégré dans la norme $\|\cdot\|_{\mathcal{S}}$, on se ramène à une formulation similaire à celle décrite par l'équation (4.14).*

Avec la formulation variationnelle du problème d'assimilation de données telle que présentée à la section (2.5) et ce nouveau terme prenant en compte l'écart aux observations de type image, un nouveau terme doit être ajouté au membre de droite du modèle adjoint défini par l'équation (2.90). Pour rappel, ce modèle est donné par :

$$\begin{cases} \frac{d\mathbf{p}}{dt} + \left[\frac{\partial \mathcal{M}}{\partial \mathbf{x}} \right]^T \mathbf{p} = \left[\frac{\partial \mathcal{H}}{\partial \mathbf{x}} \right]^T \mathbf{R}^{-1} (\mathcal{H}(\mathbf{x}) - \mathbf{y}^o), t \in [0, T], \\ \mathbf{p}(T) = 0. \end{cases} \quad (4.16)$$

En suivant le schéma de dérivation du modèle adjoint présenté à la section (2.5), le nouveau

terme sera de la forme :

$$\left[\frac{\partial \mathcal{H}_{\mathcal{X} \rightarrow \mathcal{S}}}{\partial \mathbf{x}} \right]^T \mathbf{R}_{\mathcal{S}}^{-1} (\mathcal{H}_{\mathcal{X} \rightarrow \mathcal{S}}(\mathbf{x}) - \mathcal{H}_{\mathcal{F} \rightarrow \mathcal{S}}(f)), \quad (4.17)$$

où $\left[\frac{\partial \mathcal{H}_{\mathcal{X} \rightarrow \mathcal{S}}}{\partial \mathbf{x}} \right]^T$ représente l'adjoint du linéarisé du nouvel opérateur d'observation associé aux images et $\mathbf{R}_{\mathcal{S}}$ la matrice de covariance d'erreurs d'observation dans l'espace \mathcal{S} . Le modèle adjoint devient alors :

$$\begin{cases} \frac{d\mathbf{p}}{dt} + \left[\frac{\partial \mathcal{M}}{\partial \mathbf{x}} \right]^T \mathbf{p} = \left[\frac{\partial \mathcal{H}}{\partial \mathbf{x}} \right]^T \mathbf{R}^{-1} (\mathcal{H}(\mathbf{x}) - \mathbf{x}^o) \\ \quad + \left[\frac{\partial \mathcal{H}_{\mathcal{X} \rightarrow \mathcal{S}}}{\partial \mathbf{x}} \right]^T \mathbf{R}_{\mathcal{S}}^{-1} (\mathcal{H}_{\mathcal{X} \rightarrow \mathcal{S}}(\mathbf{x}) - \mathcal{H}_{\mathcal{F} \rightarrow \mathcal{S}}(f)), t \in [0, T], \\ \mathbf{p}(T) = 0. \end{cases} \quad (4.18)$$

Du fait de la nécessité de l'adjoint de l'opérateur d'observation, il est souhaitable de choisir ou de définir des opérateurs différentiables. Cependant, la différentiabilité de tels opérateurs dépend de la quantité physique observée et du processus d'observation. Ce problème est également rencontré avec toute observation constituée de mesures indirectes des variables d'état du système. Avec le formalisme que nous venons d'introduire et une bonne définition des opérateurs associés, il est possible de prendre en compte plusieurs séquences d'images différentes dans un même processus d'assimilation de données en complétant la fonction coût par des termes appropriés.

4.2.3 Opérateurs d'observation pour les images

Pour aller plus loin avec l'utilisation des observations de type image dans la fonction coût, il faut étudier la relation entre les variables d'état du système et les images. Par la suite, nous distinguerons deux analyses différentes des images :

- images vues comme fonctions des variables du modèle ;
- images vues comme concentration d'un traceur passif transporté par la dynamique du système.

Le premier correspond au cas le plus complexe avec une description fortement dépendante du système étudié. Le deuxième est plus indépendant du système étudié dès qu'il s'agit d'un système dynamique. Cette indépendance permet d'étudier le problème d'assimilation d'images dans un contexte particulier et de l'étendre facilement à d'autres contextes. En plus, elle correspond à une bonne approximation de la réalité dans la plupart des cas.

Images analysées comme une fonction des variables du modèle

Telles que décrites au chapitre précédent, les images satellites correspondent à des observations plus ou moins indirectes des variables des modèles atmosphériques ou océanographiques : les radiances pour les canaux infrarouges et la luminosité pour les canaux visibles. Dans l'un ou l'autre cas, les quantités observées peuvent s'exprimer par des fonctions plus ou moins complexes des variables du modèle. L'opérateur d'image synthétique $\mathcal{H}_{\mathcal{X} \rightarrow \mathcal{F}}$ exprime la composition du processus d'observation et la dépendance de la quantité observée aux variables du modèle. Le processus d'observation dans la plupart des cas étant modélisé par le transfert radiatif.

Exemple 1 : observation de l'atmosphère. Les observations de l'atmosphère dans le canal visible sont des mesures de la brillance de la surface des nuages des couches supérieures. La dépendance aux variables du modèle atmosphérique se réduit à la formation des nuages et le processus d'observation peut se réduire à la propagation du rayonnement de la surface du nuage à l'instrument de mesure. L'observation de l'atmosphère dans le canal vapeur d'eau est similaire à l'observation dans le canal visible. La quantité observée dans ce cas est l'intégration de la quantité d'eau dans une épaisseur de l'atmosphère et le processus d'observation correspond au transfert radiatif entre l'atmosphère et l'instrument de mesure. L'information présente en trois dimensions est ainsi ramenée à une information en deux dimensions et **la mesure ne donne aucune information sur l'altitude du phénomène sinon très peu.**

Exemple 2 : observation de la mer. L'observation de la mer par satellite fournit, entre autres, la température de surface et la couleur de surface. Dans le cas de la température de surface, la quantité observée est fonction d'une petite épaisseur de surface de la mer. Dans le cas de la couleur de surface, la quantité observée dépend essentiellement de la concentration du phytoplancton dans la couche supérieure de la mer. Le processus d'observation combine l'absorption par la surface de la mer et la propagation du rayonnement réfléchi vers l'instrument de mesure. L'information ainsi recueillie est limitée à une épaisseur non représentable par les modèles numériques actuels.

Si les images sont considérées sous cette forme, leur assimilation directe peut être sujette à plusieurs problèmes parmi lesquels :

- La non-linéarité des fonctions de dépendance entre les variables du modèle et les quantités observées. Comme nous l'avons déjà mentionné à la sous-section (3.2.3). Les quantités observées par les télémessures résultent en général de processus très complexes faisant intervenir plusieurs variables du modèle. Ces processus ont pour caractéristique principale la non-linéarité. C'est le cas de la microphysique atmosphérique à l'origine des nuages ;
- la non différentiabilité des processus à seuil intervenant dans l'expression des quantités observées et le processus d'observation. La relation entre la quantité observée et les variables du modèle est en général caractérisée par des seuils de même que le processus d'acquisition. Il en est de même des instruments de mesure.

Les deux problèmes sus-cités sont des sources de difficulté pour la différentiation. Si la linéarisation autour d'un état de référence permet de résoudre le problème posé par les non-linéarités au prix d'un surcoût de stockage mémoire pour le code adjoint, les processus à seuil sont tout simplement non différentiables. Ce sont des problèmes auxquels il faut faire face si les télémessures doivent être utilisées de cette façon. Si on peut contourner ces problèmes, l'utilisation directe des images en particulier et des télémessures en général permet d'obtenir des informations sur toutes les variables d'état du modèle influençant la quantité observée ou le processus d'observation. Ceci permet en assimilation de données d'améliorer la qualité de toutes ces variables.

Images analysées comme un traceur passif

En physique expérimentale, il est courant d'ajouter des traceurs passifs à des fluides en mouvement pour étudier la dynamique des systèmes. Les travaux de Flór [Flór and Eames(2002)] sur l'étude de la dynamique des tourbillons en sont une bonne illustration. Dans ces expériences

qui constituent une source de données pour certaines de nos expériences numériques, un traceur passif est ajouté à un fluide et les photographies de l'état du traceur à des moments précis de l'expérience permettent d'étudier la dynamique du phénomène. Cette considération ne se limite pas à la physique expérimentale. En effet, moyennant quelques approximations, certaines images peuvent être considérées comme des observations de traceur passif transporté par la dynamique du système étudié. C'est l'hypothèse faite implicitement dans l'utilisation des images pour la dérivation des pseudo-observations de champs de vitesse pour les modèles atmosphériques. Sur une échelle de temps limitée, cette hypothèse peut être considérée comme réaliste. La figure (4.2) en est une bonne illustration. Les nuages observés sur ces images peuvent être assimilés à un traceur transporté par la dynamique du système atmosphérique au dessus de l'Europe de l'Ouest et de l'atlantique Nord. Sur une échelle de temps raisonnable, le traceur peut être considéré comme passif. Une telle hypothèse peut rapidement être biaisée si on tient compte des variations de luminosité et de la position relative de l'instrument de mesure par rapport à la cible. En excluant les observations de nuit où les canaux visibles ne produisent presque pas de données, les variations légères de luminosité ou d'exposition peuvent être gérées en considérant le niveau d'interprétation structuré de l'image. En effet à ce niveau d'interprétation, ce n'est pas la valeur individuelle de chaque mesure qui est importante mais leur organisation spatiale. Comme l'illustré la figure (4.1), cette structuration peut être identique sur des images provenant de canaux d'observation différents. Ce qui justifie une fois de plus l'importance du niveau d'interprétation des images.

Ainsi formulée, l'utilisation des images comme traceur passif nous permet d'obtenir de l'information sur la dynamique du système. En assimilation de données, elle permettra d'améliorer la qualité de l'information dynamique obtenue avec les sources d'observation classiques. Pour les expériences numériques des sections (4.3) et (4.4), nous utilisons cette formulation.

Opérateurs d'observation L'opérateur d'image synthétique se réduit au transport de l'image initiale par la dynamique du système. Il existe des cas où les images ne peuvent pas être considérées comme des traceurs ou alors, leur dynamique n'obéit pas à un simple transport par la dynamique du système. Dans de tels cas, il faut se ramener à la considération des images comme fonction des variables du modèle ou bien développer des opérateurs d'images synthétiques appropriés. Les travaux de Michel [Michel(2008)] illustrent le développement de tels opérateurs pour les images de tourbillons potentiels.

4.2.4 Régularisation

Les images étant considérées comme observation d'un traceur passif, le problème d'assimilation directe de séquences d'images consiste à déterminer le contrôle optimal qui minimise la fonction coût (équation 4.15). C'est un problème mal posé comme tout problème inverse en général. La régularisation est donc nécessaire pour obtenir une solution unique. La variable de contrôle incluant la vitesse pour les systèmes dynamiques, l'assimilation d'images est affectée par le **problème d'ouverture**⁸ comme tout problème traitant de la dynamique des images. Le problème de régularisation en traitement d'images comme dans presque tous les domaines est un problème difficile. La littérature est abondante sur le sujet mais il n'existe

8. Seule la dynamique suivant la normale au contour peut être déterminé avec précision

pas de solution générale; le problème reste ouvert. Le chapitre (5) y est entièrement consacré; nous y présentons un état de l'art de la régularisation pour les problèmes de traitement d'images. Nous introduisons ensuite une nouvelle approche d'utilisation de l'information de régularisation pour la résolution des problèmes inverses.

4.3 Application à un modèle d'écoulement en eaux peu profondes

4.3.1 Cadre général de l'application

Les expériences de laboratoire constituent en général le meilleur point de départ pour vérifier la faisabilité des techniques nouvellement introduites. Elles permettent d'approcher la complexité du problème réel dans un cadre contrôlable et reproductible. Nous avons appliqué cette méthodologie à l'approche d'assimilation d'images introduite à la section (4.2). Ceci nous permet de tester l'approche d'utilisation de l'information image en assimilation de données dans un cadre suffisamment proche de la réalité.

Contexte

Nous nous sommes basés sur une expérience d'étude de la dérive d'un tourbillon (l'équivalent d'un écoulement cyclonique dans l'atmosphère) sur une table tournante simulant la force de Coriolis. Cette force est une composante importante des écoulements atmosphériques et océanique au delà des tropiques. L'expérience est réalisée à la plate-forme Coriolis⁹ de Grenoble. Le tourbillon est créé en aspirant instantanément un certain volume du fluide à un point donné. Il est ensuite rendu visible par ajout d'un traceur passif (fluorescine). Les photographies de la vue du dessus du tourbillon sont prises à partir d'une caméra fixée sur la plate-forme et constituent la séquence d'images observées. La figure (4.3) montre deux photographies de la table tournante de la plate-forme Coriolis. Voir [Flór and Eames(2002)] pour plus de détails sur la mise en œuvre et les motivations originales de l'expérience de la plate-forme Coriolis.

Modèle

Dans la configuration expérimentale de la plate-forme Coriolis, avec une mince couche de fluide, l'écoulement peut être modélisé par les équations de Saint-Venant, aussi connu comme modèle d'écoulement en eaux peu profondes (shallow-water equations en anglais) :

$$\begin{cases} \partial_t u - (f + \zeta)v + \partial_x B & = -ru + \nu \Delta u, \\ \partial_t v + (f + \zeta)u + \partial_y B & = -rv + \nu \Delta v, \\ \partial_t h + \partial_x(hu) + \partial_y(hv) & = 0. \end{cases} \quad (4.19)$$

Dans ce modèle, la variable d'état est donnée par $\mathbf{x}(t) = (u(t), v(t), h(t))$ où $u = u(t, x, y)$ représente la composante zonale et $v = v(t, x, y)$ la composante méridionale de la vitesse du

9. La plate-forme tournante Coriolis de Grenoble (France) est une plate-forme expérimentale construite en 1960 à la demande d'EDF pour des besoins de modélisation à l'échelle 1/50000 des courants et déplacements de sédiments créés par l'implantation d'une usine marémotrice dans la baie du Mont St Michel. C'est une machine de 14m de diamètre, 300 tonnes qui peut atteindre une vitesse de rotation d'un tour toutes les 20s. Il était ainsi possible de représenter les marées dans la Manche, en tenant compte des effets de la rotation terrestre et de la topographie des fonds marins dans la région. Pour plus d'informations sur la plate-forme Coriolis, voir <http://coriolis.legi.grenoble-inp.fr>



(a) Dispositif pour une simulation de la mer

(b) Dispositif complet avec l'armature supérieure

FIGURE 4.3 – Photographie de la plate-forme Coriolis.

courant et $h = h(t, x, y)$ représente l'élévation de la surface libre. Ce sont des fonctions du temps t et de la variable d'espace physique $\mathbf{x} = (x, y)$. Nous définissons la vorticit  relative par :

$$\zeta = \partial_x v - \partial_y u \quad (4.20)$$

et le potentiel de Bernoulli :

$$B = gh + \frac{1}{2}(u^2 + v^2), \quad (4.21)$$

avec g la constante de gravit . Le param tre de Coriolis sur le plan β cr e par la table tournante est donn e par :

$$f = f_0 + \beta y, \quad (4.22)$$

o  ν repr sente le coefficient de diffusion et r le coefficient de friction au fond.

Dans nos exp riences num riques, et conform ment aux exp riences de la plate-forme Coriolis, les valeurs des diff rents param tres sont fix es ainsi qu'il suit : $r = 0.9 \times 10^{-7} \text{s}^{-1}$, $\nu = 0 \text{m}^2 \text{s}^{-1}$, $f_0 = 0.25 \text{s}^{-1}$, $g = 9.81 \text{m.s}^{-2}$ et $\beta = 0.0406 \text{m}^{-1} \text{s}^{-1}$. Le domaine physique de la simulation num rique est d fini par une boite rectangulaire $\Omega =]0, L[\times]0, H[$ repr santant un sous-domaine carr  de la table tournante de dimensions $L = H = 2.525 \text{m}$. C'est aussi l'espace couvert par la photographie. Le domaine est discr tis  en une grille de taille $N \times N = 128 \times 128$ uniforme de type C d'Arakawa. Un sch ma de diff rences finies est utilis  en espace et un sch ma de Runge-Kutta d'ordre 4 en temps. Le pas de temps consid r  est de 0.01s. Ce qui correspond   14.4s pour l' quivalent dans atmosph re du ph nom ne simul . Le code num rique associ  au mod le (4.19) est emprunt    la biblioth que de codes de l' quipe MOISE¹⁰ ainsi que son adjoint. Ils proviennent des travaux de la th se de Arthur Vidard [Vidard(2001)].

4.3.2 Processus d'assimilation d'images

Le probl me consid r  est celui de la d termination de l'* tat initial* du fluide $\mathbf{x}_0(x, y) = (u, v, h)(0, x, y)$ qui constitue la variable de contr le. Les seules observations consid r es sont

10. Equipe INRIA au sein de laquelle cette th se a  t  pr par e en partie.

les images. L'opérateur $\mathcal{H}_{\mathcal{F} \rightarrow \mathcal{S}}$ de passage du niveau d'interprétation élémentaire au niveau structuré est basé sur les caractéristiques spectrales de la fonction image. Nous utilisons la composition d'un seuillage adapté et la transformation multi-échelle et multi-directionnelle en curvelets. Le seuillage permet de prendre en considération un sous-ensemble restreint des coefficients de l'image dans la frame des curvelets. La fenêtre d'assimilation considérée est de 30s (soit 3000 pas de temps) pour les expériences jumelles et 7.5s (soit 750 pas de temps) pour les expériences avec images Coriolis. Ce qui correspond respectivement à une fenêtre de 12 heures et 3 heures pour l'équivalent du phénomène dans l'atmosphère. Pour indication, la fenêtre d'assimilation de données dans les centres météorologiques opérationnels est de 6 heures actuellement. Les observations images sont prises toutes les 1.25s (soit 125 pas de temps) pour les expériences jumelles et 0.25s (25 pas de temps) pour les expériences réelles. Cette fréquence d'acquisition correspond respectivement à une image chaque 30 minutes et une image chaque 6 minutes pour l'équivalent du phénomène dans l'atmosphère. Nous avons ainsi un ensemble de 24 images pour les expériences jumelles et 30 images pour les expériences avec les images Coriolis. L'image au temps initial est considérée comme champs de traceur au temps initial. Il n'y a aucun contrôle sur le traceur.

L'application du formalisme d'assimilation d'images sur ce cas expérimental conduit à la minimisation d'une fonction coût exprimée comme

$$J(\mathbf{x}_0) = \underbrace{\alpha_o \int_0^T \|\mathcal{H}_{\mathcal{F} \rightarrow \mathcal{S}}[f] - \mathcal{H}_{\mathcal{X} \rightarrow \mathcal{S}}[\mathcal{M}_t(\mathbf{x}_0)]\|_{\mathcal{S}}^2 dt}_{J_o} + \underbrace{\alpha_b \|\mathbf{x}_0 - \mathbf{x}_b\|_{\mathcal{X}}^2}_{J_b}, \quad (4.23)$$

avec \mathcal{M}_t le modèle intégré du temps 0 au temps t . α_o et α_b représentent les paramètres de pondération associés respectivement au terme d'écart aux observations image J_o et au terme d'écart à l'ébauche J_b . Dans la suite, nous explicitons un peu ces termes. Les autres termes de régularisation sont volontairement masqués pour des besoins de lisibilité. Nous y reviendrons au chapitre (5).

Terme d'écart aux observations : opérateurs d'image

Opérateur d'image synthétique Comme nous l'avons mentionné plus haut, une façon d'écrire l'opérateur d'observation $\mathcal{H}_{\mathcal{X} \rightarrow \mathcal{S}}$ associé aux images est de le décomposer en deux parties : un opérateur d'image synthétique $\mathcal{H}_{\mathcal{X} \rightarrow \mathcal{F}}$ (défini de l'espace d'état du système vers l'espace des images interprétées au niveau élémentaire) et un opérateur de structures $\mathcal{H}_{\mathcal{F} \rightarrow \mathcal{S}}$ (défini du niveau élémentaire d'interprétation des images vers le niveau structuré). De cette façon on utilise le même opérateur de structures pour les images observées et pour les sorties du modèle. C'est ainsi que nous avons procédé. L'opérateur d'image synthétique $\mathcal{H}_{\mathcal{X} \rightarrow \mathcal{F}}$ utilisé est obtenu par le modèle d'advection du traceur passif avec le champ de vitesse du système. Cet opérateur est défini par :

$$\mathcal{H}_{\mathcal{X} \rightarrow \mathcal{F}}[u, v, h] = q, \quad (4.24)$$

où $q = q(t, x, y)$ représente la concentration du traceur passif et vérifie l'équation de conservation dans le domaine physique Ω pour les temps $t \geq 0$:

$$\partial_t q + u \partial_x q + v \partial_y q = 0. \quad (4.25)$$

Le champ de vitesse d'advection (u, v) est donné par l'état du modèle et satisfait les équations d'écoulement en eaux peu profondes (4.19). On obtient ainsi une image synthétique au

temps t correspondant à la concentration du traceur $(x, y) \mapsto q(t, x, y)$ comme la fonction de luminosité.

Opérateur de structures : décomposition en curvelets Comme nous l'avons mentionné au paragraphe (4.1.2), une façon simple d'extraire les structures caractéristiques d'une image consiste à utiliser sa représentation dans une base spectrale en utilisant notamment une transformation multi-échelle. Nous avons utilisé la transformation en curvelets de Candès et Donoho [Candes and Donoho(2004), Candes and Donoho(2005), Candes and Donoho(2005), Candes et al.(2006)]. Avec cette transformation, une fonction image f s'exprime par :

$$f = \sum_{j,l,\mathbf{k}} \langle \varphi_{j,l,\mathbf{k}}, f \rangle \varphi_{j,l,\mathbf{k}}, \quad (4.26)$$

$(\varphi_{j,l,\mathbf{k}})_{j,l,\mathbf{k}}$ représentant les fonctions de la frame de curvelets. L'opérateur $\langle \cdot, \cdot \rangle$ représente le produit scalaire usuel de $L^2(\mathbb{R}^2)$. Il existe une formulation et une implémentation linéaire et isométrique de la transformation en curvelets ; ce qui en fait une opération unitaire. Son adjoint requis pour l'assimilation variationnelle de données est exprimé par la transformation inverse (reconstruction). C'est un avantage non négligeable pour l'utilisation de la transformation en curvelets. En vue d'extraire uniquement l'information pertinente des images (idéalement les structures), nous utilisons une fonction de seuillage pour ne retenir que quelques coefficients de l'image transformée $\langle \varphi_{j,l,\mathbf{k}}, f \rangle$. L'opérateur de structures est ainsi défini comme le seuillage de la transformée en curvelets de l'image f :

$$\mathcal{H}_{\mathcal{F} \rightarrow \mathcal{S}}[f] = \mathcal{T}(\text{DCT}(f)), \quad (4.27)$$

\mathcal{T} dénotant l'opérateur de seuillage. Il est appliqué sur les coefficients de la transformée en curvelets calculés par l'algorithme de transformation en curvelets discrets DCT¹¹ [Candes et al.(2006)]. Cet algorithme est inclus dans la bibliothèque **CurveLab** qui est disponible, pour des besoins de recherche, sur le site web www.curvelet.org.

Il existe plusieurs fonctions de seuillage ; la plus simple étant la fonction de seuillage brut (3.3) que nous avons présentée au paragraphe (3.3.5). Le seuillage brut appliqué aux coefficients des curvelets est défini par :

$$\tau(c_{j,l,\mathbf{k}}) = \begin{cases} c_{j,l,\mathbf{k}} & \text{si } |c_{j,l,\mathbf{k}}| \geq \sigma, \\ 0 & \text{si } |c_{j,l,\mathbf{k}}| < \sigma, \end{cases} \quad (4.28)$$

où $c_{j,l,\mathbf{k}}$ est le coefficient de la fonction curvelet discrète à l'échelle d'indice j , orientation d'indice l , position spatiale $\mathbf{k} = (k_1, k_2)$ et σ représente le seuil unique pour tous les coefficients. Pour les besoins d'expérimentation, nous avons introduit la notion de seuillage échelle par échelle qui est mieux adapté à notre avis aux décompositions multi-échelle. Le seuillage échelle par échelle se définit comme une stratification à la seule différence qu'on ne considère pas les sous-ensembles du domaine de définition du signal mais les différentes échelles de la décomposition. La fonction de seuillage échelle par échelle se définit par :

$$\tau(c_{j,l,\mathbf{k}}) = \begin{cases} c_{j,l,\mathbf{k}} & \text{si } |c_{j,l,\mathbf{k}}| \geq \sigma_j, \\ 0 & \text{si } |c_{j,l,\mathbf{k}}| < \sigma_j, \end{cases} \quad (4.29)$$

11. Discrete Curvelet Transform

où σ_j représente le seuil à l'échelle d'indice j .

L'espace des images structurées \mathcal{S} est défini comme l'espace des coefficients seuillés de la transformée en curvelets. Comme nous l'avons mentionné au paragraphe (3.3.5), la définition d'un seuil adapté peut se révéler fastidieux. Pour contourner ce problème, nous faisons des seuillages au ratio : le seuil est fixé par un ratio qui définit le rapport entre le nombre de coefficients utiles et le nombre total de coefficients. Nos expériences sont basées sur un ratio de 10%. Le lecteur pourra se référer à [Antoniadis(2007)] pour les fonctions de seuillage classiques. Pour la réduction des effets de Gibbs, on peut faire usage d'une fonction de seuillage basé sur une minimisation de la variation totale [Durand and Froment(2003)]. Dans le cas du seuillage pour des transformations en curvelets, [Ma et al.(2008)] constitue une bonne référence.

L'opérateur de structure est défini comme la composition d'un opérateur de seuillage (équation 4.28 ou 4.29) et de l'opérateur d'image synthétique (équation 4.24) :

$$\mathcal{H}_{\mathcal{X} \rightarrow \mathcal{S}} = \mathcal{H}_{\mathcal{F} \rightarrow \mathcal{S}} \circ \mathcal{H}_{\mathcal{X} \rightarrow \mathcal{F}} = \mathcal{T} \circ \text{DCT}. \quad (4.30)$$

Régularisation et préconditionnement

Régularisation Deux principaux termes de régularisation sont utilisés : la pénalisation de l'écart à l'ébauche afin que le problème soit bien posé, et la régularisation par diffusion généralisée pour faire face au problème d'ouverture. Le terme d'écart à l'ébauche est donné par :

$$J_b(\mathbf{x}_0) = \|\mathbf{x}_0 - \mathbf{x}_b\|_{\mathcal{X}}^2, \quad (4.31)$$

où \mathbf{x}_b représente l'ébauche de la variable de contrôle et la norme $\|\cdot\|_{\mathcal{X}}$ est définie par une matrice diagonale d'adimensionalisation des composantes de la variable de contrôle. Cette adimensionalisation est très importante pour prendre en compte les différences d'ordre de grandeur entre les différentes composantes de la variable de contrôle. Cet opérateur est également utilisé pour le préconditionnement [Vidard et al.(2004)] du problème de la minimisation. Dans nos expériences, l'ébauche est assignée à la valeur constante $\mathbf{x}_b = (u_b, v_b, h_b) = (0, 0, h_m)$ où h_m représente l'élévation moyenne de la surface libre du fluide. Il est important de préciser que dans les applications réelles, l'ébauche provient des analyses précédentes. Une de nos pré-occupations actuelles consiste à utiliser des champs de vitesse provenant de l'estimation par flot optique comme ébauche.

Pour la régularisation du champ de vitesse, nous utilisons une approche basée sur la diffusion généralisée que nous avons définie. La version utilisée pour les expériences présentées plus loin dans ce chapitre est présentée dans [Souopgui et al.(2009)]. Elle constitue une première ébauche des travaux qui sont présentés au chapitre (5). Dans le même chapitre, nous présentons un état de l'art de la régularisation pour des problèmes de traitement d'images ainsi qu'une étude comparative avec la nouvelle approche sur le problème d'estimation du mouvement. Certaines des approches de la littérature ont été testées pour la régularisation des champs de vitesse en assimilation d'images mais aucune n'a donné de résultats comparables à la régularisation par diffusion généralisée.

Préconditionnement Très souvent, le préconditionnement utilisé pour les problèmes d'assimilation de données est basé sur une simplification qui réduit l'opérateur de préconditionnement à la matrice de covariance d'erreurs d'ébauche (voir la section 2.4.4) définissant la norme

utilisée pour l'écart à l'ébauche :

$$\|\mathbf{x}\|_{\mathcal{X}} = \mathbf{x}^T \mathbf{B}^{-1} \mathbf{x}. \quad (4.32)$$

Ce préconditionnement permet de définir le changement de variable :

$$\mathbf{x}_0' = \mathbf{B}^{-1/2} \mathbf{x}_0. \quad (4.33)$$

Dans nos expériences, nous avons combiné cet opérateur avec un opérateur de préconditionnement dérivé de la régularisation par diffusion généralisée. Le principe de ce préconditionnement ainsi que celui de la régularisation par diffusion sont présentés dans [Souopgui et al.(2009)] et expliqués en détails au chapitre (5).

Comme indiqué au paragraphe (4.2.3), l'opérateur d'observation image utilisé dans nos expériences ne permet d'obtenir de l'information que sur la dynamique (champs de vitesse du modèle). Pour répercuter l'information sur la variable h (dénivellation de la surface libre), nous utilisons un opérateur de balance basé sur l'*inversion*¹² de l'équilibre géostrophique :

$$\begin{aligned} u &= -\frac{g}{f} \frac{\partial h}{\partial y}, \\ v &= \frac{g}{f} \frac{\partial h}{\partial x}. \end{aligned} \quad (4.34)$$

Optimisation

La fonction coût J est minimisée successivement sur une suite de sous fenêtres de taille croissante (chaque sous fenêtre incluant la précédente) suivant l'approche incrémentale en temps décrite dans [Luong et al.(1998)]. Cette approche permet d'éviter la stagnation dans les minima locaux qui arrive quand la fonction coût est non convexe. Ce qui est le cas avec les images et le problème d'ouverture. Sur chaque sous fenêtre, l'ébauche correspond au résultat obtenu dans la sous fenêtre précédente avec l'ébauche par défaut pour la première sous fenêtre. L'algorithme utilisé pour la minimisation dans chaque sous fenêtre est l'algorithme MIQN3 de la bibliothèque LIBOPT [Gilbert and Jonsson(2007)]. C'est un algorithme de type quasi Newton (voir la section B.3 pour la description de cette classe d'algorithme)

4.4 Résultats des expérimentations numériques

Comme pour toute nouvelle méthode, nous avons procédé dans un premier temps à des expériences jumelles afin de valider l'approche proposée et sa faisabilité. Ensuite, nous sommes passé à des expériences avec images réelles. Dans les deux paragraphes suivants (expériences jumelles et expériences avec images réelles), nous utilisons les équivalents temps dans l'atmosphère du phénomène étudié. Les graphiques sur le domaine physique sont restreints à la sous-région contenant le vortex.

4.4.1 Expériences jumelles

En expériences jumelles, tout est connu : plus précisément la variable de contrôle recherchée est connue avec exactitude, nous l'appelons état initial vrai. Il correspond dans nos expériences

12. la détermination de la dénivellation de la surface libre à partir des composantes de la vitesse horizontale est un problème surdéterminé. Nous utilisons une approche de moindres carrés avec une résolution itérative par gradient conjugué.

à un tourbillon isolé soumis à la force de Coriolis. Les observations images sont générées en faisant tourner le modèle avec l'état initial vrai et l'opérateur d'image synthétique. Le modèle utilisé est le modèle d'écoulement en eaux peu profondes (équation 4.19). La configuration du modèle pour les expériences jumelles est calquée de la description des expériences de Flór [Flór and Eames(2002)]. Le vortex est initialisé à partir du profil radial de vitesse donné par la fonction de Whittaker :

$$v(r) = \frac{rv_m}{R_m} \exp \left[-\frac{1}{2} \left(\frac{r}{R_m} \right)^2 \right], \quad (4.35)$$

exprimée en coordonnées polaires. R_m et v_m représentent respectivement le rayon du vortex et l'intensité maximale de la vitesse azimutale. Dans nos expériences, $R_m = 0.129\text{m}$ et $v_m = 0.02\text{m.s}^{-1}$. L'élévation initiale de la surface libre du fluide est déduite de la vitesse par *inversion* de l'équilibre géostrophique (4.34). La hauteur moyenne du fluide est donnée $h_0 = 0.36\text{m}$. La concentration initiale du traceur est donnée par des petits pavés disposés dans la région contenant le tourbillon. Le modèle est alors intégré jusqu'à la stabilisation. Ce qui permet d'éliminer les bruits et les instabilités de l'initialisation imparfaite. La date de stabilisation est ensuite considérée comme date de référence et début de la fenêtre d'assimilation de données. Les images sont générées à partir de la date de référence (la fonction de luminance étant définie comme la concentration du traceur passif). Des conditions aux frontières ouvertes permettent d'éliminer les ondes afin de stabiliser l'ensemble. Des conditions aux frontières différentes sont utilisées pour le processus d'assimilation. Ce changement de conditions aux frontières représente une source d'erreur modèle. Pour l'assimilation, nous utilisons des conditions de Neumann homogènes sur la variable h et des frontières rigides sur la variable \mathbf{v} . Les images générées sont ensuite assimilées afin de déterminer l'état initial ayant permis de les générer. Les figures suivantes permettent d'analyser les résultats obtenus.

Comparaison de l'évolution de l'état analysé et de l'état vrai

La figure (4.4) montre l'évolution comparée du champ de vitesse vrai et du champ de vitesse analysé. La colonne (4.4a) montre l'état vrai et la colonne (4.4b) montre l'état analysé. Nous nous sommes limités au champ de vitesse et l'avons superposé à l'image (concentration du traceur). La ligne inférieure montre l'état au temps 0 (début de la fenêtre d'assimilation de données). La ligne du milieu montre l'état à la date 12h00 (fin de la fenêtre d'assimilation de données). La ligne supérieure montre l'état à la date 24h00 (prévision à la date $T \times 2$ où T correspond à la largeur de la fenêtre d'assimilation de données). La figure montre une bonne reconstruction de la dynamique du système.

La figure (4.5) montre l'évolution comparée des images avec l'état vrai et l'état analysé. La colonne (4.5a) montre les vraies images et la colonne (4.5b) montre les images calculées à partir de l'état analysé. L'image au temps initial $t = 0$ est identique dans les deux cas étant donné qu'on ne contrôle pas l'image. La différence se trouve dans les images aux dates $t > 0$ qui résultent de l'advection de l'image au temps $t = 0$ par la dynamique du système. La ligne inférieure montre l'image au temps 0 (début de la fenêtre d'assimilation de données). La ligne du milieu montre l'image à la date 12h00 (fin de la fenêtre d'assimilation de données) et la ligne supérieure montre l'image à la date 24h00 (prévision à la date $T \times 2$ où T correspond à la largeur de la fenêtre d'assimilation de données). Les différences ne sont pas significatives, on peut en déduire que l'assimilation directe de la séquence d'images a permis de déterminer le champ de vitesse qui transporte assez fidèlement le traceur passif.

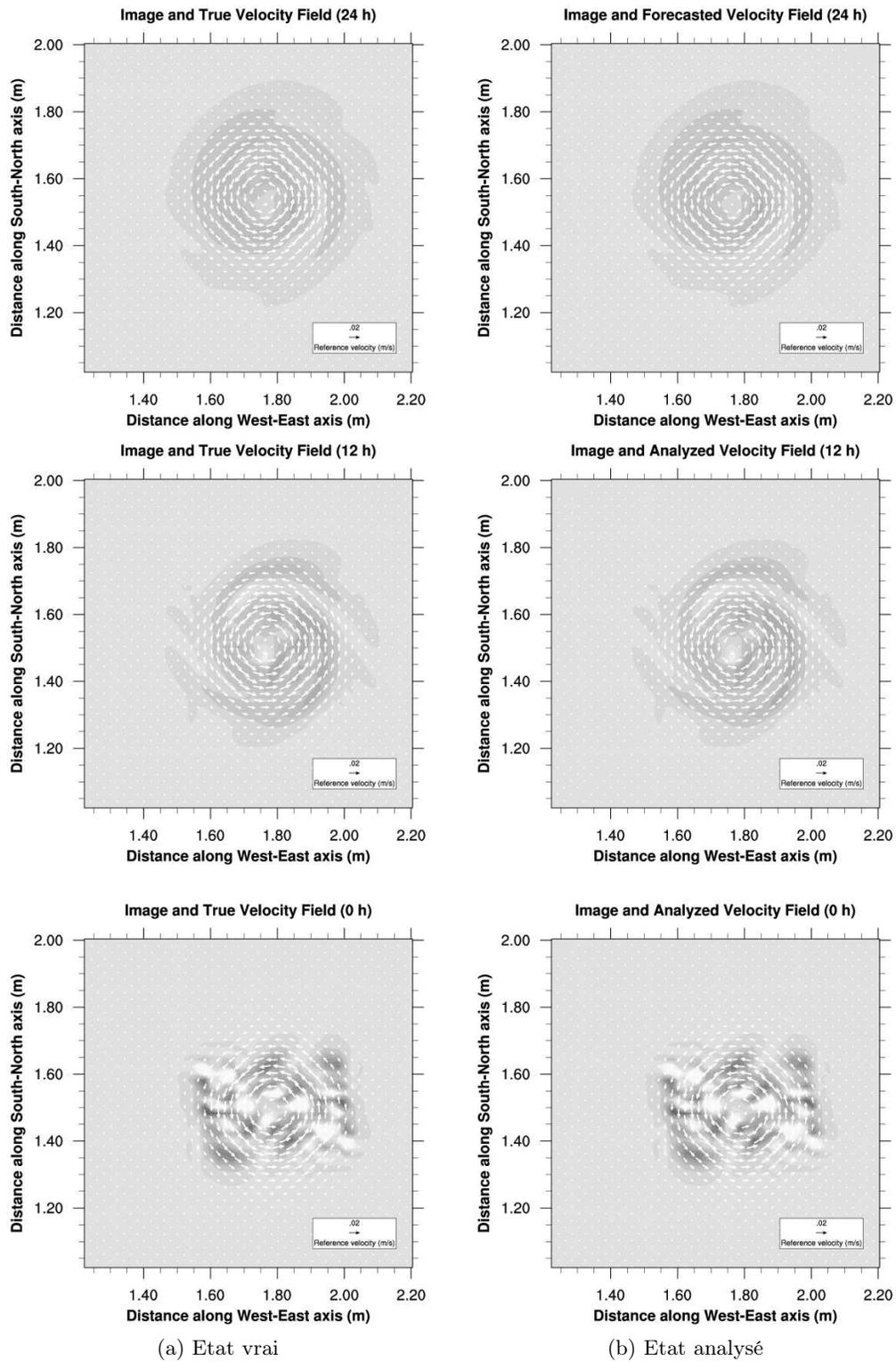


FIGURE 4.4 – Comparaison du champ de vitesse vrai et du champ de vitesse analysé, zoom sur la région contenant le tourbillon. Les champs de vitesse sont superposés à l'image correspondante au temps initial (00h), à la fin de la fenêtre d'assimilation de données (12h) et à une date de prévision (24h).

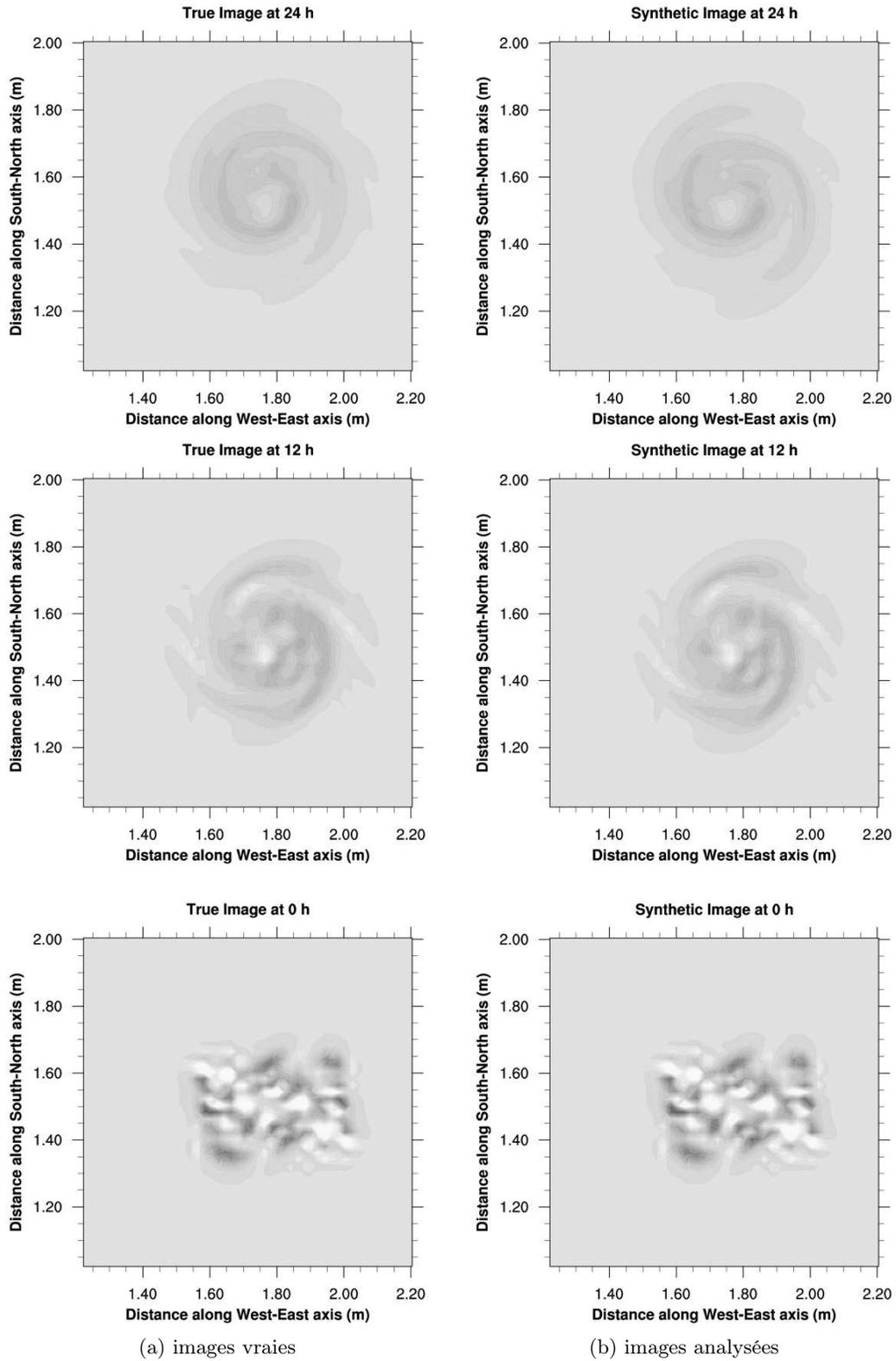


FIGURE 4.5 – Comparaison de la séquence d'images vraies et de la séquence d'images analysées aux dates $0h$, $12h$ et $24h$.

Courbes d'erreurs

Ces courbes permettent de faire une analyse quantitative des résultats obtenus. La figure (4.6) montre l'évolution de l'erreur moyenne normalisée (Normalised Root Mean Square Errors). La sous-figure (4.6a) montre l'erreur $RMS_{(u,v)}$ sur le champ de vitesse analysé et la sous-figure (4.6b) montre l'erreur $RMS_{(u,v)}$ sur l'élévation de la surface libre. Les erreurs sont calculées suivant les formules :

$$RMS_{(u,v)} = \frac{\left(\sum_{i,j=1}^N \delta u^2(i,j) + \delta v^2(i,j) \right)^{1/2}}{\left(\sum_{i,j=1}^N u_t^2(i,j) + v_t^2(i,j) \right)^{1/2}}, \quad (4.36)$$

$$RMS_h = \frac{\left(\sum_{i,j=1}^N \delta h^2(i,j) \right)^{1/2}}{\left(\sum_{i,j=1}^N (h_t(i,j) - h_0)^2 \right)^{1/2}}. \quad (4.37)$$

Avec $\delta u = u_a - u_t$, l'écart entre l'état analysé u_a et l'état vrai u_t . Chacune des sous-figures est coupées en 2 régions délimitées par un trait vertical, la première région correspond à la fenêtre d'assimilation et la deuxième région correspond à la prévision. L'aspect en U de la courbe d'erreur sur le champ de vitesse qui est la variable indirectement observée via les images est typique des courbes d'erreurs en assimilation de données ; elle est synonyme d'erreur modèle. Quand à l'erreur sur l'élévation de la surface libre, elle est très forte au temps initial. Ceci peut s'interpréter par l'absence d'information sur cette variable mais aussi de l'effet réduit de l'opérateur de balance par *inversion* de l'équilibre géostrophique (elle justifie la nécessité de laisser tourner le modèle jusqu'à stabilisation avant de l'utiliser pour des expériences jumelles). Elle décroît ensuite très vite grâce à l'équilibre induite par le modèle. Cependant cet équilibre est fait au prix d'une perte d'énergie en vitesse. La décroissance des courbes d'erreur au début de la fenêtre d'assimilation de données est le symptôme d'une matrice de covariance d'erreur d'ébauche imparfaite et non un problème lié à l'assimilation d'images.

Comparaison des images analysées et des images observées.

La figure (4.7) montre la courbe de l'écart aux observations en fonction du temps. Ce terme d'écart aux observations est calculé pour chaque image dans l'espace des pixels. Son intégrale sur toute la fenêtre d'assimilation correspond à la fonction minimisée lorsqu'on utilise directement l'espace des pixels pour l'écart aux observations. En abscisses, nous avons les pas de temps (pour rappel, l'assimilation utilise une image observée tous les 125 pas de temps) et en ordonnées l'écart aux observations normalisé calculé comme suit :

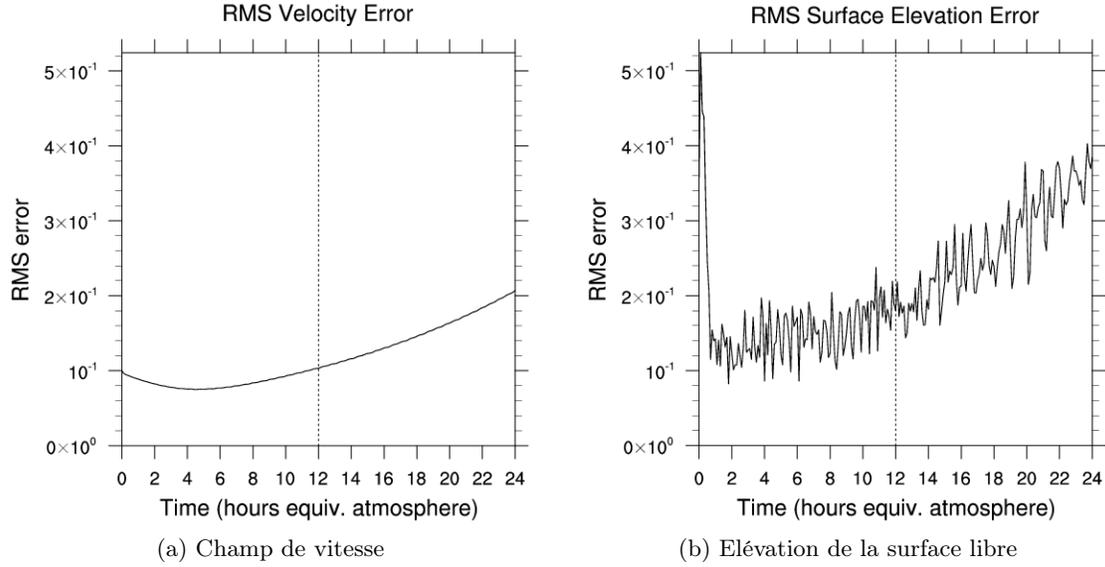


FIGURE 4.6 – Courbe d'évolution des erreurs normalisées sur la fenêtre d'assimilation de données et une fenêtre de prévision.

$$\delta \mathbf{y}(t) = \frac{\int_{\Omega} (q(t, \mathbf{x}) - f(t, \mathbf{x})) d\mathbf{x}}{\int_{\Omega} (f(t, \mathbf{x})) d\mathbf{x}}. \quad (4.38)$$

Où $\mathbf{x} = (x, y) \in \Omega$ représente la variable d'espace physique, $q(t, \mathbf{x})$ l'image donnée par l'opérateur d'observation au temps t et $f(t, \mathbf{x})$ l'image observée au temps t .

Cette donnée permet de quantifier la moyenne de l'écart aux observations par pixel sur le domaine physique. La courbe de l'écart aux observations calculé à partir de l'ébauche donne l'écart de toutes les images observées dans la fenêtre d'assimilation par rapport à l'image au temps 0. Tracée sur le même graphique que la courbe après assimilation de données, elle permet de quantifier l'effet de la minimisation. On remarque que les valeurs sont très faibles et que la fonction d'écart varie très peu au delà de 1000 pas de temps. Les faibles valeurs sont dues au fait que la fonction image est assez homogène. Son support est localisé dans la région contenant le tourbillon. La figure nous indique que l'assimilation de données améliore de façon significative l'écart aux observations.

Comparaison de l'état analysé et de l'état vrai

La figure (4.8) montre les graphiques de la variable de contrôle vraie (état initial vrai) et de la variable de contrôle analysée (état initial calculé par assimilation d'images). Il est important de rappeler qu'on commence l'expérience par une ébauche nulle en vitesse.

On constate que le tourbillon est reconstruit et bien localisé. On peut constater que l'élévation de la surface libre est légèrement sous-estimée, ce qui est en accord avec la courbe d'erreur (4.6b). Cette sous-estimation de l'élévation de la surface libre est due au fait que les observations (images) ne fournissent aucune information sur cette variable. L'étude du champ

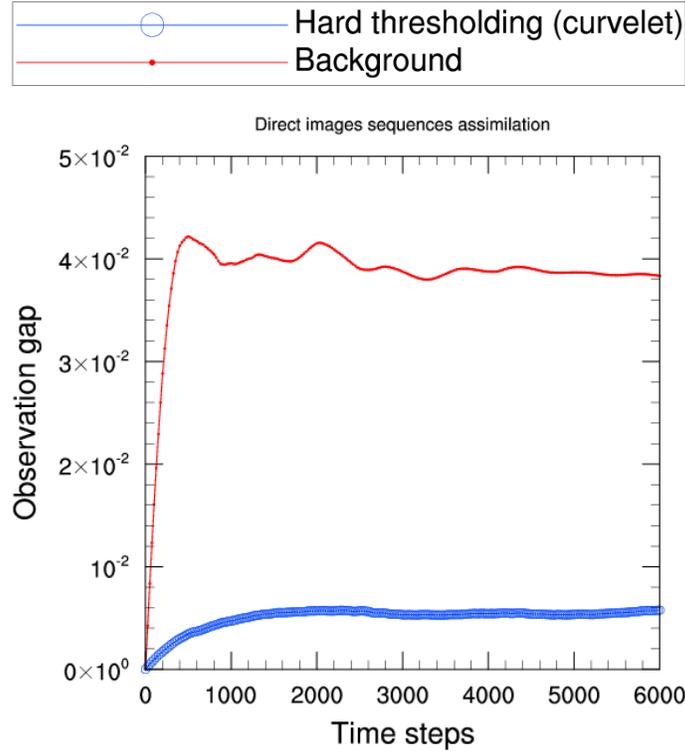


FIGURE 4.7 – Assimilation directe d’images, terme d’écart aux observations par image. La première partie du graphique ($[0, 3000]$ pas de temps) correspond à l’assimilation de données et le reste ($[3000, 6000]$) correspond à la prévision.

de vitesse analysé demande quant à elle des outils plus avancés. Nous avons utilisé l’erreur angulaire de Barron [Barron et al.(1994)] $bae_{\mathbf{u}}$ et la carte d’erreur relative $\delta\|\mathbf{u}\|$ sur le champ de vitesse. L’erreur angulaire de Barron s’exprime par :

$$bae_{\mathbf{u}}(i, j) = \arccos \left(\left\langle \frac{\mathbf{u}_a(i, j)}{\|\mathbf{u}_a(i, j)\|}, \frac{\mathbf{u}_t(i, j)}{\|\mathbf{u}_t(i, j)\|} \right\rangle \right), \quad (4.39)$$

avec $\mathbf{u} = (u, v)$. Plus de détails sur l’erreur angulaire de Barron sont donnés dans le chapitre (5) où nous proposons une formulation corrigeant certaines lacunes de la formulation initiale. L’erreur relative $\delta\|\mathbf{u}\|$ est définie par :

$$\delta\|\mathbf{u}\|(i, j) = \frac{\left| \|\mathbf{u}_a(i, j)\| - \|\mathbf{u}_t(i, j)\| \right|}{\max_{i, j} \|\mathbf{u}_t(i, j)\|}. \quad (4.40)$$

La figure (4.9) montre les diagnostics d’erreur sur le champ de vitesse estimé. La sous-figure (4.9a) montre une superposition des iso contours de l’erreur angulaire de Barron $bae_{\mathbf{u}}$ (entre le champ analysé et le champ vrai) et de la norme du champ de vitesse vrai. Ceci permet d’étudier la répartition de l’erreur angulaire. On note une faible erreur dans la région où le champ de vitesse est intense (région où est localisé le tourbillon). Elle devient importante et aléatoire dans la région où le champ de vitesse est très faible. La sous-figure (4.9b) montre la carte d’erreur relative. Elle a des valeurs faibles (moins de 7%) et est en accord avec l’erreur

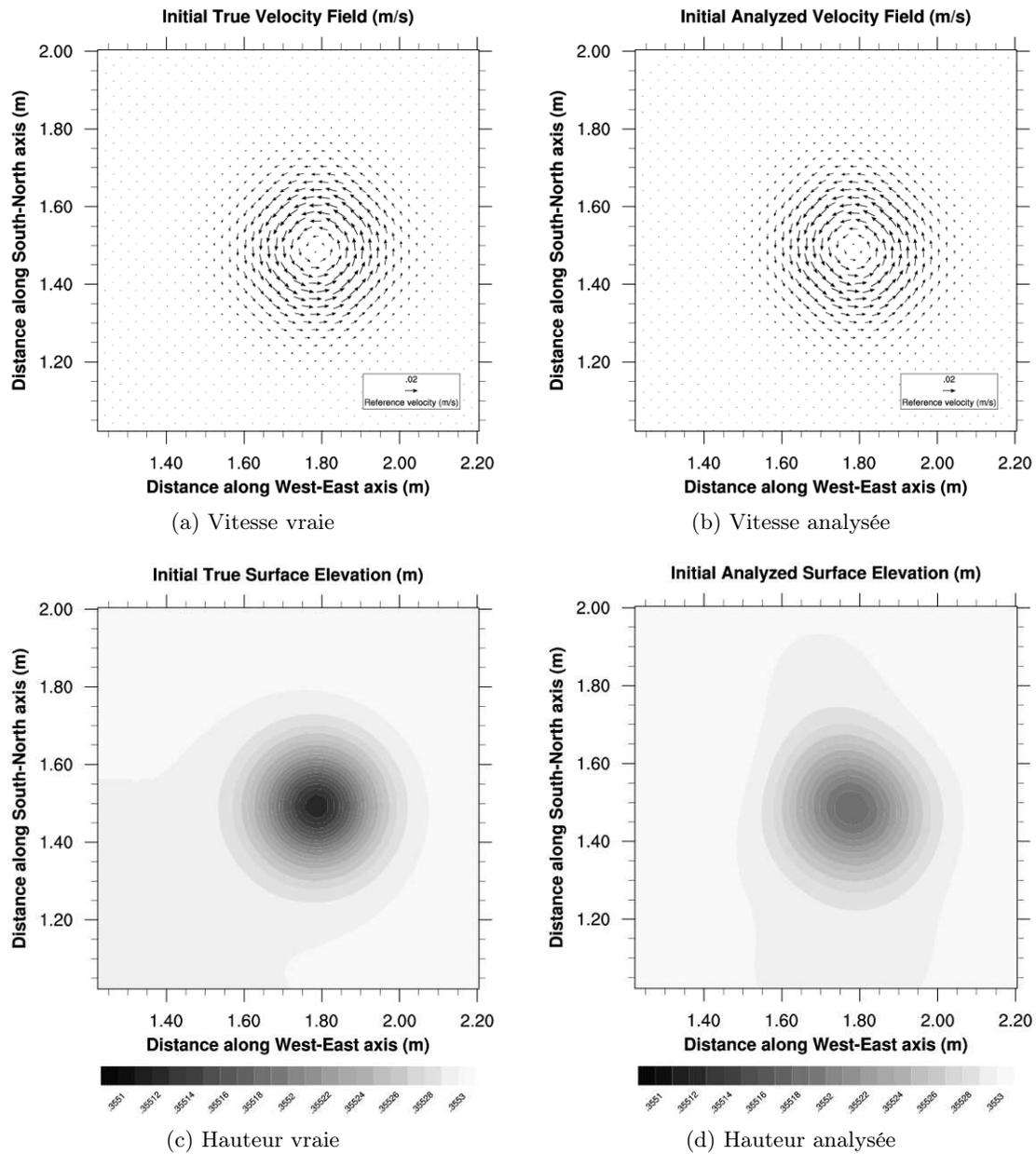


FIGURE 4.8 – Comparaison de l'état analysé et de l'état vrai

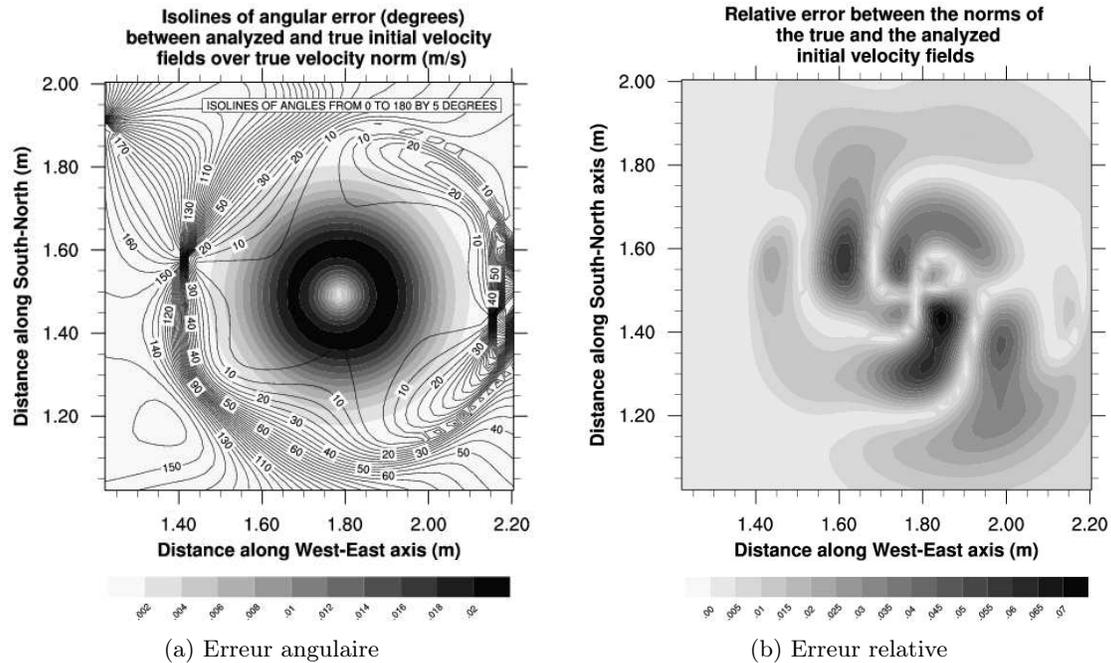


FIGURE 4.9 – Diagnostics sur le champ de vitesse analysé. (a) Iso-contours d’erreur angulaire superposés à la norme du champ de vitesse vrai, (b) carte d’erreur relative

angulaire de Barron dans la région où le champ de vitesse est intense. La structuration de l’erreur relative se justifie par la distribution initiale du champ de traceurs (c’est la seule source d’information).

Remarque 4.4.1 *Pour l’instant, nous n’avons pas réussi à observer de différences fondamentales au niveau de l’opérateur d’observation utilisé. Le seuillage dur et le seuillage échelle par échelle sur les curvelets produisent des résultats très proches. La principale différence réside au niveau du nombre d’itérations d’optimisation et donc du temps de calcul nécessaire pour converger. Le seuillage brut requiert beaucoup plus d’itérations que le seuillage échelle par échelle mais on parvient à des résultats presque identiques. Quand aux images sous forme de pixels, elles requièrent encore plus d’itérations que le seuillage brut sur les coefficients de curvelets. Les résultats sont également comparables avec un nombre suffisant d’itérations. Cependant, dès que la fenêtre d’assimilation de données devient importante (au delà de 600 pas de temps dans nos expériences), la minimisation stagne¹³ très vite avec un écart aux observations calculé dans l’espace des pixels. Des études plus approfondies sont nécessaires pour tirer des conclusions significatives sur les phénomènes observés et leurs causes. Nous pensons que la pseudo-équivalence qu’on peut observer avec les différents opérateurs d’observation sur de courtes fenêtres d’assimilation est due aux images trop lisses. En effet, avec des images trop lisses, l’effet du seuillage est assez réduit.*

En conclusion, l’assimilation directe de séquences d’images est en mesure de reconstruire le champ de vitesse initiale d’un modèle dynamique qui permet de transporter un champ de

¹³. il devient difficile pour l’algorithme de trouver une bonne direction de descente et le pas de descente associé

traceur dont la concentration est utilisée comme image. Nous l'avons montré sur un modèle d'écoulement en eaux peu profondes avec un tourbillon. Le tourbillon est notamment reconstruit au bon endroit et le champ de vitesse est reconstruit avec les bons ordres de grandeurs. Ceci montre la faisabilité de la technique.

4.4.2 Expériences avec des images réelles

Après avoir validé la méthode sur des expériences jumelles nous avons réalisé des expériences avec des images photographiques issues d'une expérimentation de la plate-forme Coriolis similaire à celle décrite dans [Flór and Eames(2002)]. Nous avons utilisé la séquence de photographies comme images observées et nous y avons appliqué l'assimilation directe de séquences d'images décrite dans ce chapitre. Le modèle utilisé est le même que celui utilisé pour les expériences jumelles. L'expertise des physiciens permet de savoir que l'ordre de grandeur de la vitesse est plus élevé que celle que nous avons utilisée dans les expériences jumelles. Très peu de paramètres physiques de l'expérience sont connus, en l'occurrence nous n'avons aucune idée du vrai champ de vitesse. Ceci entraîne une erreur modèle qui peut être très importante et que nous ne pouvons pas quantifier. A cause de l'intensité trop élevée de la vitesse dans l'expérience Coriolis, la fréquence d'acquisition nécessaire des images est beaucoup plus élevée que dans le cas des expériences jumelles (une image tous les 25 pas de temps) afin d'éviter les phénomènes d'aliasing¹⁴. Elle correspond à une fréquence d'une image toutes les 6 minutes dans l'atmosphère, ce qui correspond à peu près à la fréquence d'acquisition des satellites METEOSAT seconde génération sur sa zone nominale de couverture divisé par deux : c'est le mode de fonctionnement qui est prévu en cas de phénomène extrême tel qu'un ouragan. Les expériences sur les tourbillons correspondent parfaitement à ce contexte.

Expérience avec images en pixels

L'opérateur d'observation correspond à l'identité, les images sont utilisées à l'état *brut* (les redimensionnements ont introduit un peu de lissage). Le but de l'expérience est d'étudier le comportement du terme d'écart aux observations dans l'espace des pixels. La figure 4.10 montre la variable de contrôle estimée à différentes dates :

- date 0, début de la fenêtre d'assimilation, c'est la variable de contrôle déterminée par assimilation ;
- date T (300 pas de temps), fin de la fenêtre d'assimilation ;
- date $2T$, date deux fois supérieure à la taille de la fenêtre d'assimilation : c'est la prévision.

La colonne de gauche (sous-figure (4.10a)) montre la superposition du champ de vitesse obtenu à l'image réelle et celle de droite (sous-figure (4.10b)) l'élévation de la surface libre du fluide. La fenêtre est limitée à 300 pas de temps, ce qui correspond à 1,2 heures en équivalent temps dans l'atmosphère et un ensemble de 12 images au maximum (il est possible de ne pas prendre toute les images en compte dans le processus d'assimilation de données). Pour ce qui est de la condition initiale estimée, le tourbillon principal¹⁵ est reconstruit au bon endroit. L'ordre de grandeur des champs de vitesse correspond à celui qu'attendent les physiciens mais les orientations des vecteurs de vitesse ne sont pas bonnes dans toutes les régions du

14. C'est un phénomène qui rend des signaux différents non distinguable après un certain échantillonnage. Dans le cas présent, le mouvement de rotation du tourbillon pourrait être masqué.

15. Il y a un tourbillon secondaire créé par l'onde de Rossby.

tourbillon (voir la figure (4.11) qui montre le champ de vitesse estimé avec plus de détails). Le champ de vitesse montre un tourbillon avec des vitesses plus intenses sur le côté ouest (la gauche) alors qu'on s'attendrait à un phénomène inversé notamment des vitesses plus intenses sur le côté est (droite). La dynamique du modèle permet de corriger ces incohérences : on retrouve à la date $2T$ une plus grande cohérence dans l'orientation des vecteurs du champ de vitesse avec une légère perte d'énergie. Nous n'avons pas trouvé de réglages de paramètres pour prendre en compte une fenêtre plus grande. En plus, il est nécessaire de commencer par une première fenêtre d'assimilation prenant en compte une seule image (en plus de l'image au temps 0). Autrement dit, seul le mouvement de translation correspondant à la dérive du tourbillon est reconstruit. L'estimation du champ de vitesse est essentiellement faite dans cette première fenêtre, l'agrandissement de la fenêtre et les images suivantes n'apportent pas une grande correction. L'algorithme de minimisation peine à trouver une bonne direction et un bon pas de descente. Ceci est probablement dû à la non convexité de la fonction coût et à des minima locaux. On note également une grande dépendance à la matrice de covariance d'erreurs d'ébauche (qui est réduite dans nos expériences à un opérateur d'adimensionalisation)

Expérience avec seuillage brut sur les coefficients de curvelets

L'opérateur d'observation correspond à une composition de la transformation en curvelets et d'un seuillage brut. La figure (4.12) montre la variable de contrôle estimée à différentes dates :

- date 0, début de la fenêtre d'assimilation, c'est la variable de contrôle déterminée par assimilation ;
- date T (600 pas de temps), fin de la fenêtre d'assimilation ;
- date $2T$, date deux fois supérieure à la taille de la fenêtre d'assimilation : c'est la prévision.

Le tourbillon principal et le tourbillon secondaire sont reconstruits au bon endroit mais les ordres de grandeur sont sous-estimés. La vitesse maximale est de 2cm/s . La sous-estimation de l'ordre de grandeur de l'intensité des vecteurs du champs de vitesse est probablement due au lissage des discontinuités dans l'image. Rappelons que l'information de la dynamique est localisée au niveau des discontinuités. Les vecteurs du champ de vitesse sont tracés à une échelle appropriée pour mieux discerner les détails du tourbillon reconstruit. Les vecteurs du champ de vitesse ont une bonne orientation dans presque toutes les régions du domaine physique (voir la figure (4.13) pour les détails sur le champ de vitesse). Le champ de vitesse montre un tourbillon avec des vitesses plus intenses sur le côté est (la droite) comme on s'y attendait. La fenêtre d'assimilation peut être étendu jusqu'à 600 pas de temps, ce qui correspond à 2.4 heures pour le phénomène équivalent dans l'atmosphère (voir la figure 4.12) ce qui permet de prendre en compte jusqu'à 24 images. Nous n'avons pas trouvé de réglages pour aller au delà. Le nombre d'itérations requis pour converger est nettement inférieur à celui requis dans le cas des images en pixels. De plus, il est possible de commencer avec une première fenêtre de plus d'une image (en plus de l'image au temps 0). On note également une grande dépendance à la matrice de covariance d'erreurs d'ébauche.

La figure (4.14) montre la variable de contrôle estimée tracée à la même échelle qu'avec les images brutes (ce qui correspond à l'ordre de grandeur estimé par les physiciens). On note une grande dissymétrie entre l'est et l'ouest du tourbillon. La dynamique du modèle permet de corriger ce défaut au prix d'une perte d'énergie.

Expérience avec seuillage échelle par échelle sur les coefficients de curvelets

L'opérateur d'observation correspond à une composition de la transformation en curvelets et d'un seuillage échelle par échelle (voir l'équation (4.29)). La figure (4.15) montre la variable de contrôle estimée à différentes dates :

- date 0, début de la fenêtre d'assimilation, c'est la variable de contrôle déterminée par assimilation ;
- date T (600 pas de temps), fin de la fenêtre d'assimilation ;
- date $2T$, date deux fois supérieure à la taille de la fenêtre d'assimilation : c'est la prévision.

Le tourbillon principal et le tourbillon secondaire sont reconstruits au bon endroit avec de bons ordres de grandeur et de bonnes orientations sur tout le domaine. La figure (4.16) permet de voir le niveau de détails sur le champ de vitesse reconstruit. Le tourbillon a des vitesses légèrement plus intenses sur le côté est (la droite) comme on s'y attendait. La fenêtre d'assimilation peut être étendue au delà de 600 pas de temps, soit plus de 2.4 heures pour le phénomène équivalent dans l'atmosphère. Le nombre d'itérations requis pour converger est nettement inférieur à celui requis dans le cas des deux cas précédents (pixel et seuillage brut sur les coefficients des curvelets). De même que dans le cas du seuillage brut, il est possible de commencer avec une première fenêtre de plus d'une image (en plus de l'image au temps 0). On peut également prendre une fréquence d'acquisition d'images plus faible, notamment une image toutes les 75 pas de temps (l'écart entre deux images peut ainsi être supérieur à l'écart en temps entre deux images successives des satellites METEOSAT de seconde génération). La dépendance à la matrice de covariance d'erreurs d'ébauche est moins importante que dans les deux cas précédents. Pour l'instant, nous n'avons pas toutes les explications à ce phénomène, il est encore sous investigation. Cependant nous pensons que ce comportement est le résultat de la considération de l'image à un niveau d'interprétation structuré. Si cette hypothèse se vérifie, elle confirmera l'indépendance du niveau d'interprétation structuré de la quantité physique observée et son adéquation pour l'étude de la dynamique dans les images. Rappelons que nous avons introduit le niveau d'interprétation structuré de l'image parce qu'il nous semblait le plus approprié pour l'étude de la dynamique dans les images. Rappelons également que nous avons introduit le seuillage échelle par échelle dans le but de mieux extraire les structures de l'image.

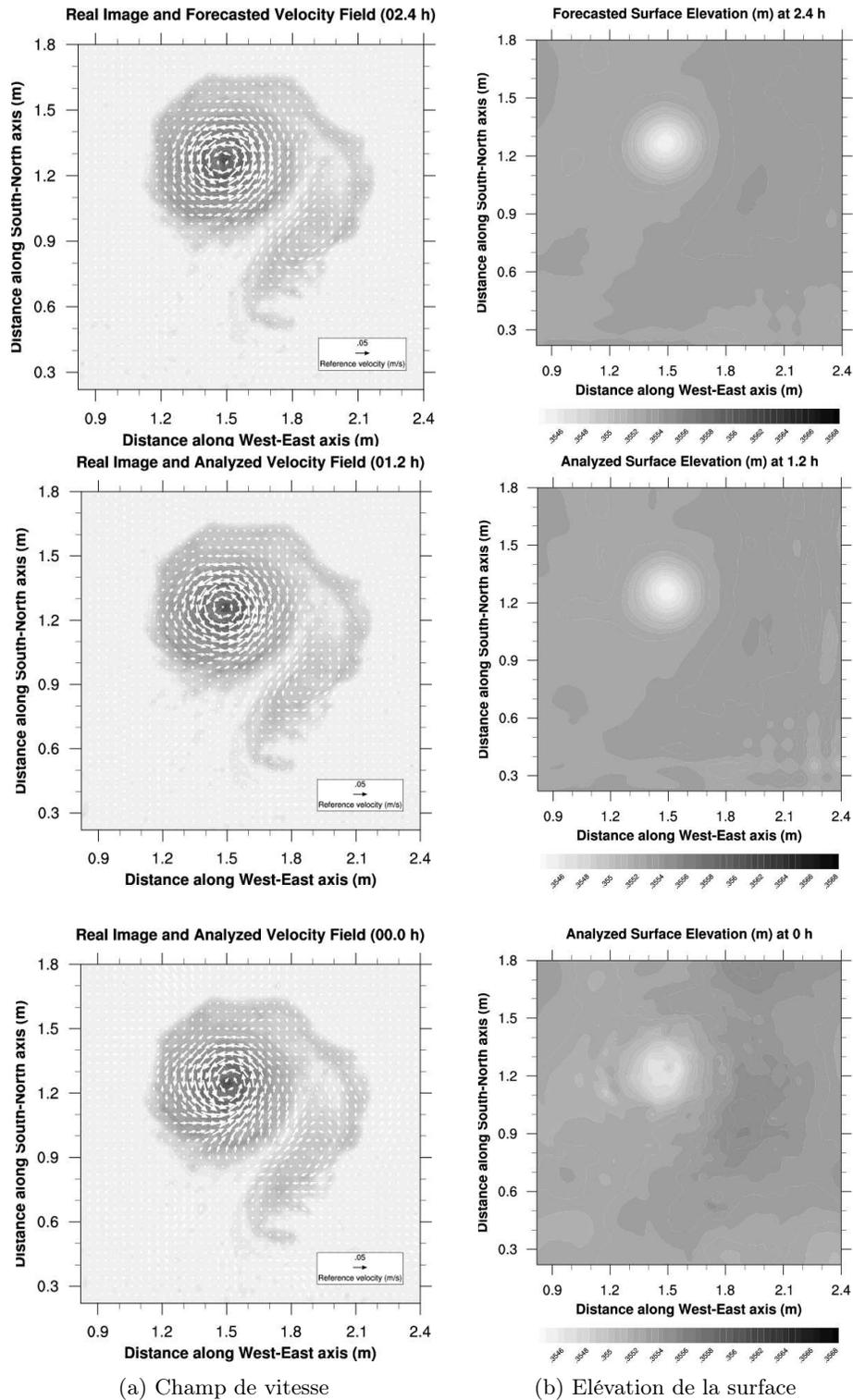


FIGURE 4.10 – Etat analysé par assimilation directe d’une séquence d’images réelles avec représentation en pixel. La figure montre un zoom sur la région contenant le tourbillon avec en (a) le champ de vitesse superposé à l’image à différentes dates et en (b) l’élévation de la surface libre à différentes dates.

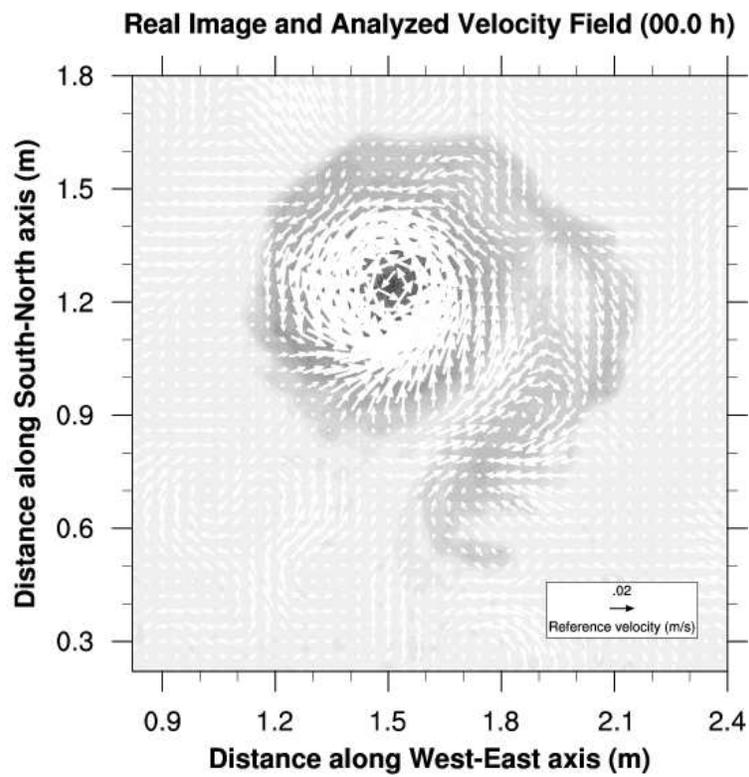


FIGURE 4.11 – Détails sur le champ de vitesse initial reconstruit par assimilation directe d'une séquence d'images réelles avec représentation pixels. La figure montre un zoom sur la région contenant le tourbillon

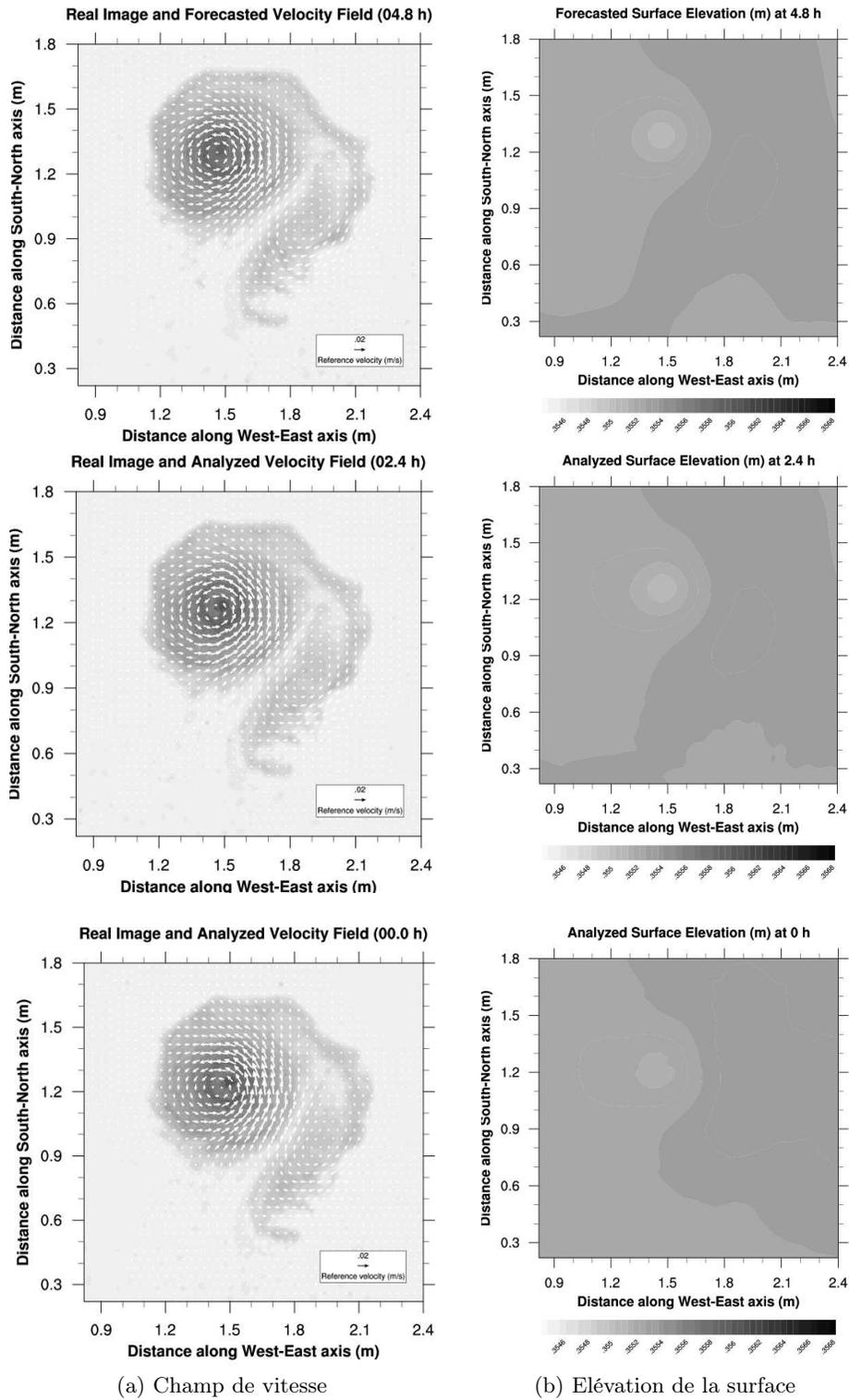


FIGURE 4.12 – Etat analysé par assimilation directe d’une séquence d’images réelles avec décomposition en curvelets et seuillage brut. La figure montre un zoom sur la région contenant le tourbillon avec en (a) le champ de vitesse superposé à l’image à différentes dates et en (b) l’élévation de la surface libre à différentes dates.

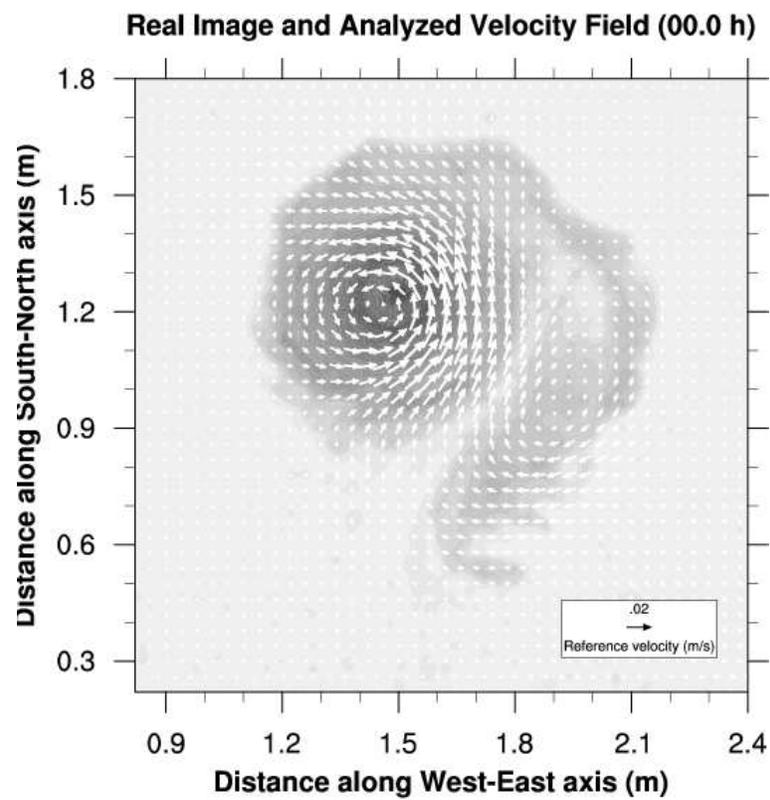


FIGURE 4.13 – Détails sur le champ de vitesse initial reconstruit par assimilation directe d'une séquence d'images réelles avec décomposition en curvelets et seuillage brut. La figure montre un zoom sur la région contenant le tourbillon

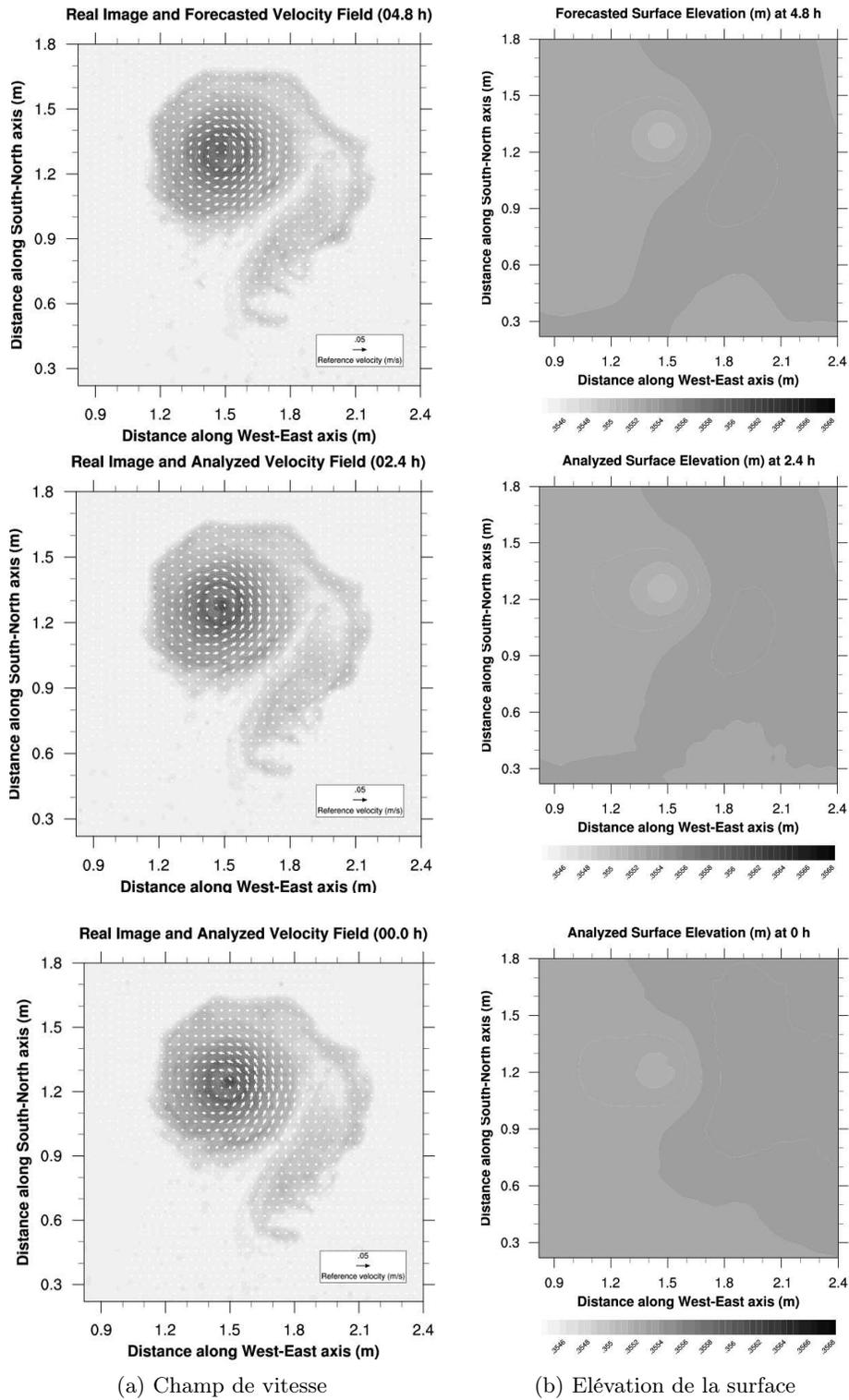


FIGURE 4.14 – Etat analysé par assimilation directe d’une séquence d’images réelles avec décomposition en curvelets et seuillage brut. La figure montre un zoom sur la région contenant le tourbillon avec en (a) le champ de vitesse superposé à l’image à différentes dates et en (b) l’élévation de la surface libre à différentes dates.

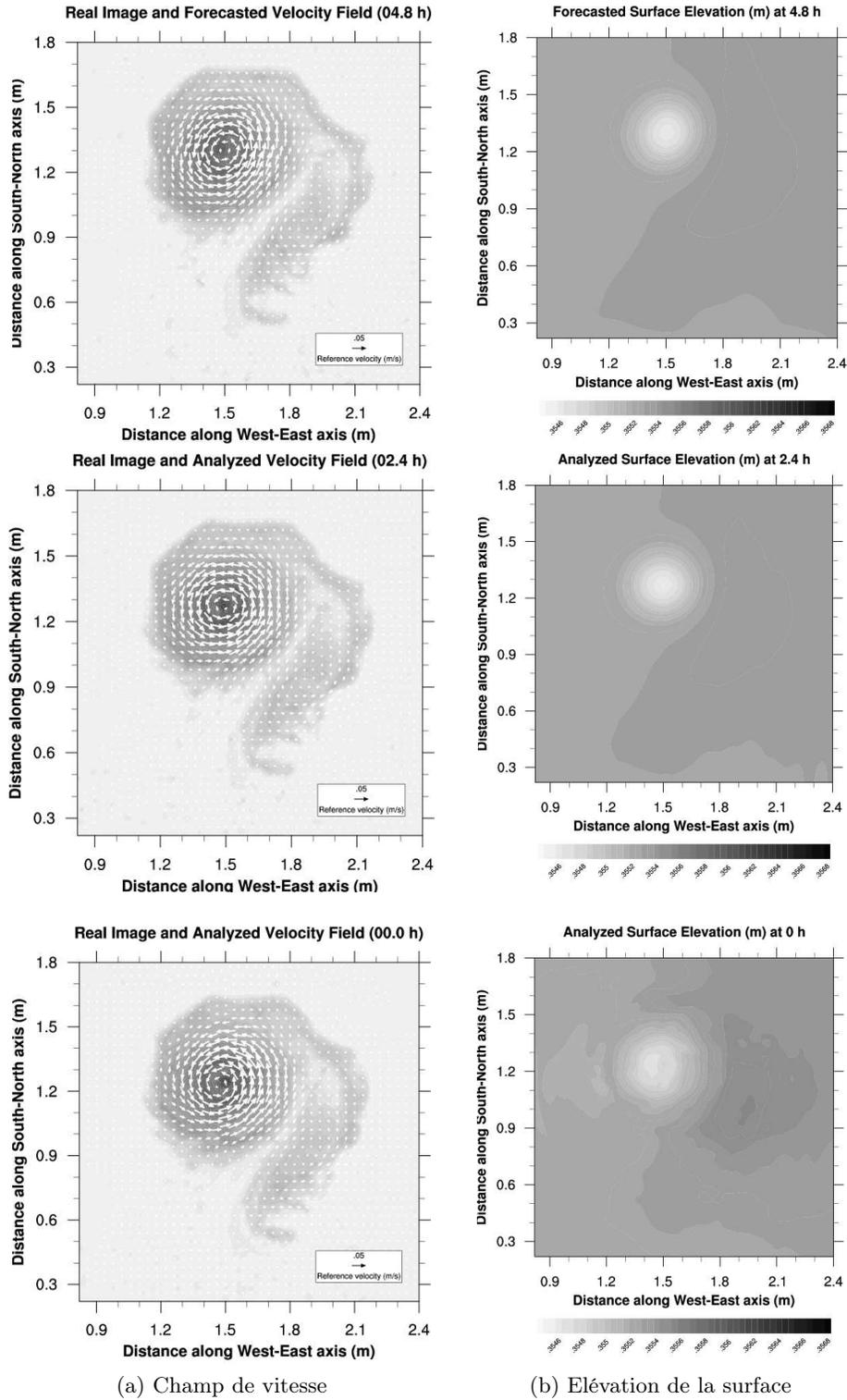


FIGURE 4.15 – Etat analysé par assimilation directe d’une séquence d’images réelles avec décomposition en curvelets et seuillage échelle par échelle. La figure montre un zoom sur la région contenant le tourbillon avec : (a) champ de vitesse superposé à l’image à différentes dates et (b) l’élévation de la surface libre à différentes dates.

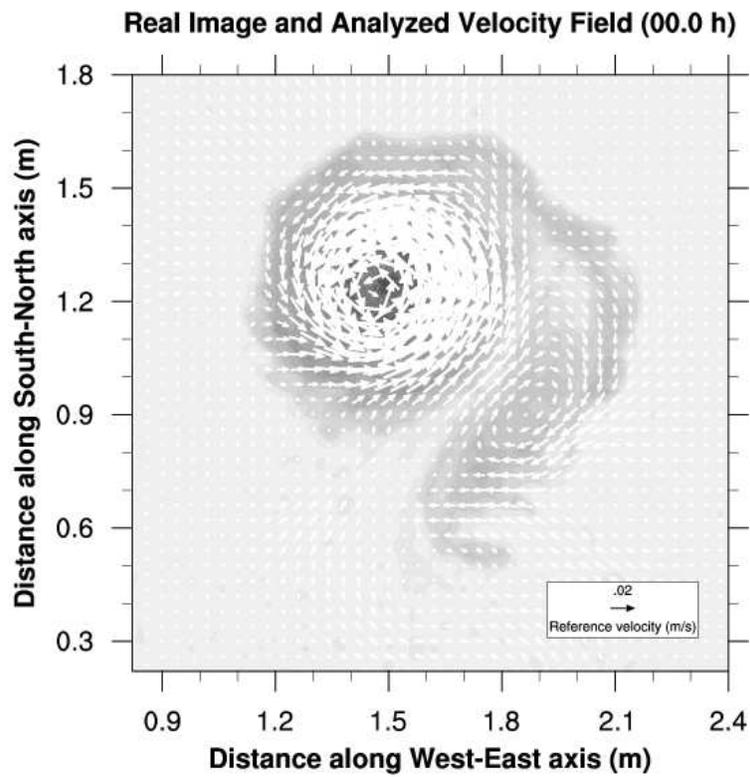


FIGURE 4.16 – Détails sur le champ de vitesse initial reconstruit par assimilation directe d’une séquence d’images réelles avec décomposition en curvelets et seuillage échelle par échelle. La figure montre un zoom sur la région contenant le tourbillon

Comparaison des opérateurs d'observation

La figure (4.17) montre le champ de vitesse estimé respectivement avec :

- des images représentées sous forme de pixels ;
- un seuillage brut sur les coefficients de curvelets ;
- un seuillage échelle par échelle sur les coefficients de curvelets ;
- un seuillage brut sur les coefficients de curvelets avec élimination de l'échelle la plus grossière.

Les vecteurs sont tracés avec une échelle relative au résultat obtenu avec le seuillage brut. Cette échelle permet notamment d'étudier ce qui se passe dans les régions un peu éloigné du centre du tourbillon principal. Comme nous l'avons dit dans les trois paragraphes précédents, les réalisateurs de l'expérience sont plus satisfaits par le résultat produit avec le seuillage échelle par échelle (voir la sous-figure (4.17c)) sur les coefficients de curvelets. Cet opérateur d'observation permet de reconstruire le tourbillon principal et le tourbillon secondaire aux bons endroits, de retrouver aussi bien les ordres de grandeur que la cohérence spatiale du champ de vitesse sur tout le domaine. Les images sous forme de pixels (voir la sous-figure (4.17a)) permettent de reconstruire le tourbillon principal et d'obtenir de bons ordres de grandeur pour le champ de vitesse mais les orientations des vecteurs du champ de vitesse ne sont pas correctes sur tout le domaine. On note principalement l'absence du tourbillon secondaire et des mouvements relativement importants dans les régions où la concentration du traceur n'est pas élevée et où on ne prévoit pas trop de dynamique. On peut penser que ces problèmes sont dus en partie au bruit dans l'image et donc qu'un lissage les corrigerait au moins partiellement. Le seuillage brut sur les coefficients de curvelets a le même effet qu'un lissage ou un débruitage [Starck et al.(2002)]. Si le seuillage brut sur les coefficients des curvelets (voir la sous-figure (4.17b)) permet de corriger les orientations des vecteurs du champ de vitesse, les ordres de grandeur ne sont plus satisfaisants (on perd plus de 50% sur l'intensité maximale) et le tourbillon principal est très déséquilibré. Le seuillage brut corrige en effet les incohérences dues au bruit mais lisse également les discontinuités. Dans le cas particulier des curvelets, un seuillage bruit élimine en effet plus de coefficients des échelles fines au profit des coefficients des échelles grossières. Les fonctions des échelles grossières permettent d'obtenir de l'information globale (la forme) et les coefficients des échelles fines l'information locale (le détail). Les détails étant détruits par le seuillage, l'information la plus importante qui reste est au niveau global, ce qui permet de reconstruire la dynamique globale de translation (beaucoup plus faible) au détriment de la dynamique locale de rotation plus importante. Ceci peut être la principale cause du déséquilibre du tourbillon principal. Pour vérifier cette hypothèse, nous avons implémenté une version de seuillage brut qui élimine tous les coefficients de l'échelle la plus grossière (voir la sous-figure (4.17d)). Avec ce seuillage, nous obtenons de bons ordres de grandeur sur les vecteur du champ de vitesse, les deux tourbillons sont reconstruits mais nous avons également des mouvements importants dans les régions où la concentration du traceur n'est pas élevée et où on ne prévoit pas une dynamique importante. Nous avons ainsi un résultat intermédiaire entre la représentation brute en pixels et le seuillage brut sur les coefficients des curvelets. De plus avec ce seuillage, la minimisation peut se faire sur des fenêtres plus grandes qu'avec la représentation brute en pixels et des résultats comparables sont obtenus avec peu d'itérations. Le seuillage échelle par échelle semble l'alternative la plus appropriée. En procédant échelle par échelle, il ne traite ensemble que des fonctions comparables (des fonctions de la même échelle). La discrimination peut être étendue aux orientations. Cette extension pourra permettre de corriger certains problèmes induits par l'approximation de la

rotation dans un domaine discret. On garde ainsi l'information pertinente à chaque échelle.

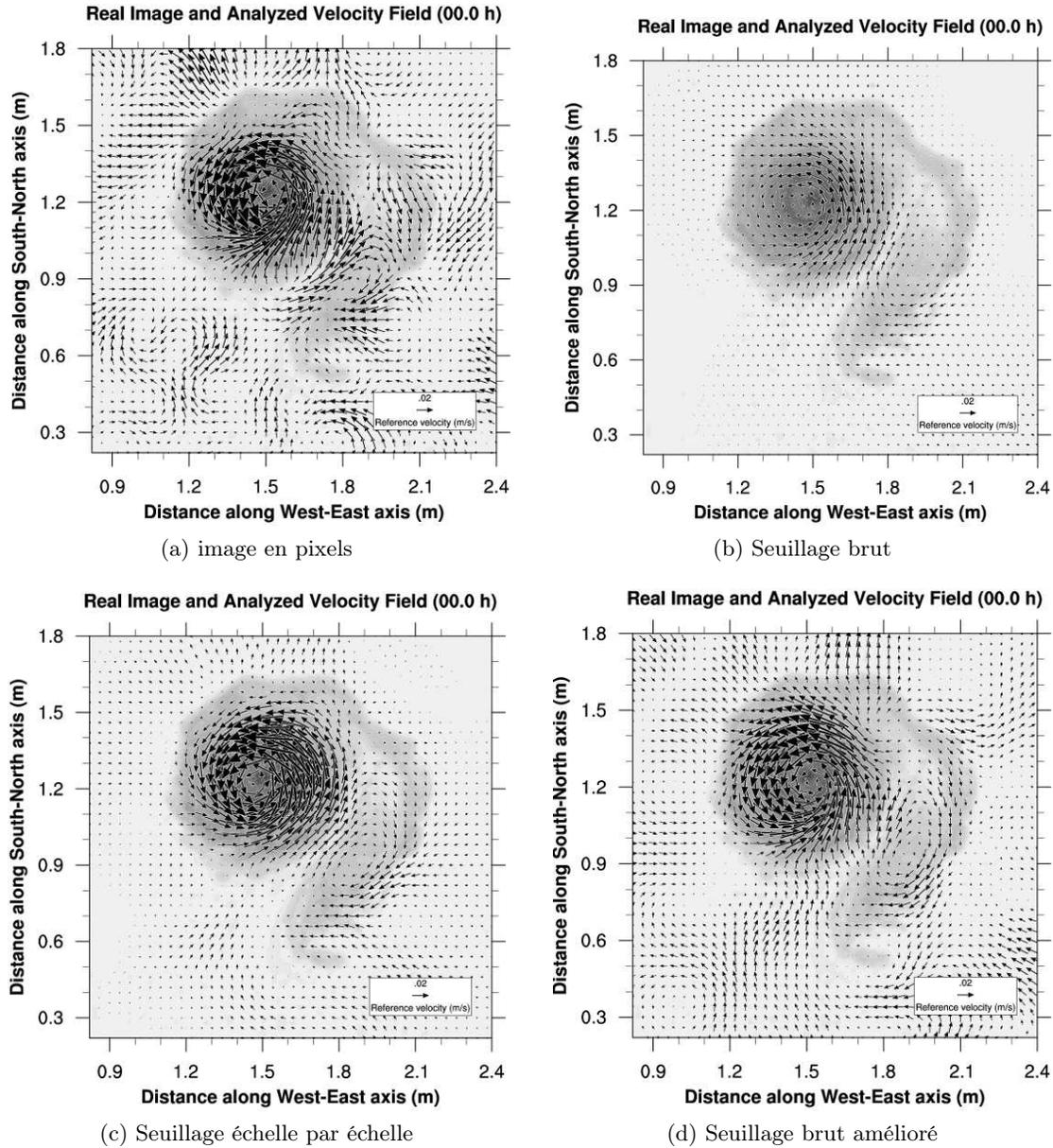


FIGURE 4.17 – Champ de vitesse initial reconstruit par assimilation directe d'une séquence d'images réelles avec différents opérateurs d'observation. Images sous forme de pixels (b); seuillage brut sur la décomposition en curvelets (a); seuillage échelle par échelle sur la décomposition en curvelets (c) et seuillage brut sur la décomposition en curvelets avec élimination de l'échelle grossière (b).

Comparaison des opérateurs d'observation (pixels et curvelets)

La figure (4.18) compare le terme d'écart aux observations avec les opérateurs d'observation respectifs : images sous forme de pixel et seuillage échelle par échelle sur les coefficients de curvelets. Ce terme d'écart aux observations est calculé selon la formule de l'équation (4.38)

pour une fenêtre d'assimilation de 300 pas de temps. Il correspond à l'écart en terme de la norme \mathcal{L}^2 . Son intégrale sur toute la fenêtre d'assimilation correspond à la fonction minimisée lorsqu'on utilise directement l'espace des pixels pour l'écart aux observations. L'étude de cette figure permet de dégager deux principales observations :

1. la minimisation n'améliore pas assez le terme d'écart aux observations contrairement à ce qui a été observé avec les expériences jumelles ;
2. l'opérateur d'observation basé sur un seuillage des coefficients des curvelets produit un meilleur résultat que le calcul dans l'espace des pixels.

La première observation peut avoir plusieurs sources d'explications parmi lesquelles :

- l'utilisation d'un terme d'écart aux observations non approprié (ce problème suggère l'utilisation d'une distance autre que celle basée sur la norme \mathcal{L}^2) ;
- une erreur modèle très importante ;
- une erreur d'observation très importante (elle inclut la différence de luminosité entre les différentes images de la séquence).

Ces sources ne sont pas exclusives. La deuxième observation quant à elle nous montre assez clairement que le calcul de l'écart aux observations dans l'espace des pixels n'est pas adapté au problème. La fonction de diagnostic utilisée pour la comparaison correspond au terme d'écart aux observations utilisé dans l'espace des pixels. On se serait attendu à ce que l'opérateur d'observation dans l'espace des pixels fournisse de meilleurs résultats selon ce diagnostic.

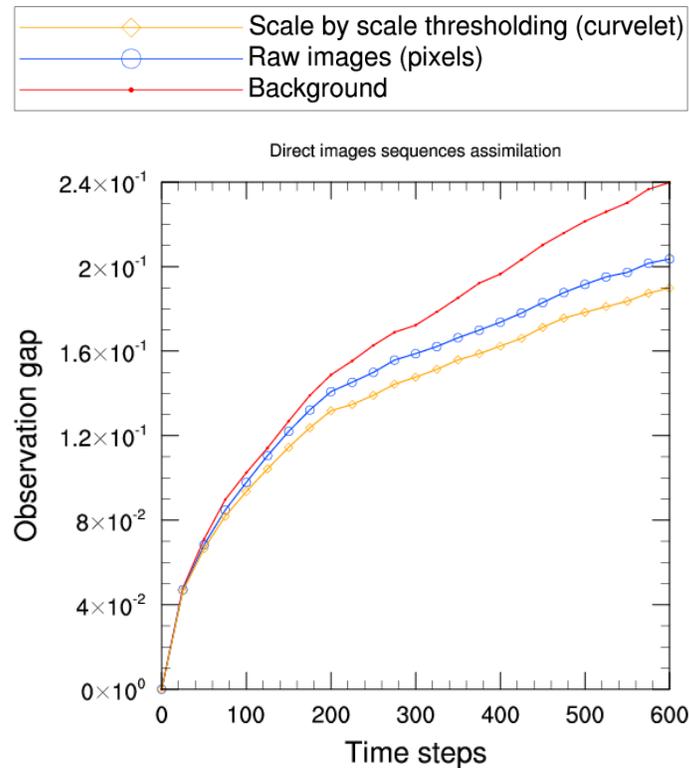


FIGURE 4.18 – Assimilation directe d'images, comparaison de l'écart aux observations avec différents opérateurs d'observation. La première partie du graphique (300 premiers pas de temps) correspond à l'assimilation de données et le reste à la prévision.

Comparaison des différents seuillages sur les coefficients de curvelets

La figure (4.19) compare le résidu (terme d'écart aux observations) obtenu avec différents seuillages sur les coefficients de curvelets. Ce terme d'écart aux observations est calculé pour chaque image dans l'espace des pixels afin d'avoir quelque chose de comparable. Ce diagnostic est calculé pour une fenêtre d'assimilation de 600 pas de temps qui représente la plus grande fenêtre que nous avons pu traiter avec le seuillage brut. Si on ne considère que ce diagnostic pour évaluer les opérateurs d'observation, on conclurait immédiatement que le seuillage brut est le meilleur. Rappelons qu'avec le seuillage brut, on ne reconstitue que la dynamique globale. Laquelle dynamique globale est une approximation très grossière dans les écoulements aussi complexes que la dérive d'un tourbillon qui combine une légère dynamique globale de translation avec une dynamique locale intense de rotation. Si l'on prend en considération le fait qu'un seuillage brut sur les coefficients des curvelets correspond à un lissage dans l'espace des pixels, alors la figure (4.19) nous fait comprendre une fois de plus que le niveau d'interprétation élémentaire est inadapté pour l'étude de la dynamique dans les images. Elle nous fait aussi comprendre également que le terme d'écart aux observations donné par la distance associée à la norme \mathcal{L}^2 n'est pas un critère approprié pour les observations de type image.

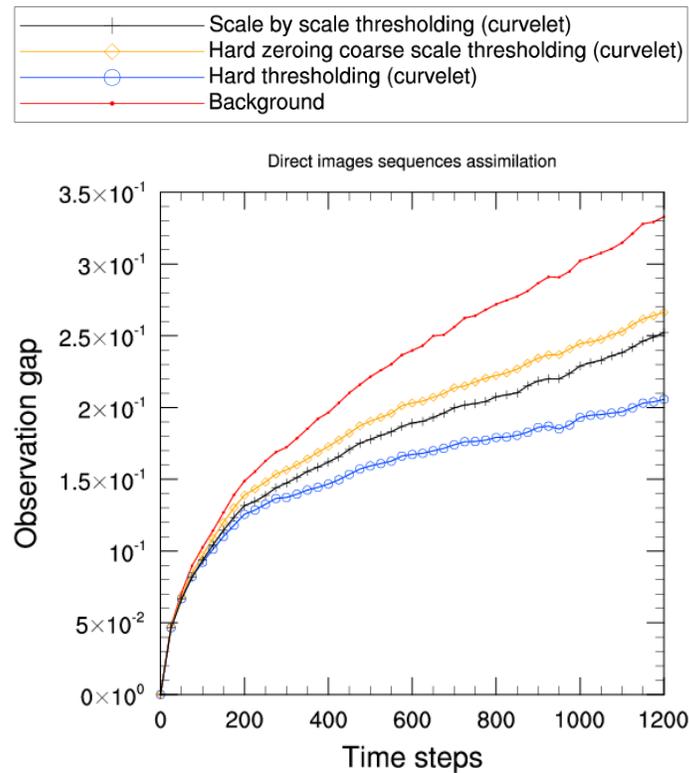


FIGURE 4.19 – Assimilation directe d'images, comparaison de l'écart aux observations pour différents seuillages sur les coefficients des curvelets. La première partie du graphique (600 premiers pas de temps) correspond à l'assimilation de données et le reste à la prévision.

4.5 Conclusion

La prévision des systèmes géophysiques nécessite l'utilisation de toutes les sources d'information disponibles sur ces systèmes. Ces sources d'information, hétérogènes en nature, en qualité et en quantité incluent les modèles mathématiques, les mesures physiques qui constituent les observations, les statistiques d'erreurs, etc. L'assimilation de données permet d'utiliser toutes ces sources d'information de façon optimale pour réaliser la meilleure prévision. Chaque nouvelle source d'observation nécessite la définition d'opérateurs adaptés (opérateur d'observation et opérateur de distance). Les observations de type image constituent l'une des sources d'observation dont l'utilisation quantitative est récente. Elles constituent pour certains systèmes la seule source d'observation. C'est notamment le cas de l'astrophysique où les observations conventionnelles sont presque inexistantes. C'est aussi le cas de la médecine où l'on ne peut pas se permettre n'importe quel type de mesure.

Dans le cas des systèmes géophysiques que sont l'atmosphère et la mer, les mesures par satellite constituent une des sources les plus importantes actuellement. Elles produisent des images (observations indirectes) qui ont une résolution inégalée par les modèles numériques de prévision et une couverture qui défient toutes les autres sources d'observation. Malgré le coût très important de la mise en service des satellites, les images qu'ils produisent étaient jusque là sous utilisées pour la prévision. Pendant longtemps, cette source d'observation a été utilisée de façon qualitative. L'approche d'utilisation par pseudo-observations permettant d'extraire l'information pertinente et de la considérer comme observation conventionnelle ne s'est pas montrée plus probante. Il se pose notamment des problèmes de consistance physique et de corrélation d'erreur avec les autres sources d'information notamment l'ébauche.

L'assimilation directe de cette classe d'observation demande la définition d'opérateurs d'observation et de distance appropriés. Des études récentes ont déjà permis de démontrer la supériorité de cette approche par rapport à l'approche pseudo-observation. Nous avons présenté dans ce chapitre un formalisme général d'assimilation d'observations de type séquences d'images. Un cadre approprié a été défini afin de tester la faisabilité du formalisme proposé. Nous avons montré que l'approche immédiate qui consiste à utiliser les images sous leur forme brute (pixel) n'est pas la plus appropriée. Nous avons introduit la notion de niveau d'interprétation de l'information de type image qui permet de ne considérer de l'image que l'information pertinente. Cette notion nous a permis de définir la notion d'opérateurs d'images qui a ensuite été utilisé pour définir l'opérateur d'observation pour les mesures de type image. Nous avons proposé un formalisme de définition de l'espace des images qui permettrait de mieux exploiter ce type d'information en assimilation de données. Une norme simple est ainsi définie dans cet espace d'images pour mesurer l'écart aux observations. Nous avons expérimenté cette proposition sur la représentation spectrale des images en nous basant sur la décomposition multi-échelle et multi-orientation dans la frame des curvelets. L'utilisation d'un seuillage approprié qui restreint l'image à l'information pertinente qui y est contenue permet d'obtenir des résultats de qualité supérieure à ceux obtenus avec la représentation en pixels. Nous avons notamment défini le seuillage échelle par échelle sur la frame des curvelets qui s'est montré plus adapté pour l'interprétation des images dans un contexte d'assimilation de données.

Nous avons également souligné le problème de régularisation qui est incontournable dans l'utilisation des images (à cause du problème d'ouverture) et le problème de conditionnement. Ces problèmes sont étudiés dans le chapitre suivant où nous proposons une nouvelle approche de régularisation pour les problèmes inverses et son application aux observations de type image.

Chapitre 5

Régularisation pour l'assimilation d'images : une nouvelle approche

Les travaux exposés dans le présent chapitre ont fait l'objet des publications suivantes : [Souopgui et al.(2010c)]; [Souopgui et al.(2010b)] et [Souopgui et al.(2010a)].

5.1 Introduction

5.1.1 Contexte

La description d'un système physique lorsqu'elle est connue peut être exprimée à l'aide d'un modèle et de ses entrées que nous appellerons paramètres¹. L'état du système dans une région de l'espace physique et/ou temporel peut être calculé à partir du modèle mathématique et des paramètres physiques : on parle de problème direct (encore connu sous les expressions modélisation ou simulation). Si dans la vie réelle le modèle mathématique donne une approximation raisonnable de la réalité, les paramètres quant à eux sont en général peu ou mal connus. Dans ce cas ils peuvent être estimés à partir des mesures de l'état du système : on parle alors de problème inverse. L'assimilation de données en général et l'assimilation d'images en particulier en sont des exemples. La nécessité de faire usage des problèmes inverses se rencontrent dans des domaines aussi variés que la géophysique, la télédétection et la vision par ordinateur pour ne citer que ceux-ci. En géophysique et plus particulièrement en océanographie et en météorologie, la prévision est fondée sur la reconstruction d'un état du système à partir des observations : des méthodes éprouvées connues sous le terme assimilation de données (décrit dans les chapitres précédents) permettent d'y arriver. En télédétection, la possibilité d'extraire de l'information significative à partir de mesures dégradées obtenues des appareils de télémessure est une importance capitale. Cette extraction d'informations inclut la restauration d'images. En vision par ordinateur et en traitement d'images, il est courant d'estimer le déplacement entre deux images successives d'une séquence dans un but de compression (vidéogramme) ou du calcul du champ de vitesse ; on parle alors d'estimation du mouvement.

Dans un cas déterministe, le problème direct a une solution unique. Le problème inverse quant à lui n'a de solution unique que dans des situations idéalisées qu'on ne rencontre presque

1. Nous entendons par **paramètres**, les paramètres physiques, les conditions initiales et les conditions aux frontières du domaine physique

pas dans la réalité [Snieder and Trampert(2000)]. La principale cause de cette non unicité de la solution est le caractère sous-déterminé et/ou surdéterminé de ce type de problème. Les données d'observation qui permettent de résoudre le problème inverse sont en quantité très limitée par rapport à la dimension du système. A titre d'illustration, les modèles numériques en météorologie sont actuellement décrits par environ 10^9 variables tandis que les observations sont de l'ordre de 10^6 à 10^7 données. On dit que le problème est sous-déterminé.

Dans certains cas, Il peut y avoir plus d'observations que de paramètres à déterminer, avec des sous-ensembles d'observations redondants et même incohérents. On est en présence de problèmes surdéterminés sans solution (à cause de l'inconsistance des données). On obtient des systèmes surdéterminés notamment quand la résolution des observations est plus fine que la résolution des modèles. A cela s'ajoute la diversité des sources d'observation qui sont à la fois source de surdétermination et d'incohérence. Ces facteurs font que le problème ne soit pas bien posé au sens d'Hadamard : on dit qu'il est mal posé. C'est un défaut caractéristique de la plupart des problèmes inverses. Un autre défaut majeur des problèmes inverses est le mauvais conditionnement ² : de légères variations dans les données entraînent de grandes variations dans la solution du problème. Lorsqu'il s'agit d'un problème inverse, les données correspondent aux observations (plus précisément à la fonction coût) et la solution correspond aux paramètres estimés.

Pour faire face au caractère mal posé des problèmes inverses, on utilise autant que possible les connaissances *a priori* sur les propriétés de la solution au problème. En cas de mauvais conditionnement, les méthodes de préconditionnement (voir la section 2.4.4) sont utilisées. L'une des deux pratiques n'exclut pas l'autre. Les connaissances *a priori* les plus souvent utilisées et dont nous parlerons dans ce chapitre sont :

- l'ébauche (approximation initiale) et les statistiques d'erreurs associées ;
- la régularité de la solution.

Celle qui nous intéresse étant principalement la régularité. Ce chapitre est entièrement consacré au problème de régularisation des problèmes inverses mal posés.

5.1.2 Rappels sur les problèmes inverses et leur résolution

Présentation des problèmes inverses

Problème direct Soit un système physique dont l'état caractérisé par la variable d'état $\mathbf{y} \in \mathcal{Y}$ peut être défini en fonction de la variable dite de contrôle $\mathbf{v} \in \mathcal{V}$ par l'équation suivante :

$$\begin{aligned} \mathcal{M} : \mathcal{V} &\rightarrow \mathcal{Y}, \\ \mathbf{v} &\mapsto \mathbf{y} = \mathcal{M}(\mathbf{v}); \end{aligned} \tag{5.1}$$

Avec \mathcal{M} , le modèle mathématique permettant de passer de l'espace de contrôle \mathcal{V} à l'espace d'état \mathcal{Y} . La variable d'état est notée \mathbf{y} au lieu de \mathbf{x} pour des raisons de convenance ; \mathbf{x} sera utilisé pour noter la variable d'espace physique.

On appelle problème direct le problème qui consiste à déterminer l'état du système à partir d'une réalisation de la variable de contrôle en utilisant le modèle mathématique \mathcal{M} . Dans le cas déterministe, l'état \mathbf{y} du système déterminé à partir d'une réalisation donnée du contrôle

2. En analyse numérique, le conditionnement est une notion qui permet de mesurer la dépendance de la solution d'un problème numérique par rapport aux données du problème, ceci afin de contrôler la validité d'une solution calculée par rapport à ces données.

\mathbf{v} est unique. Dans la réalité, la variable de contrôle est en général peu, partiellement ou mal connue. En revanche, il est courant d'avoir des systèmes d'observation qui permettent d'estimer au moins en partie l'état du système. On définit le problème inverse comme la détermination de la variable de contrôle à partir des données d'observation.

Opérateur d'observation Dans la réalité, il est impossible ou très difficile d'observer entièrement l'état du système. De plus, les observations sont très souvent des mesures indirectes. Ainsi, les observations et l'état du système n'appartiennent pas au même espace. On définit en général l'opérateur d'observation comme l'opérateur permettant de projeter l'espace d'état sur l'espace des observations. Nous le noterons \mathcal{H} et le définirons en termes mathématiques comme suit :

$$\begin{aligned} \mathcal{H} : \mathcal{Y} &\rightarrow \mathcal{O}, \\ \mathbf{y} &\mapsto \mathcal{H}(\mathbf{y}). \end{aligned} \quad (5.2)$$

Definition 5.1.1 *Le problème inverse associé au problème direct défini par l'équation (5.1) et l'opérateur d'observation de l'équation (5.2) est défini en termes de problème aux moindres carrés comme suit :*

$$\text{trouver } \mathbf{v}^* = \text{ArgMin}(J(\mathbf{v})), \mathbf{v} \in \mathcal{V}, \quad (5.3)$$

avec :

$$J(\mathbf{v}) = J_o(\mathbf{v}) = \frac{1}{2} \|\mathcal{H}(\mathcal{M}(\mathbf{v})) - \mathbf{y}^o\|_{\mathcal{O}}^2. \quad (5.4)$$

où $\|\cdot\|_{\mathcal{O}}$ est une norme appropriée définie sur l'espace des observations \mathcal{O}

Le problème défini par l'équation (5.3) est connu comme le problème sans contraintes. Il est de très peu d'utilité à cause des caractères mal posé et mal conditionné qu'on rencontre en général avec les problèmes inverses. Selon la terminologie, la fonction J définie par l'équation (5.4) est appelée fonction coût (en assimilation de données) ou fonction objectif (pour la communauté optimisation en général)

L'existence et l'unicité de la solution au problème sans contrainte (équation 5.3) sont garanties si :

$$\begin{cases} J \text{ est strictement convexe, semi continue inférieure et} \\ \text{coercive } (\lim_{\|\mathbf{v}\| \rightarrow +\infty} J(\mathbf{v}) \rightarrow +\infty). \end{cases} \quad (5.5)$$

Sous ces conditions (équation 5.5), si J est différentiable, alors la solution au problème inverse sans contrainte (équation 5.3) est donnée par la solution de l'équation d'Euler-Lagrange suivante :

$$\nabla J(\mathbf{v}) = 0. \quad (5.6)$$

Les conditions de l'équation (5.5) sont garanties pour les modèles et les opérateurs d'observation linéaires.

Comme nous l'avons dit en introduction, le problème inverse tel que défini par l'équation (5.3) est en général mal posé et/ou mal-conditionné. Pour la prise en compte du caractère mal posé du problème, on utilise dans la mesure du possible les connaissances *a priori* sur le système. Dans le cas du mauvais conditionnement, on utilise des méthodes de préconditionnement. Les deux pouvant s'utiliser pour la résolution du même problème si besoin est. Dans la suite de cette section, nous allons introduire de façon sommaire l'utilisation des connaissances *a priori* et des méthodes de préconditionnement pour les problèmes inverses.

Utilisation des connaissances *a priori* pour la résolution des problèmes inverses

Les connaissances *a priori* sont exprimées en optimisation comme des contraintes sur la solution du problème. Ces contraintes définissent un sous-ensemble $\mathcal{W} \subset \mathcal{V}$ de solutions admissibles. Le problème inverse sans contrainte (équation 5.3) est ainsi transformé en un problème sous contraintes tel que défini par l'équation suivante :

$$\text{trouver } \mathbf{v}^* = \text{ArgMin}(J(\mathbf{v})), \quad (5.7)$$

$$\mathbf{v} \in \mathcal{W}.$$

Nous nous intéressons particulièrement aux cas où le sous-ensemble de solutions admissibles peut s'exprimer comme :

$$\mathcal{W} = \{\mathbf{v} \in \mathcal{V} / g(\mathbf{v}) = 0\}, \quad (5.8)$$

avec la fonction g à définir selon les circonstances.

Le problème sous contraintes (équation 5.7) se ramène alors à un problème sans contrainte avec pénalisation comme suit :

$$\text{trouver } \mathbf{v}_\epsilon^* = \text{ArgMin}(J_o(\mathbf{v}) + \frac{1}{\epsilon_c} J_c(\mathbf{v})), \mathbf{v} \in \mathcal{V}. \quad (5.9)$$

La fonction coût est devenue $J(\mathbf{v}) = J_o(\mathbf{v}) + \frac{1}{\epsilon_c} J_c(\mathbf{v})$ où J_o défini par l'équation (5.4) est appelé terme d'écart aux observations et J_c le terme de contrainte ou de pénalisation défini par :

$$J_c(\mathbf{v}) = \frac{1}{2} \|g(\mathbf{v})\|^2. \quad (5.10)$$

Sous certaines conditions, la solution $\mathbf{v}_{\epsilon_c}^* \rightarrow \mathbf{v}^*$ quand $\epsilon_c \rightarrow 0$. Une formulation équivalente consiste à considérer un paramètre multiplicatif $\alpha_c = \frac{1}{\epsilon_c}$. Le premier formalisme est mieux approprié pour l'étude théorique de la pénalisation des problèmes inverses. Le deuxième étant utilisé dans la pratique. Dans la suite de ce document, nous utiliserons le deuxième formalisme pour des besoins d'ordre pratique et pour la conformité à ce qui est fait dans la communauté du traitement d'images (nos cas pratiques étant empruntés à cette communauté).

Remarque 5.1.1 *Il est connu que la pénalisation pure telle que définie ci-dessus n'est pas efficace dans la pratique avec la résolution numérique; à la place, il est conseillé d'utiliser les algorithmes de type Lagrangien augmentés tels que décrits par Glowinski et Le Tallec [Glowinski and Le Tallec(1987)].*

Nous limiterons l'étude ici aux connaissances *a priori* sur l'ébauche et à la régularité de la solution.

Ebauche et statistiques d'erreurs d'ébauche

Si d'une manière ou d'une autre, on connaît une première approximation de la variable de contrôle à déterminer et les statistiques d'erreurs associées, une contrainte consiste à pénaliser l'écart à cette ébauche relativement aux connaissances des statistiques d'erreurs associées. Le problème pénalisé se définit alors comme :

$$\text{trouver } \mathbf{v}_{\alpha_b}^* = \text{ArgMin}(J_o(\mathbf{v}) + \alpha_b J_b(\mathbf{v})), \mathbf{v} \in \mathcal{V}, \quad (5.11)$$

avec α_b le paramètre de pondération associé au terme d'écart à l'ébauche $J_b(\mathbf{v})$ défini comme suit :

$$J_b(\mathbf{v}) = \frac{1}{2} \|\mathbf{v} - \mathbf{v}^b\|_{\mathcal{V}}^2. \quad (5.12)$$

La forme générale de l'utilisation de l'ébauche est connue comme la régularisation de Tikhonov [Tikhonov(1963)]. \mathbf{v}^b est l'ébauche (première approximation) de la solution et $\|\cdot\|_{\mathcal{V}}$ est une norme appropriée prenant éventuellement en compte les statistiques d'erreurs d'ébauche. Les informations sur l'ébauche sont très utiles dans la résolution des problèmes inverses ; c'est l'approche élémentaire pour faire face au caractère mal posé de ce type de problèmes. Dans le cas où il n'y a pas de connaissance sur l'ébauche, il est courant de considérer une ébauche nulle, ce qui permet de pénaliser la norme de la solution et donc de chercher la solution de norme minimale. Dans les applications réelles et plus précisément en prévision, l'ébauche et les statistiques d'erreurs associées proviennent des analyses précédentes ; c'est ce qui se passe dans les centres opérationnels de prévision. Pour simplifier les écritures, il est courant d'écrire la fonction coût J en fonction de l'écart à l'ébauche $\delta\mathbf{v} = \mathbf{v} - \mathbf{v}^b$ aussi appelé incrément, ce qui conduit à la formulation suivante :

$$J(\delta\mathbf{v}) = \frac{1}{2} \|\mathcal{H}(\mathcal{M}(\mathbf{v}^b + \delta\mathbf{v})) - \mathbf{y}^o\|_{\mathcal{O}}^2 + \frac{1}{2} \alpha_b \|\delta\mathbf{v}\|_{\mathcal{V}}^2. \quad (5.13)$$

Régularité de la solution

Dans certains cas, la physique du problème définit la régularité des solutions admissibles. C'est le cas des écoulements laminaires, irrotationnels ou encore incompressibles. Les informations de régularité sont en général définies en termes de dérivés de la fonction décrivant la variable de contrôle. Dans ce cas, le problème inverse pénalisé se définit comme :

$$\text{trouver } \delta\mathbf{v}_{\alpha_r}^* = \text{ArgMin}(J_o(\delta\mathbf{v}) + \alpha_b J_b(\delta\mathbf{v}) + \alpha_r J_r(\delta\mathbf{v})), \mathbf{v} \in \mathcal{V}, \quad (5.14)$$

où α_r est le paramètre de pondération associé au terme de régularisation $J_r(\delta\mathbf{v})$ défini par :

$$J_r(\delta\mathbf{v}) = \frac{1}{2} \|\Phi(\delta\mathbf{v})\|^2. \quad (5.15)$$

Φ est la fonction de régularisation définie en termes de dérivés de la variable de contrôle. Nous reviendrons plus en détail sur la régularisation dans la section (5.2).

Le reste de ce chapitre est organisé selon le plan décrit dans les lignes suivantes.

Dans la section (5.2), nous présenterons la régularisation pour les problèmes inverses en mettant un accent particulier sur la régularisation des champs de vecteurs en traitement d'images. Nous étudierons en particulier la complexité de la détermination des paramètres optimaux de pondération des termes de régularisation et le comportement de ces méthodes à travers un ensemble de fonctions de diagnostic.

A la section (5.4), nous introduirons la nouvelle approche de régularisation que nous proposons. La présentation sera faite dans un cadre général avec un accent sur le cas particulier des champs de vecteurs. Nous fournirons les détails sur l'utilisation de quelques opérateurs de régularisation avec la nouvelle approche ; notamment la pénalisation du gradient, la pénalisation du premier et du second ordre de la divergence et du rotationnel. Une implémentation spéciale basée sur la **diffusion généralisée** est présentée pour ces cas classiques et les bases de son

extension aux autres cas sont exposées. Cette implémentation est la clé de l'utilisation des opérateurs de régularisation en opérateur de préconditionnement et constitue un des principaux atouts de cette approche. Nous démontrerons au passage l'équivalence entre la pénalisation du gradient et l'équi-pénalisation du premier ordre de la divergence et du rotationnel énoncé par Gupta et Prince [Gupta and Prince(Feb 1996)].

A la section (5.5), nous présenterons l'application de l'approche introduite ici à l'estimation du mouvement. Nous montrerons une étude expérimentale du comportement de cette approche en fonction des paramètres de pondération. Nous présenterons une étude expérimentale et comparative de cette méthode avec les autres sur un certain nombre de fonctions de diagnostic telles que l'évolution de la fonction objectif, l'évolution de l'erreur sur le champ de vitesse estimé et sa vorticité ainsi que l'erreur angulaire par rapport au champ vrai.

5.2 Etat de l'art : régularisation par pénalisation

Comme nous l'avons dit dans la section (5.1.2), la régularisation est une façon d'utiliser les connaissances *a priori* sur la régularité de la solution au problème. Dans la résolution des problèmes inverses par des méthodes de moindres carrés, la régularisation est faite par ajout à la fonction coût d'un terme de pénalisation de l'écart à la régularité. Comme présenté dans la section (5.1.2), le terme de pénalisation de l'écart à la régularité se présente sous la forme :

$$J_r(\mathbf{v}) = \frac{1}{2} \int_{\Omega} \|\Phi(\mathbf{v}(\mathbf{x}))\|^2 d\mathbf{x}. \quad (5.16)$$

Pour son intégration dans la fonction coût, ce terme de régularisation J_r est pondéré par un paramètre de régularisation α_r à déterminer. Φ est la fonction de régularisation à définir selon la régularité recherchée et $\|\cdot\|$ la norme quadratique dans l'espace approprié. La solution \mathbf{v}^* est ainsi contrainte à être lisse (dans un sens défini par la fonction Φ) : $\|\Phi(\mathbf{v}(\mathbf{x}))\| \approx 0$. Comme nous l'avons dit dans la section (5.1.2), la fonction de régularisation Φ est basée sur les dérivées de la variable de contrôle \mathbf{v} . L'ordre des dérivées prises en compte définit l'ordre de la régularisation. On parlera de régularisation d'ordre k lorsque Φ est défini en termes des dérivées allant jusqu'à l'ordre k . Dans la pratique, on se limite en général à l'ordre 2.

Remarque 5.2.1 *Dans ce document, nous prendrons en compte uniquement des fonctions de régularisation linéaires et distributives par rapport à l'addition ; ce sont les fonctions Φ telles que $\Phi(\mathbf{v} + \lambda\mathbf{w}) = \Phi(\mathbf{v}) + \lambda\Phi(\mathbf{w})$ avec $\mathbf{v}, \mathbf{w} \in \mathcal{V}$ et $\lambda \in \mathbb{R}$.*

5.2.1 Régularisation des champs de vecteurs

Pour commencer cette section, nous allons donner quelques précisions sur les notations génériques définies à la section (5.1.2) et utilisées jusqu'à présent notamment en ce qui concerne l'espace de contrôle.

Notations Supposons que le domaine physique du système étudié soit noté Ω , un sous-ensemble ouvert de \mathbb{R}^m ($\Omega \subset \mathbb{R}^m$), nous sommes intéressés ici par des espaces de contrôle du type $\mathcal{V} = (L^2(\Omega))^n$. La variable de contrôle se définit alors comme $\mathbf{v}(\mathbf{x}) = (v_i(\mathbf{x}))_{1 \leq i \leq n}$ où la variable d'espace physique \mathbf{x} est définie par $\mathbf{x} = (x_i)_{1 \leq j \leq m} \in \Omega$.

Régularisation du premier ordre

Fonctions de régularisation usuelles

Les méthodes de régularisation du premier ordre définissent les fonctions de régularisation Φ comme une fonction des dérivées premières de la variable de contrôle \mathbf{v} :

$$\Phi(\mathbf{v}) = \Phi^{(1)} \left(\frac{\partial v_i}{\partial x_j} \right)_{1 \leq i, j \leq n} . \quad (5.17)$$

L'exposant (1) dans $\Phi^{(1)}$ dénote le premier ordre. La fonction de régularisation $\Phi^{(1)}$ est précisée selon les cas et les applications ; Une revue de telles fonctions peut se trouver dans le livre de Corpetti [Corpetti(2004)]. La méthode de régularisation du premier ordre la plus utilisée est la pénalisation du gradient. Elle a été formulée par Horn et Schunck dans la formulation du flot optique [Horn and Schunck(1981)] pour l'estimation du mouvement. Le terme de régularisation de Horn et Schunck se définit comme suit :

$$J_{grad}(\mathbf{v}) = \frac{1}{2} \int_{\Omega} \sum_{i=1}^n \|\nabla v_i\|^2 d\mathbf{x}. \quad (5.18)$$

La fonction de régularisation associée est définie par :

$$\Phi_{grad}(\mathbf{v}) = ((\nabla v_1)^T, (\nabla v_2)^T, \dots, (\nabla v_n)^T)^T . \quad (5.19)$$

Pour les écoulements incompressibles ou irrotationnels, on fait recours à la pénalisation de la divergence ou du rotationnel du champ de vecteur, ce qui définit la régularisation par pénalisation de la divergence :

$$J_{div}(\mathbf{v}) = \frac{1}{2} \int_{\Omega} \|\mathit{div}(\mathbf{v})\|^2 d\mathbf{x}, \quad (5.20)$$

pour les fluides incompressibles ou la régularisation par pénalisation du rotationnel :

$$J_{curl}(\mathbf{v}) = \frac{1}{2} \int_{\Omega} \|\mathit{curl}(\mathbf{v})\|^2 d\mathbf{x}, \quad (5.21)$$

pour les écoulements irrotationnels.

Les fonctions de régularisation respectives sont :

$$\Phi_{div}(\mathbf{v}) = \mathit{div}(\mathbf{v}) = \nabla \cdot \mathbf{v}, \quad (5.22)$$

et :

$$\Phi_{curl}(\mathbf{v}) = \mathit{curl}(\mathbf{v}) = \nabla \times \mathbf{v}. \quad (5.23)$$

Forme générale des régularisations du premier ordre

Dans le but de prendre en considération les informations sur les données dans la régularisation, la fonction de régularisation Φ peut être définie comme une fonction de la variable de contrôle \mathbf{v} à travers ses dérivées $(\frac{\partial v_i}{\partial x_j})$ et des données \mathbf{y}^o (observations). Ceci est notamment le cas en estimation de mouvement par flot optique :

$$\Phi = \Phi(\mathbf{v}, \mathbf{y}^o) = \Phi_{(g)} \left(\frac{\partial v_i}{\partial x_j}, \mathbf{y}^o \right)_{1 \leq i, j \leq n}. \quad (5.24)$$

g pour général dans $\Phi_{(g)}$. Les principaux termes de régularisation pour le flot optique de cette forme et leurs propriétés sont résumés dans le tableau (5.1). Une revue de leur dérivation pouvant se trouver dans les travaux de Weickert et des ses co-auteurs [Weickert and Schnörr(2001), Weickert et al.(2006)].

TABLE 5.1 – Termes généraux de régularisation pour le flot optique et fonctions de régularisation associées

terme de régularisation	fonction de régularisation
$J_{r1}(\mathbf{v}, f) = \int_{\Omega} \sum_{i=1}^n \ \nabla v_i\ ^2 d\mathbf{x}$ Homogène de Horn et Schunck [Horn and Schunck(1981)]	$\Phi_1(\mathbf{v}) = ((\nabla v_1)^T, (\nabla v_2)^T, \dots, (\nabla v_n)^T)^T$ Homogène
$J_{r2}(\mathbf{v}, f) = \int_{\Omega} g(\ \nabla f\ ^2) \sum_{i=1}^n \ \nabla v_i\ ^2 d\mathbf{x}$ isotropique basé sur l'image [Alvarez et al.(1999)]	$\Phi_2(\mathbf{v}) = g(\ \nabla f\ ^2) \Phi_1(\mathbf{v})$ linéaire et isotropique
$J_{r3}(\mathbf{v}, f) = \int_{\Omega} \sum_{i=1}^n \nabla v_i D(\nabla f) \nabla v_i d\mathbf{x}$ anisotropique basé sur l'image de Nagel [Nagel(1983)]	$\Phi_3(\mathbf{v}) = D(\nabla f)^{\frac{1}{2}} \Phi_1(\mathbf{v})$ linéaire et anisotropique
$J_{r4}(\mathbf{v}, f) = \int_{\Omega} \Psi \left(\sum_{i=1}^n \ \nabla v_i\ ^2 \right) d\mathbf{x}$ isotropique basé sur le champ de Schnörr [Schnörr(1994)]	$\Phi_4(\mathbf{v}) = \Psi \left(\sum_{i=1}^n \ \nabla v_i\ ^2 \right)$ non linéaire et isotropique
$J_{r5}(\mathbf{v}, f) = \int_{\Omega} \text{trace} \Psi \left(\sum_{i=1}^n \nabla v_i (\nabla v_i)^T \right) d\mathbf{x}$ anisotropique basé sur le champ de Weickert et Schnörr [Weickert and Schnörr(2001)]	$\Phi_5(\mathbf{v}) = \text{trace} \Psi \left(\sum_{i=1}^n \nabla v_i (\nabla v_i)^T \right)$ non linéaire et anisotropique

Explications pour la lecture du tableau 5.1 f dénote les données d'observations (notamment l'image dans le cas du flot optique) au temps initial. Ce qui correspond à $\mathbf{y}^o(t = 0)$. Les fonctions Ψ et g sont des fonctions robustes³ utilisées pour gérer les discontinuités. C'est un concept emprunté au domaine de la statistique [Huber(1981)] et qui a montré son utilité pour l'estimation du mouvement par flot optique [Black and Anandan(1991), Hinterberger et al.(2002), Mémin and Pérez(1998)]. $\Psi(s^2)$ est une fonction différentiable, croissante et convexe en s et $g(s^2)$ est une fonction décroissante en s^2 . Des exemples de ces fonctions

3. Les fonctions robustes ont des propriétés de régularisation qui permettent de pénaliser les effets des valeurs extrêmes/anormales apparaissant dans une série de données

peuvent être trouvés dans les travaux de Weickert et Schnörr [Weickert and Schnörr(2001)] :

$$\Psi(s^2) = \epsilon s^2 + (1 - \epsilon)\lambda^2 \sqrt{(1 + s^2/\lambda^2)}; \quad (5.25)$$

$$g(s^2) = \epsilon + \frac{(1 - \epsilon)}{\sqrt{1 + s^2/\lambda^2}}; \quad (5.26)$$

avec $0 < \epsilon \ll 1$ et $\lambda > 0$. Leur extension aux matrices symétriques⁴ est définie comme suit :

$$\Psi(\mathbf{A}^2) = \sum_{i=1}^n \Psi(\mu_i^2) \mathbf{w}_i \mathbf{w}_i^T, \quad (5.27)$$

$$g(\mathbf{A}^2) = \sum_{i=1}^n g(\mu_i^2) \mathbf{w}_i \mathbf{w}_i^T, \quad (5.28)$$

où \mathbf{A} est une matrice symétrique $n \times n$, μ_i la valeur propre de \mathbf{A} associée au vecteur propre unitaire \mathbf{w}_i . $D(\nabla f)$ est un tenseur de diffusion défini comme :

$$D(\nabla f) = \frac{1}{\|\nabla f\|^2 + 2\lambda^2} \left((\nabla f^\perp)(\nabla f^\perp)^T + \lambda^2 I \right), \quad (5.29)$$

où ∇f^\perp est un vecteur orthogonal au gradient ∇f de la donnée d'observation avec une norme identique ($\|\nabla f^\perp\| = \|\nabla f\|$). $D(\nabla f)^{\frac{1}{2}}$ est défini par la factorisation de Cholesky comme $D(\nabla f) = D(\nabla f)^{\frac{1}{2}} D(\nabla f)^{\frac{T}{2}}$. Par définition, l'opérateur $D(\nabla f)$ est symétrique et défini positif, donc une telle factorisation est possible.

Régularisation du second ordre

Les méthodes de régularisation du second ordre sont basées sur les dérivées jusqu'à l'ordre 2 de la variable de contrôle \mathbf{v} :

$$\Phi(\mathbf{v}) = \Phi^{(2)} \left(\frac{\partial^2 v_i}{\partial x_j \partial x_k}, \frac{\partial v_r}{\partial x_s} \right)_{1 \leq i, r \leq n; 1 \leq j, k, s \leq m}. \quad (5.30)$$

L'exposant (2) dans $\Phi^{(2)}$ dénote le second ordre. L'exemple le plus utilisé est la régularisation de Suter [Suter(1994)] basée sur les dérivées premières de la divergence et du rotationnel. Le terme de régularisation associé est défini par :

$$J_{suter}(\mathbf{v}) = \frac{1}{2} \int_{\Omega} \alpha \|\nabla \operatorname{div}(\mathbf{v})\|^2 + \beta \|\nabla \operatorname{curl}(\mathbf{v})\|^2 dx. \quad (5.31)$$

Remarque 5.2.2 *Les dérivées d'ordre supérieur peuvent également être utilisées pour définir les fonctions de régularisation. Chen et Suter [Chen and Suter(1998)] ont ainsi étendu la régularisation de Suter (équation 5.31) aux dérivées d'ordre m de la divergence et du rotationnel :*

$$J_m(\mathbf{v}) = \frac{1}{2} \int_{\Omega} \alpha \|\nabla^m \operatorname{div}(\mathbf{v})\|^2 + \beta \|\nabla^m \operatorname{curl}(\mathbf{v})\|^2 dx. \quad (5.32)$$

Cependant, dans la pratique, on se limite en général à l'ordre 2.

4. rappelons que les matrices symétriques ont des valeurs propres positives

5.2.2 Techniques d'amélioration de l'impact de la régularisation

L'utilisation des fonctions robustes permet de prendre en compte une bonne partie des connaissances *a priori* sur le comportement du système. On rencontrera dans la littérature de la régularisation des développements basés sur la pénalisation du gradient (équation 5.18) notamment ceux résumés dans le tableau (5.1) ou encore sur la régularisation de Suter (équation 5.31) Du côté de la communauté de la vision par ordinateur, des efforts ont également été faits dans le but de la préservation des discontinuités (qui sont en général lissées par la régularisation); les travaux de Alvarez [Alvarez et al.(1999)] et de Corpetti [Corpetti(2004)] représentent de bonnes références sur ce sujet. Dans le cas de la dynamique des fluides, des fonctions de régularisation spéciales ont été développées afin de mieux prendre en compte les caractéristiques des écoulements fluides telles que le rotationnel, la divergence ou encore le cisaillement : [Mémmin and Pérez(1998)] propose notamment des termes de régularisation basés sur les structures propres aux écoulements fluides ; [Corpetti et al.(2002)] introduit des fonctions scalaires intermédiaires servant d'approximation des caractéristiques d'intérêt. Notons également que plusieurs termes de régularisation peuvent être combinés afin d'obtenir un meilleur résultat. Dans tous les cas, le choix du paramètre de pondération des termes de régularisation reste crucial.

5.2.3 Limites de la régularisation par pénalisation

Comme nous l'avons dit plus haut, le but de la régularisation pour les problèmes inverses est d'obtenir une solution avec une certaine régularité. Théoriquement, cet objectif est atteint en réduisant l'espace des solutions candidates au sous-espace des candidats réguliers. Dans la pratique, cette contrainte de régularité est prise en compte en complétant la fonction coût par un terme de pénalisation de l'écart à la régularité comme présenté à la section (5.1.2). Dans la plupart des cas, cette pratique est affectée par trois principaux problèmes :

- interprétation physique des termes de régularisation ;
- détermination des paramètres de pondération des termes de régularisation ;
- convergence des algorithmes de minimisation.

Interprétation physique des termes de régularisation

Telle qu'utilisée dans la plupart des cas, la régularisation pour l'estimation de mouvement n'a pas toujours une justification physique valide notamment dans le cas des écoulements fluides. La régularisation devrait dépendre de l'application ; voir les travaux de Amodei [Amodei and Benbourhim(1991)] pour la régularisation des champs de vitesse en général et ceux de Vigan [Vigan et al.(2000)] pour l'estimation du mouvement en océanographie. Dans la pratique, ce n'est pas toujours le cas : seule la régularisation par pénalisation des caractéristiques spéciales non rencontrées dans un écoulement particulier peut être justifiée avec des arguments relatifs à la physique du problème. A titre d'exemple, nous avons la pénalisation de la divergence dans le cas d'un écoulement incompressible ou encore de la pénalisation du rotationnel dans le cas d'un écoulement irrotationnel.

La pénalisation homogène des variations spatiales telle que la régularisation de Horn et Schunck [Horn and Schunck(1981)] ou encore la régularisation du second ordre de Suter n'ont pas de justification physique valable pour les écoulements fluides en général. Font exception à ce constat : les écoulements laminaires dans le cas de la régularisation de Horn et Schunck et les écoulements à rotationnel et à divergence constant dans le cas de la régularisation de Suter. Ce

problème rencontré avec la régularisation par pénalisation des variations spatiales peut aussi se rencontrer lorsqu'on traite des mouvements solides en présence de plusieurs objets mouvant indépendamment les uns des autres : la pénalisation des variations spatiales peut être considérée à l'intérieur d'un objet solide mais pas à l'extérieur dans le cas général. Pour faire face au problème du déplacement d'objets, Papadakis et Mémin[Papadakis and Mémin(2008b)] proposent de coupler l'estimation du mouvement avec le suivi de contour des objets. Cette approche permet de résoudre le problème dans le cas d'objets solides mais son application serait assez limitée dans le cas des fluides où il n'y a pas d'objet proprement dit. Plusieurs autres propositions ont été faites dans le but de résoudre les problèmes inhérents à la pénalisation homogène des variations spatiales, on peut noter entre autres : la régularisation isotropique basée sur l'image [Alvarez et al.(1999)], la régularisation anisotropique basée sur l'image [Nagel(1983)], la régularisation isotropique basée sur le champ [Schnörr(1994)] ou encore la régularisation anisotropique basée sur le champ [Nagel(1983)].

Les améliorations notées avec les dernières adaptations résultent de l'utilisation d'un terme d'écart aux observations modifié et d'un schéma itératif adapté au contexte. Des adaptations qui permettent de passer du schéma de minimisation d'une fonctionnelle à un processus de diffusion. Bien que produisant de très bons résultats dans le cas de l'estimation d'un champ de vitesse statique, ces adaptations ne peuvent pas être admises dans le cas général. Elles nécessiteraient dans le meilleur des cas de reformuler le modèle et le schéma d'optimisation pour ramener la résolution du problème à un processus de diffusion. Cette pratique doit être envisagée selon les cas car :

- il n'y a *a priori* pas de garantie que le changement de schéma de résolution du problème soit sans effet sur les autres types d'observation déjà pris en compte ;
- on ne peut pas toujours se permettre la réécriture des modèles pour la prise en compte d'une observation particulière. Les modèles de prévision nécessitent en général plusieurs années pour être opérationnels.

Paramètres de pondération

Le paramètre optimal de pondération d'un terme de régularisation dépend des données d'observations. Or dans la formulation des problèmes inverses, les observations n'apparaissent que dans l'expression de la fonction coût. De ce fait, l'étude de la dépendance ou de la sensibilité aux observations ne peut se faire qu'en considérant le système d'optimalité⁵ de l'assimilation variationnelle de données. C'est en effet le système qui contient toute l'information (le modèle direct et son adjoint et les observations). Son étude se fait au second ordre car nécessitant sa différentiation et donc celle du modèle adjoint encore appelé adjoint au premier ordre : c'est la différentiation au second ordre du modèle.

La détermination du paramètre optimal de pondération nécessite donc une analyse au second ordre et ce pour chaque jeu d'observations. Ce qui est actuellement peu envisageable dans la pratique. La littérature n'est pas abondante sur ce sujet de détermination du paramètre optimal de régularisation dont l'importance est cruciale pour la résolution des problèmes inverses mal posés. Nous étudierons ce problème dans un cadre expérimental dans la section suivante.

5. En assimilation variationnelle de données, le système d'optimalité désigne le système définissant le problème général de minimisation de la fonction coût. Il a la particularité de contenir toute l'information sur le problème : le modèle, la fonction coût à minimiser, les observations via la fonction coût et le modèle adjoint permettant de calculer le gradient de la fonction coût

5.3 Etude expérimentale de la régularisation par pénalisation

L'étude présentée dans cette section est reprise en annexe (E) avec plus d'illustrations

Afin de mieux comprendre la régularisation par pénalisation, nous avons fait une étude expérimentale d'estimation du mouvement en utilisant différents termes de régularisation. L'annexe (D) donne une description de ce problème ainsi que des techniques utilisées dans ce document ; les données expérimentales utilisées y sont présentées également. Nous présentons dans cette section les résultats de cette étude. Il s'agit notamment du comportement des solveurs régularisés :

- en fonction du paramètre de pondération ;
- au cours de la minimisation ;
- en fonction de différents termes de régularisation.

Ce comportement est étudié par l'intermédiaire d'un ensemble de fonctions de diagnostic.

5.3.1 Fonctions de diagnostic

Fonctions de diagnostic utilisées

Un ensemble de diagnostics est calculé afin d'étudier la qualité du résultat obtenu et le comportement des différents termes de régularisation. Les diagnostics utilisés sont :

- le terme d'écart aux observations de la fonction coût (encore appelé résidu) ; il permet d'évaluer la reconstruction de la deuxième image de la séquence à partir du champ estimé et de la première image ;
- la fonction coût globale (somme du terme d'observation et des termes de régularisation) ; c'est la fonctionnelle qui est minimisée pendant le processus itératif. Ce diagnostic permet d'évaluer l'impact des termes supplémentaires sur la convergence globale de la méthode ;
- l'erreur sur le champ de vitesse estimé ; elle permet de mesurer la capacité de la méthode à reconstruire le champ ;
- l'erreur sur la vorticité ; la vorticité étant un indicateur caractéristique majeur dans un écoulement fluide. Ce diagnostic permet de mesurer la capacité de la méthode à retrouver les caractéristiques d'un écoulement fluide ;
- l'erreur angulaire ; c'est l'écart angulaire entre le champ vrai et le champ estimé. Depuis les travaux de Barron [Barron et al.(1994)], cet outil de diagnostic s'est montré très utile pour l'étude des champs de vecteurs. Dans le paragraphe (5.3.2), nous donnons quelques détails sur l'écart angulaire et nous proposons une formulation numérique qui comble certaines lacunes de la formulation de Barron couramment utilisée.

Normalisation des fonctions de diagnostic

Sauf indication contraire, tous les graphiques que nous présentons ici représentent les courbes des fonctions de diagnostic normalisées. Il s'agit de :

- NRMSE (Normalized Root Mean Square Error) pour l'erreur sur le champ estimé et sa vorticité. L'erreur moyenne est normalisée par l'erreur moyenne sur l'ébauche qui dans nos expériences est un champ nul ;
- le terme d'observation de la fonction coût et l'erreur angulaire sont normalisés par rapport à leurs équivalents calculés avec comme variable de contrôle l'ébauche.

Ainsi normalisées, les fonctions de diagnostic permettent de quantifier aisément l'impact de la régularisation utilisée et de son paramètre de pondération selon la convention suivante :

- $NRMSE < 1 \Rightarrow$ l'estimation de mouvement a globalement amélioré la qualité de l'approximation par rapport à l'ébauche selon le diagnostic utilisé ;
- $NRMSE = 1 \Rightarrow$ l'estimation de mouvement n'a globalement aucun effet sur l'approximation selon le diagnostic utilisé ;
- $NRMSE > 1 \Rightarrow$ l'estimation de mouvement a globalement détérioré la qualité de l'approximation par rapport à l'ébauche selon le diagnostic utilisé.

Cette interprétation est valable également pour les autres diagnostics que sont la fonction coût, le terme d'observation de la fonction coût et l'erreur angulaire dans leur version normalisée. Pour rendre la compréhension plus aisée, la lecture sera faite en termes de pourcentage, ceci correspondra aux valeurs normalisées multipliées par 100. Afin de ne pas se limiter aux valeurs moyennes, nous présentons également les valeurs maximales pour toutes les fonctions de diagnostic sauf dans le cas de la fonction coût et de son terme d'écart aux observations. La valeur maximale ici représente la valeur maximale sur le domaine physique. Dans tous les cas, cette valeur maximale est normalisée par son équivalent calculé sur l'ébauche afin de quantifier l'impact de la méthode utilisée. L'interprétation des valeurs normalisées est similaire à celle des valeurs moyennes normalisées.

Remarque 5.3.1 *Il faut garder à l'esprit que nous utilisons une ébauche nulle (l'erreur sur l'ébauche est donc de 100%). Ce qui n'est pas nécessairement la bonne information pour tester une méthode dans le cas général. Cependant, dans les problèmes de flot optique en général, on n'a généralement pas d'autres choix qu'une ébauche nulle. Dans tous les cas, un bon usage n'est fait de l'ébauche que si on a une connaissance des statistiques d'erreurs associées. Enfin, l'effet d'une ébauche non nulle peut toujours être atténué par un paramètre de pondération adéquat sur le terme d'écart à l'ébauche de la fonction coût.*

5.3.2 Erreur angulaire

Pour le cas particulier de l'erreur angulaire, nous avons utilisé une version modifiée de l'erreur angulaire telle que proposée par Barron [Barron et al.(1994)]. Idéalement, l'erreur angulaire entre un vecteur et son approximation est définie par l'angle entre les deux vecteurs. Elle se définit comme :

$$ae(\mathbf{v}^a) = \arccos \left(\frac{\langle \mathbf{v}^a, \mathbf{v}^t \rangle}{\|\mathbf{v}^a\| \|\mathbf{v}^t\|} \right), \quad (5.33)$$

où \mathbf{v}^a représente le vecteur estimé, \mathbf{v}^t le vrai vecteur et ae une abréviation de l'expression anglaise angular error. \arccos est la fonction cosinus inverse définie sur l'intervalle $[-1, 1]$. Lorsqu'on est en dimension 2 avec le vecteur \mathbf{v} défini par $\mathbf{v} = v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2$ où $\{\mathbf{e}_1, \mathbf{e}_2\}$ représente la base orthonormée de l'espace vectoriel d'intérêt, la formulation de l'erreur angulaire de l'équation (5.33) se réécrit alors :

$$ae(\mathbf{v}^a) = \arccos \left(\frac{v_1^a v_1^t + v_2^a v_2^t}{\sqrt{(v_1^{a2} + v_2^{a2})(v_1^{t2} + v_2^{t2})}} \right). \quad (5.34)$$

Telle que définie par les équations (5.33) et (5.34), l'erreur angulaire n'a de sens que pour des vecteurs non nuls, c'est à dire $\mathbf{v}^t \neq 0$ et $\mathbf{v}^a \neq 0$. Barron et ses co-auteurs [Barron et al.(1994)]

proposent d'ajouter une composante unitaire au vecteur et de passer en dimension supérieure. Si on se trouve initialement en dimension 2, on définit alors le vecteur modifié :

$$\mathbf{w} = \mathbf{w}(\mathbf{v}) = v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 + \mathbf{e}_3, \quad (5.35)$$

$\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ définissant une base orthonormée. L'erreur angulaire devient :

$$bae(\mathbf{v}^a) = \arccos \left(\frac{\langle \mathbf{w}^a, \mathbf{w}^t \rangle}{\|\mathbf{w}^a\| \|\mathbf{w}^t\|} \right) = \arccos \left(\frac{v_1^a v_1^t + v_2^a v_2^t + 1}{\sqrt{(v_1^{a2} + v_2^{a2} + 1)(v_1^{t2} + v_2^{t2} + 1)}} \right), \quad (5.36)$$

où *ae* une abréviation de l'expression anglaise Barron angular error. Avec cette définition, l'estimation de l'erreur angulaire est très peu précise lorsque l'un des vecteurs est de petite norme. Pour faire face à ce problème, nous proposons d'ajouter un paramètre multiplicatif très petit ($0 < \epsilon \ll 1$) à la composante de la dimension supplémentaire ajoutée par Barron. Le vecteur modifié s'écrit alors :

$$\mathbf{w} = \mathbf{w}(\mathbf{v}) = v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 + \epsilon \mathbf{e}_3. \quad (5.37)$$

Un paramètre suffisamment petit est mieux approprié pour l'estimation de l'erreur angulaire même en présence de vecteurs de petite norme. En effet, avec cette formulation, l'erreur angulaire est définie dans tous les cas et la modification reste négligeable si le paramètre est suffisamment petit ; l'estimation de l'erreur est alors suffisamment précise dans tous les cas. L'erreur angulaire modifiée (notée *mae* de l'expression anglaise modified angular error en anglais) s'écrit :

$$mae(\mathbf{v}^a) = \arccos \left(\frac{\langle \mathbf{w}^a, \mathbf{w}^t \rangle}{\|\mathbf{w}^a\| \|\mathbf{w}^t\|} \right) = \arccos \left(\frac{v_1^a v_1^t + v_2^a v_2^t + \epsilon^2}{\sqrt{(v_1^{a2} + v_2^{a2} + \epsilon^2)(v_1^{t2} + v_2^{t2} + \epsilon^2)}} \right). \quad (5.38)$$

Pour toutes les expériences présentées dans ce document, nous avons choisit :

$$\epsilon = 5.10^{-4} \max(\|\mathbf{v}^t(\mathbf{x})\|), \mathbf{x} \in \Omega, \quad (5.39)$$

où la fonction *max* retourne le maximum d'une tableau de scalaires. Nous supposons ainsi que tout déplacement dont le rapport de l'amplitude à celui du plus grand déplacement dans l'écoulement est inférieur à 5.10^{-4} peut être considéré comme négligeable.

5.3.3 Expérience sans régularisation

Une telle expérience n'a théoriquement aucun sens pour un problème mal posé mais numériquement elle est réalisable ; elle est utilisée ici afin d'étudier l'impact des différents termes de régularisation. La figure (5.1) montre l'évolution au cours des itérations des fonctions de diagnostic normalisées pour une expérience sans régularisation. Les fonctions de diagnostic sur le terme d'écart aux observations, l'erreur sur le champ de vitesse estimé et l'erreur angulaire décroissent au cours des itérations, l'erreur sur la vorticit e quant  a elle croit plut ot. La fonction c ot n'est pas repr esent ee  etant donn e qu'elle est restreinte au terme d' ecart aux observations. A la convergence, l'erreur angulaire est de 43%, l'erreur sur le champ de vitesse estim e est de 58% et l'erreur sur la vorticit e est de 410%. Le terme d' ecart aux observations de la fonction

coût est de 10^{-5} . On note une estimation plutôt bonne en termes de décroissance de la fonction coût. L'erreur sur le champ estimé ainsi que l'erreur angulaire évoluent de façon similaire et dans le même sens que la fonction coût, mais la décroissance est plus modeste. L'erreur sur la vorticité quant à elle explose. On peut en déduire que même si la décroissance de la fonction coût est synonyme d'amélioration du champ de vitesse estimé en termes d'erreur moyenne et d'erreur angulaire, la principale caractéristique de l'écoulement (la vorticité) ne suit pas. Le graphique de droite (courbe des maxima des fonctions de diagnostic) montre que bien qu'il y ait une amélioration au niveau global (courbe des moyennes) sur certains diagnostics, il y a plutôt une détérioration au niveau des valeurs maximales de fonctions de diagnostic.

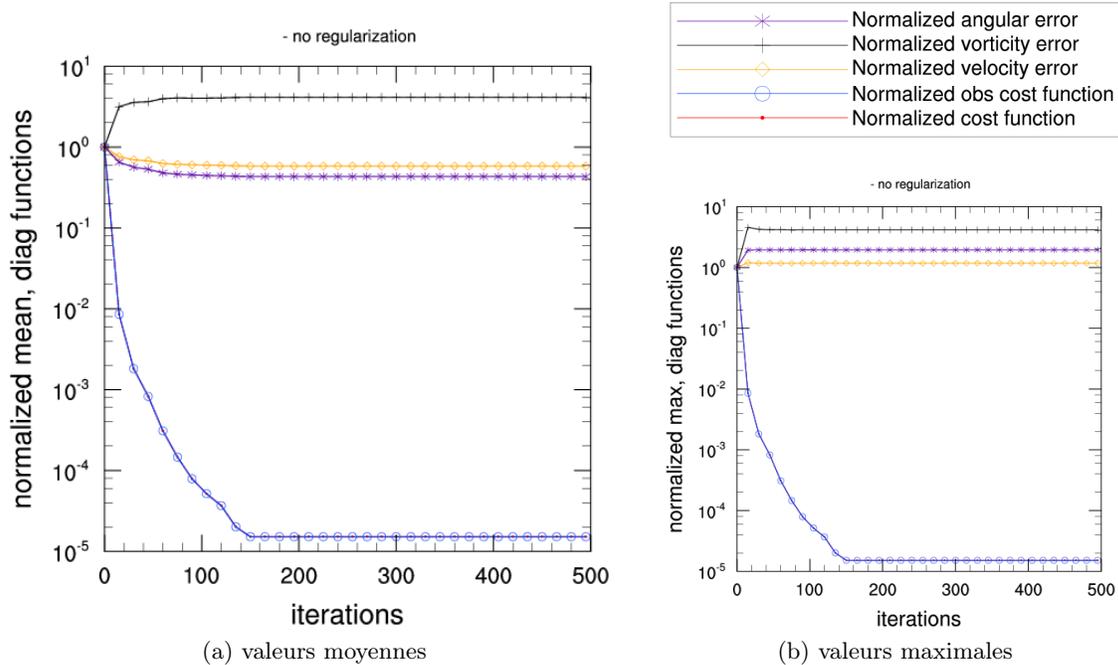


FIGURE 5.1 – Estimation du mouvement sans régularisation : évolution des fonctions de diagnostic normalisées en fonction des itérations ; valeurs moyennes (a) et valeurs maximales (b).

5.3.4 Evolution des diagnostics en fonction du paramètre de pondération

Ici, nous étudions le comportement des fonctions de diagnostic à convergence en fonction du paramètre de pondération pour différents termes de régularisation. L'objectif est d'avoir une idée de la difficulté relative du choix du paramètre optimal de pondération. Les courbes sont tracées en coordonnées logarithmiques pour mieux lire les détails sur les faibles valeurs de la fonction coût. Les autres fonctions de diagnostic présentent moins de détails mais suffisamment pour illustrer l'évolution des courbes. La figure (5.2) montre l'évolution des fonctions de diagnostic en fonction du paramètre de pondération avec régularisation de Tikhonov. Sur le graphique sont représentées les valeurs atteintes par les fonctions de diagnostic quand la méthode a convergé. En abscisse nous avons le paramètre de pondération. En se basant sur l'évolution de la fonction coût et du terme d'écart aux observations, on peut diviser la courbe en deux parties :

- la première partie correspond aux faibles valeurs (jusqu'à $\sim 10^{-5}$) du paramètre de

pondération où les courbes de la fonction coût et du terme d'écart aux observations sont presque identiques. Dans cette partie du graphique, l'effet de la régularisation est négligeable et la fonction coût est dominée par le terme d'écart aux observations d'où la superposition des deux courbes. Les autres fonctions de diagnostic sont quasi constantes et coïncident avec les valeurs à convergence obtenues sans régularisation comme l'illustre la figure (5.1) ;

- la seconde partie correspond aux grandes valeurs du paramètre de pondération où la fonction coût passe sensiblement au dessus du terme d'écart aux observations. La fonction coût est progressivement dominée par le terme de régularisation. Les fonctions de diagnostic notamment l'erreur sur le champ estimé et l'erreur angulaire croissent et se rapprochent des diagnostics calculés sur l'ébauche.

A la jonction de ces deux parties du graphique, on a les valeurs pour lesquelles le terme d'écart aux observations et le terme de régularisation s'équilibrent. Ce comportement est en accord avec ce que l'on attend de la pénalisation.

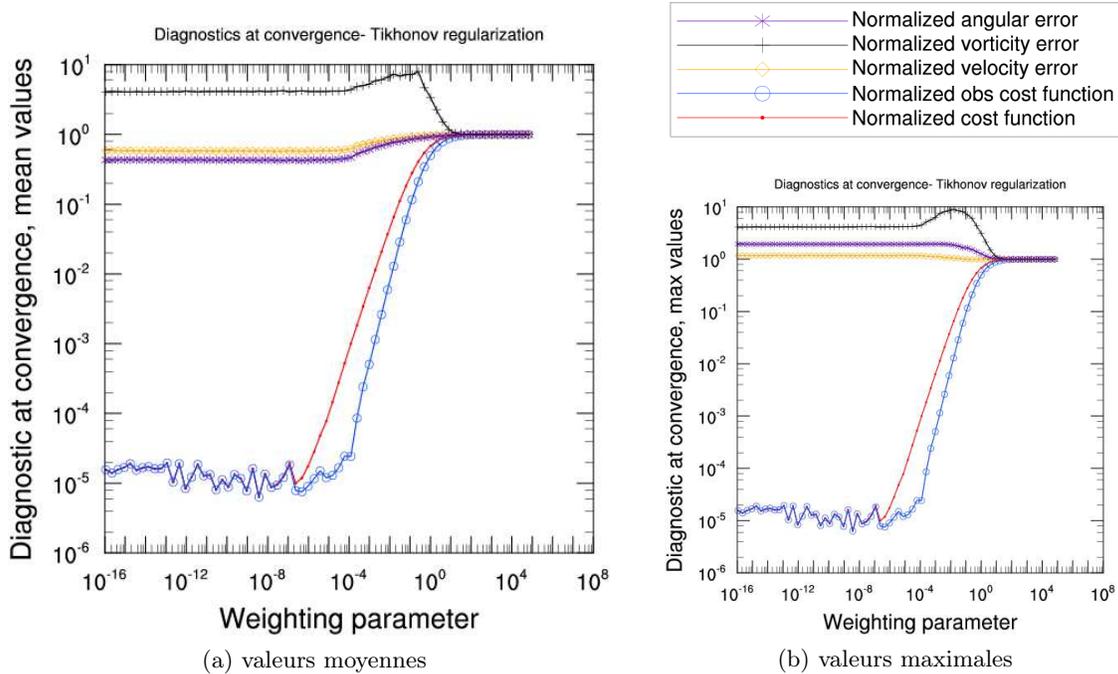


FIGURE 5.2 – Estimation du mouvement avec régularisation de Tikhonov : évolution des fonctions de diagnostic à convergence en fonction du paramètre de pondération ; valeurs moyennes (a) et valeurs maximales (b).

Les figures (5.3) et (5.4) montrent l'évolution des fonctions de diagnostic (en fonction du paramètre de pondération) avec les régularisations respectives :

- homogène de Horn et Schunck par pénalisation du gradient du champ ;
- pénalisation du rotationnel du champ ;
- pénalisation de la divergence du champ ;
- pénalisation du gradient du rotationnel du champ ;
- pénalisation du gradient de la divergence du champ ;
- isotropique basée sur les données ;

- isotropique basée sur le champ.

Ces deux figures illustrent le comportement général observé. L'annexe (E) montre des illustrations sur le comportement des autres termes de régularisation notamment :

- pénalisation du rotationnel du champ ;
- pénalisation de la divergence du champ ;
- pénalisation du gradient du rotationnel du champ ;
- pénalisation du gradient de la divergence du champ ;
- isotropique basée sur le champ.

De la même façon que dans le cas de la régularisation de Tikhonov, les graphiques représentent les valeurs atteintes par les fonctions de diagnostic quand la méthode a convergé. En abscisse nous avons le paramètre de pondération. Les courbes sont tracées avec une échelle logarithmique. Tous ces graphiques montrent des comportements presque identiques sur l'erreur commise sur le champ estimé et l'erreur angulaire.

En se basant sur l'évolution de la fonction coût et du terme d'écart aux observations, on peut diviser la courbe en deux parties comme dans le cas de la régularisation de Tikhonov. Mais on ne peut pas se limiter au comportement de la fonction coût. En effet les autres fonctions de diagnostic notamment l'erreur sur le champ estimé et l'erreur angulaire suivent une évolution très différente de celle de la régularisation de Tikhonov. En prenant en compte le comportement de l'erreur sur le champ estimé, tous les graphiques peuvent alors se diviser en trois principales parties (les limites de chaque partie dépendant du type de régularisation) :

- la première partie correspond aux faibles valeurs du paramètre de pondération où on note un comportement similaire à celui observé avec la régularisation de Tikhonov. Les courbes de la fonction coût et du terme d'écart aux observations sont presque identiques. Dans cette partie du graphique, l'effet de la régularisation est négligeable et la fonction coût est dominée par le terme d'écart aux observations d'où la superposition des deux courbes. Les autres fonctions de diagnostic sont quasi constantes. La constante correspondant dans chaque cas à la valeur obtenue sans régularisation lorsque l'algorithme de minimisation a convergé, (voir figure (5.1)) ;
- la deuxième partie avec les valeurs intermédiaires du paramètre de régularisation où l'erreur sur le champ de vitesse estimé passe en deçà des valeurs optimales obtenues sans régularisation. Dans cette région du graphique, la courbe de la fonction coût se détache sensiblement de la courbe du terme d'écart aux observations. La fonction coût est donnée par un équilibre entre le terme d'écart aux observations et le terme de régularisation. Les courbes des fonctions de diagnostic décrivent une concavité. Il est évident que le paramètre optimal se trouve au minimum de cette vallée. Cette dernière est plus ou moins mise en évidence selon la régularisation utilisée. Dans tous les cas, la concavité est moins marquée sur la vorticit . L'effet de la régularisation y est moins accentu  que sur l'erreur angulaire et l'erreur sur le champ estimé ;
- la troisième partie correspond aux grandes valeurs du paramètre de pondération du terme de régularisation. On note ici un comportement similaire à celui de la deuxième partie de la figure (5.2) de la régularisation de Tikhonov. Les fonctions de diagnostic notamment l'erreur moyenne sur le champ estimé et l'erreur angulaire moyenne croissent avec le paramètre de pondération et tendent vers une borne correspondant aux diagnostics sur l'ébauche. Pour ces valeurs du paramètre de pondération, la fonction coût est essentiellement dominée par le terme de régularisation qui limite la minimisation.

La deuxième partie des graphiques montre une différence significative par rapport à la régularisation de Tikhonov. C'est ce qui fait la différence par rapport à ce terme qui sert

essentiellement à rendre le problème bien posé. En plus d'agir sur le caractère mal posé du problème, les autres types de régularisation améliorent de façon significative le résultat moyennant le choix d'un bon paramètre de pondération.

Les courbes moyennes montrent dans tous les cas une amélioration par rapport à la régularisation de Tikhonov avec un bon choix du paramètre de pondération. Dans le cas des valeurs maximales on note plutôt une détérioration ou une amélioration peu significative sauf dans le cas de la régularisation homogène de Horn et Schunck par pénalisation du gradient ou l'effet est bien visible sur la courbe d'erreur sur le champ estimé.

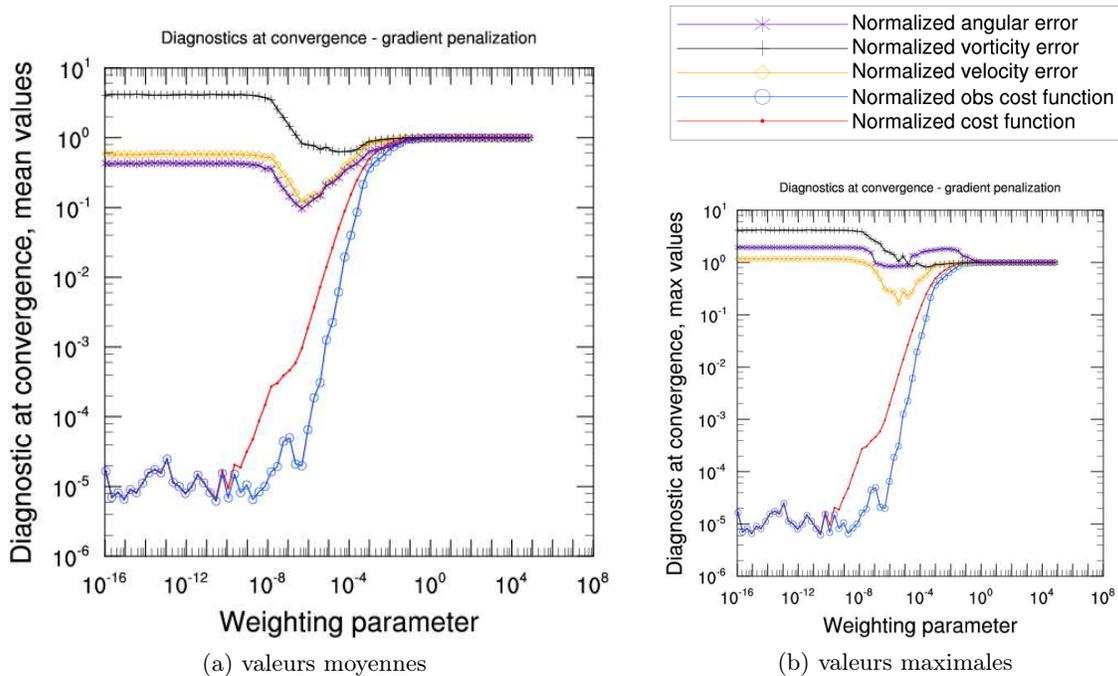


FIGURE 5.3 – Estimation du mouvement avec pénalisation du gradient : évolution des fonctions de diagnostic à convergence en fonction du paramètre de pondération ; valeurs moyennes (a) et valeurs maximales (b).

5.3.5 Choix du paramètre de pondération

L'étude expérimentale avec illustration sur le comportement des fonctions de diagnostic montre qu'il est difficile, voire impossible, de choisir le paramètre optimal de pondération de façon déterministe. En effet, sa position n'est visible qu'à partir des courbes d'erreurs sur le champ estimé. Cependant dans les applications réelles, on ne dispose pas du champ vrai pour calculer cette erreur. Les seules données accessibles dans une application réelle sont la fonction coût et ses composantes. Cependant, les illustrations graphiques montrent qu'elles ne donnent presque aucune information sur la localisation du paramètre optimal. Elles auraient permis de le déterminer à partir d'une étude au second ordre. Même dans un cas académique où l'on connaîtrait la vérité sur la variable de contrôle, la détermination de ce paramètre resterait difficile. En effet, bien que sa position puisse être localisée sur les graphiques, nous sommes en présence de fonctions non convexes. Il faudrait au préalable pouvoir localiser la région convexe dans laquelle se trouve ce paramètre optimal. Cette localisation serait tout

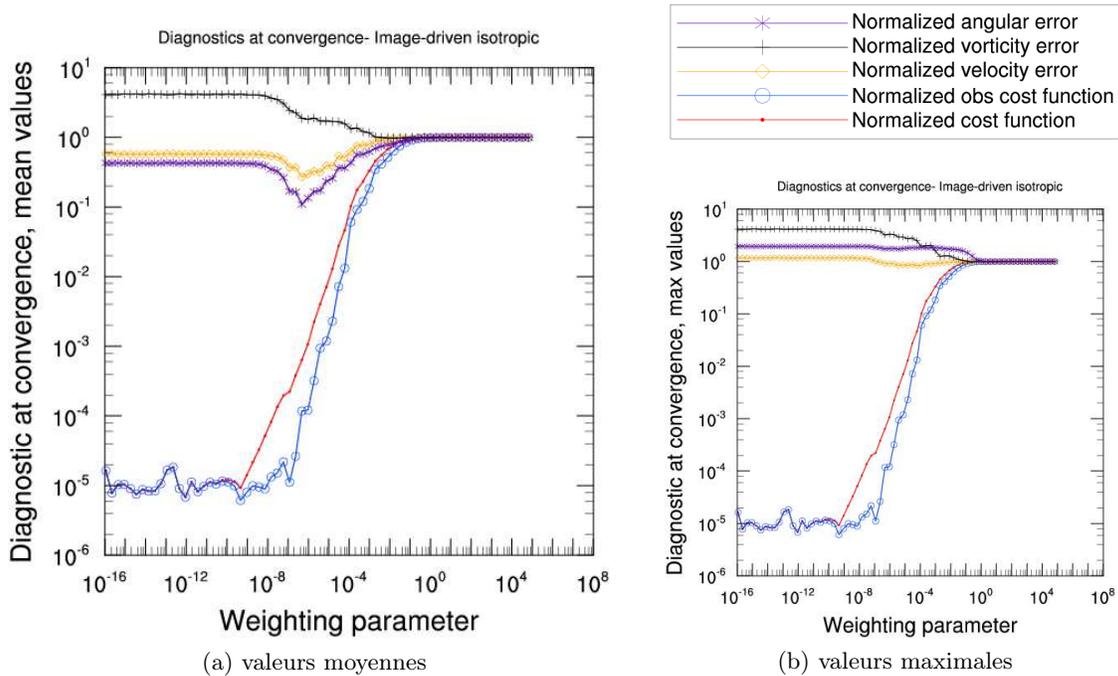


FIGURE 5.4 – Estimation du mouvement avec régularisation isotropique basée sur les données : évolution des fonctions de diagnostic à convergence en fonction du paramètre de pondération ; valeurs moyennes (a) et valeurs maximales (b).

aussi difficile à cause de la taille très réduite de ces intervalles. Dans les exemples traités ici, ces intervalles ont des longueurs de l'ordre de 10^{-9} pour la régularisation au second ordre (gradient du rotationnel et gradient de divergence, voir l'annexe (E) pour les graphiques) et de l'ordre de 10^{-5} à 10^{-4} dans les autres cas. La seule méthode de détermination du paramètre optimal reste donc empirique.

5.3.6 Evolution des fonctions de diagnostic au cours de la minimisation

Nous étudions dans ce paragraphe le comportement des fonctions de diagnostic au cours de la minimisation avec le *meilleur* paramètre de pondération dans le cas des principaux termes de régularisation de la littérature. Ce meilleur paramètre est déterminé à partir de l'étude précédente. L'étude du comportement des fonctions de diagnostic au cours de la minimisation nous permet de nous faire une idée du lien entre la décroissance de la fonction coût et l'amélioration de la solution au problème. L'étude de la figure (5.1) nous a permis de conclure que dans le cas de l'expérience sans régularisation, la décroissance de la fonction coût était synonyme d'amélioration du résultat au moins en termes de l'erreur sur le champ et de l'erreur angulaire. Avec la régularisation de Tikhonov, on obtient exactement le même comportement.

La figure (5.5) montre l'évolution des fonctions de diagnostic au cours des itérations avec la régularisation par pénalisation du gradient associé au meilleur paramètre. Les illustrations sur les autres termes de régularisation se trouvent en annexe. Les courbes sont tracées avec une échelle logarithmique en ordonnées. On peut noter des oscillations sur la courbe d'évolution du terme d'écart aux observations de la fonction coût. Pour expliquer ces oscillations qui ne sont pas observées sur la fonction coût globale, il faut se rappeler que la fonction coût globale

est constituée de plusieurs termes dont un terme d'écart aux observations et un ou plusieurs termes de régularisation. L'algorithme itératif de minimisation calcule une suite d'approximations convergeant vers la solution minimisant la fonction coût. A chaque itération, la nouvelle approximation est la somme d'une composante améliorant le terme d'écart aux observations et d'une composante améliorant le(s) terme(s) de régularisation. Les deux composantes pouvant être contradictoires. Ainsi la détérioration apportée par la composante de régularisation sur le terme d'écart aux observations peut dominer sur l'amélioration apportée par le terme associé. Ce qui crée des oscillations sur le terme d'écart aux observations. Un comportement similaire est observé avec les autres termes de régularisation.

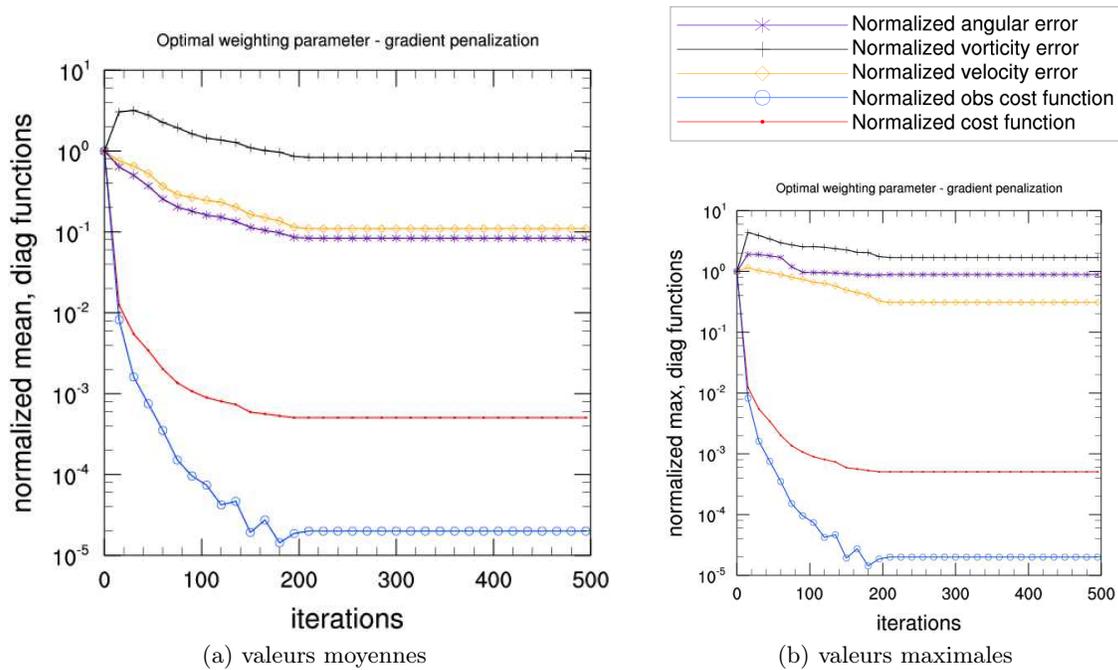


FIGURE 5.5 – Estimation du mouvement, régularisation par pénalisation du gradient avec le meilleur paramètre de pondération : évolution des fonctions de diagnostic normalisées en fonction des itérations ; valeurs moyennes (a) et valeurs maximales (b).

Les figures (5.6, 5.7, 5.8, 5.9) montrent l'évolution comparée du terme d'écart aux observations, de l'erreur sur le champ estimé, de l'erreur angulaire et de l'erreur sur la vorticité avec les contraintes suivantes :

- aucune régularisation ;
- régularisation homogène de Horn et Schunck par pénalisation du gradient ;
- régularisation second ordre de Suter (gradient de divergence et du rotationnel) ;
- régularisation isotropique basée sur les données ;
- régularisation isotropique basée sur le champ.

Le terme de régularisation étant pondéré dans chaque cas par le paramètre optimal. Dans le cas de la régularisation de Suter, les paramètres optimaux obtenus avec l'étude sur le terme de pénalisation du gradient du rotationnel et le terme de pénalisation du gradient de divergence.

Sur la figure (5.6), on peut constater que l'évolution du terme d'écart aux observations est comparable pour tous les termes de régularisation. On note des oscillations sauf dans le cas de l'expérience sans régularisation ; ce qui témoigne des effets contradictoires du terme d'écart aux

observations et du terme de régularisation. La décroissance avec la régularisation isotropique basée sur les données et la régularisation isotropique basée sur le champ s'arrête très vite par rapport aux autres cas. Le terme de régularisation dans ces cas domine la fonction coût et la minimisation est arrêtée comme le confirme l'évolution des autres fonctions de diagnostic (Figures 5.7, 5.8 et 5.9). Dans les autres cas (pénalisation du gradient et second ordre de Suter) le terme de l'écart aux observations atteint la valeur minimale obtenue sans régularisation. L'impression de l'arrêt très brutal de la minimisation est due à l'échelle logarithmique utilisée en ordonnées.

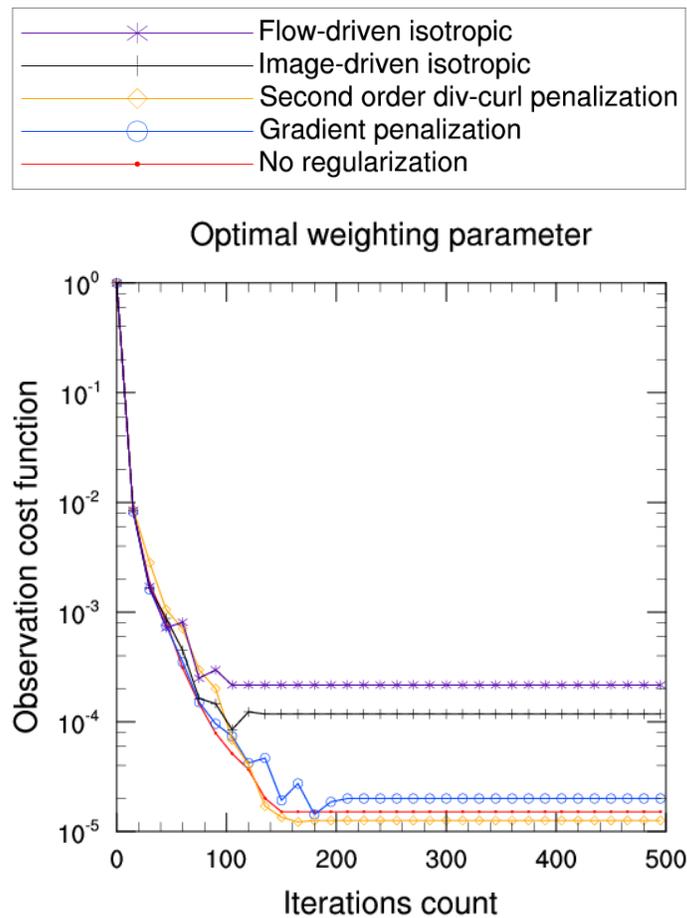


FIGURE 5.6 – Estimation du mouvement avec différents termes de régularisation, comparaison de l'évolution du terme d'écart aux observations au cours de la minimisation

Sur le graphique de gauche (moyennes normalisées) de la figure (5.7), la progression de l'erreur sur le champ estimé est comparable dans les différents cas. A convergence, on a une erreur sur le champ estimé de l'ordre de :

- 60% sans régularisation ;
- 30% avec la régularisation isotropique basée sur les données et la régularisation isotropique basée sur le champ ;
- 10% avec la régularisation par pénalisation du gradient et la régularisation au second ordre de Suter.

En terme de valeurs maximales, sur le graphique de droite de la figure (5.7), on note à la

convergence une erreur sur le champ estimé de l'ordre de :

- 120% sans régularisation ;
- 90% avec la régularisation isotropique basée sur les données et la régularisation isotropique basée sur le champ ;
- 30% avec la régularisation par pénalisation du gradient et la régularisation au second ordre de Suter.

Toutes ces valeurs sont récapitulées dans la table (5.2).

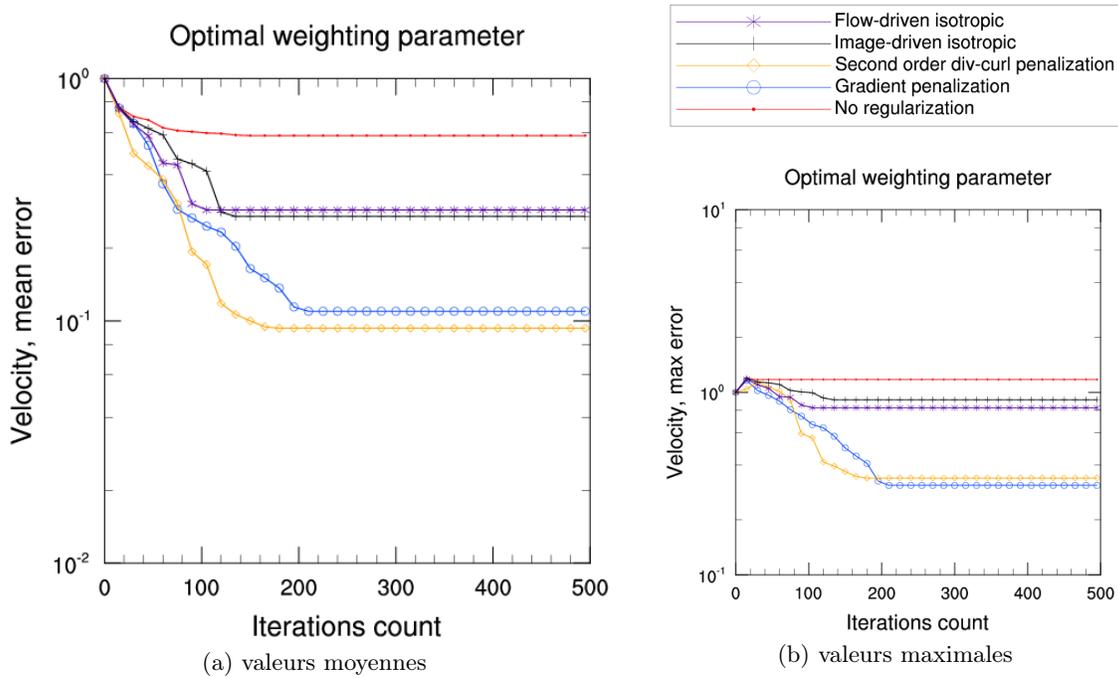


FIGURE 5.7 – Estimation du mouvement avec différents termes de régularisation, comparaison de l'évolution de l'erreur moyenne normalisée sur le champ estimé au cours de la minimisation ; valeurs moyennes (a) et valeurs maximales (b).

En terme d'erreur angulaire (figure 5.8), l'évolution des courbes est similaire à celle observée avec l'erreur sur le champ estimé. Les valeurs moyennes étant légèrement plus faibles dans le cas de l'erreur angulaire. On note à convergence une erreur moyenne de l'ordre de :

- 40% sans régularisation ;
- 20% avec la régularisation isotropique basée sur le champ ;
- 10% avec la régularisation par pénalisation du gradient, la régularisation au second ordre de Suter et la régularisation isotropique basée sur les données.

En termes d'erreur maximale cependant, il n'y a pas de comparaison possible avec l'erreur sur le champ estimé. A convergence, on note une erreur maximale de l'ordre de 90% avec la régularisation par pénalisation du gradient et 150 à 200% dans les autres cas. Toutes ces valeurs sont résumées dans la table (5.2).

En terme de vorticit  (figure 5.9), l'am lioration reste mitig e m me en termes de moyenne. Au d but de la minimisation, l'erreur moyenne sur la vorticit  croit dans tous les cas avant de d croitre dans certains cas. A convergence, on note une erreur moyenne (graphique de gauche de la figure 5.9) de l'ordre de :

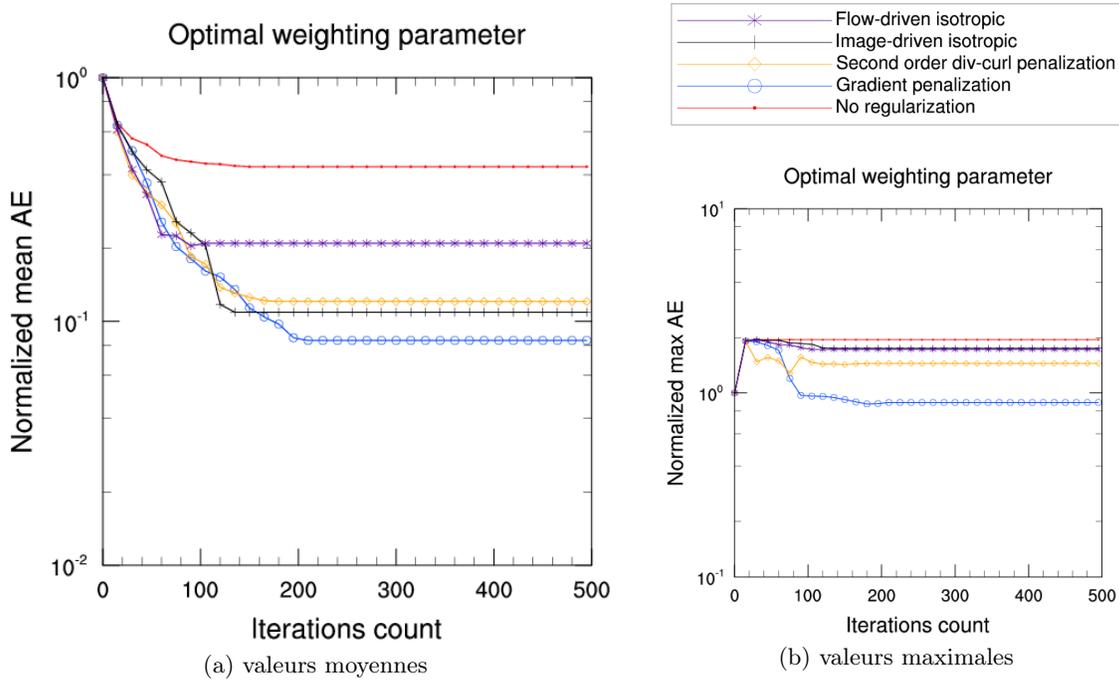


FIGURE 5.8 – Estimation du mouvement avec différents termes de régularisation, comparaison de l'évolution de l'erreur angulaire moyenne normalisée au cours de la minimisation ; valeurs moyennes (a) et valeurs maximales (b).

- 400% sans régularisation ;
- 200% avec la régularisation isotropique basée sur le champ et la régularisation isotropique basée sur les données ;
- 80% avec la régularisation par pénalisation du gradient et la régularisation au second ordre de Suter.

En valeurs maximales (graphique de droite de la figure 5.9), seule la régularisation au second ordre de Suter converge avec une erreur légèrement inférieure à 100%. Toutes ces valeurs sont résumées dans la table (5.2).

5.3.7 Comparaison des diagnostics pour différents termes de régularisation

Ce paragraphe présente un récapitulatif de l'étude comparative des fonctions de diagnostic pour différents termes de régularisation. Le tableau (5.2) récapitule les valeurs des différentes fonctions de diagnostic à convergence. Toutes les valeurs sont normalisées par rapport à leurs équivalents calculés sur l'ébauche sauf dans le cas de l'erreur angulaire où nous donnons les valeurs normalisées (E A normalisée) et les valeurs réelles (E A). La colonne moyenne donne les valeurs moyennes sur le domaine physique et la colonne maxima les valeurs maximales sur le domaine. Toutes les valeurs sont données en pourcentage sauf les valeurs réelles de l'erreur angulaire qui sont données en degré et le terme d'écart aux observations dont les valeurs relatives sont données directement.

De cette étude, nous pouvons tirer quelques remarques de l'usage de l'approche classique de régularisation :

TABLE 5.2 – Estimation du mouvement avec différents termes de régularisation, comparaison des fonctions de diagnostic à convergence

Écart aux observations		
Terme de régularisation	valeur moyenne	valeur maximale
sans régularisation	1.51e-05	-
Horn & Schunck	1.99e-05	-
Suter	1.26e-05	-
Isotropique basée sur les données	1.17e-04	-
Isotropique basée sur le champ	2.16e-04	-

Diagnostics d'erreur relatives		
Terme de régularisation	valeur moyenne(%)	valeur maximale(%)
Erreur sur le champ de vitesse		
sans régularisation	58	117
Horn & Schunck	10	30
Suter	9	33
Isotropique basée sur les données	26	91
Isotropique basée sur le champ	28	82
Erreur sur la vorticité		
sans régularisation	411	414
Horn & Schunck	83	169
Suter	68	95
Isotropique basée sur les données	188	321
Isotropique basée sur le champ	184	306
Erreur angulaire normalisée		
sans régularisation	43	194
Horn & Schunck	8	88
Suter	12	144
Isotropique basée sur les données	10	174
Isotropique basée sur le champ	20	172
Erreur angulaire en degré		
Terme de régularisation	valeur moyenne(°)	valeur maximale(°)
sans régularisation	25	173
Horn & Schunck	5	79
Suter	7	129
Isotropique basée sur les données	6	156
Isotropique basée sur le champ	12	153

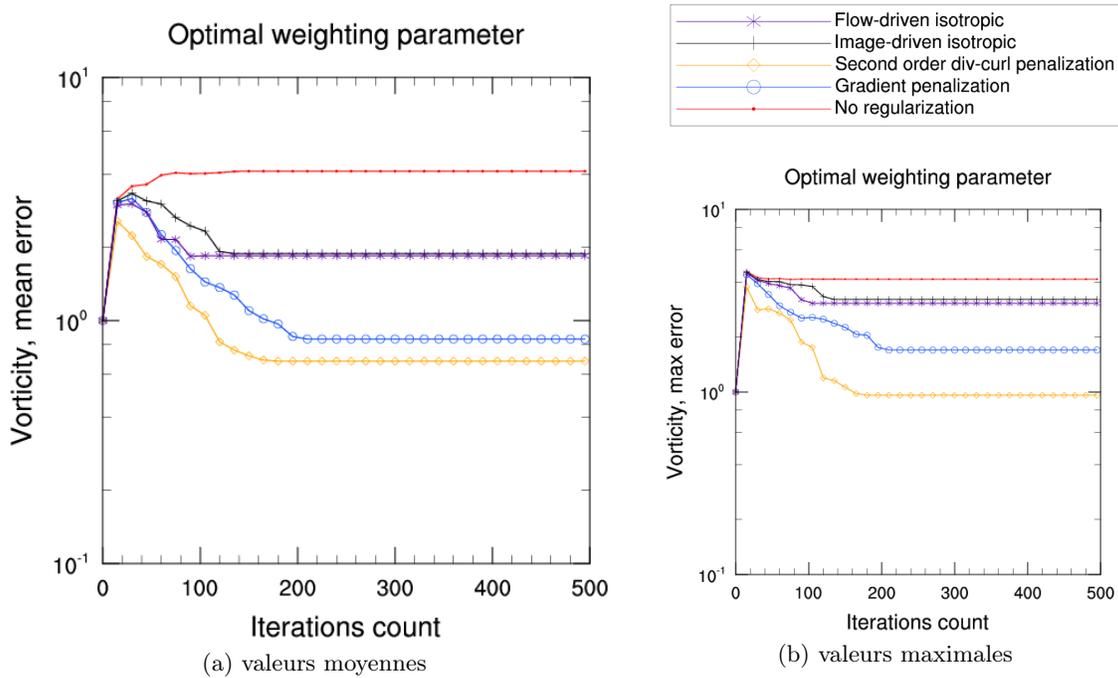


FIGURE 5.9 – Estimation du mouvement avec différents termes de régularisation, comparaison de l'évolution de l'erreur moyenne normalisée sur la vorticité au cours de la minimisation ; valeurs moyennes (a) et valeurs maximales (b).

1. La décroissance de la fonction coût n'est pas forcément synonyme d'une amélioration pour toutes les caractéristiques de la variable de contrôle estimée. L'erreur sur la vorticité en est une illustration pour les exemples présentés ici. Le meilleur des cas dans notre étude a donné une erreur de 68%. Il est important de mentionner que ce meilleur résultat est obtenu au prix d'un nombre élevé d'itérations. Il faut donc être prudent sur le nombre d'itérations ;
2. Même si dans certains cas les valeurs moyennes des fonctions de diagnostic sur la variable de contrôle estimée sont satisfaisantes, il n'en n'est pas de même pour les valeurs maximales de ces mêmes fonctions. Dans notre étude, seule l'erreur maximale sur le champ de vitesse estimé n'est pas aggravée par les différentes techniques de régularisation étudiées ;
3. La régularisation peut se faire au prix de la dégradation du terme d'écart aux observations. Dans l'étude faite ici, nous observons ce comportement avec les oscillations sur les courbes d'évolution des termes d'écart aux observations en fonction des itérations.

Remarque 5.3.2 *Les aberrations observées avec certaines fonctions de diagnostic peuvent être dues aux approximations algorithmiques. En effet, avec les algorithmes du type quasi-Newton comme celui utilisé pour ces expériences, le pas de descente n'est pas calculé de façon exacte, seule une approximation est calculée. Mais on ne peut pas faire autrement, le calcul de la valeur exacte du pas de descente nécessiterait un temps de calcul astronomique.*

5.3.8 Interprétation mathématique et problèmes liés à l'usage de la régularisation par pénalisation

Interprétation mathématique de la régularisation par pénalisation

Les conclusions tirées de l'étude du comportement des fonctions de diagnostic nous amènent à nous poser des questions sur l'interprétation mathématique de la régularisation. Rappelons que le but de la régularisation est de contrôler le caractère mal posé des problèmes inverses. En présence d'un problème mal posé, la fonction d'écart aux observations n'est pas strictement convexe, elle dispose alors de plusieurs minima. C'est ce qu'on rencontre en général avec des opérateurs (modèle ou opérateur d'observation) non linéaires. Dans cette situation, en l'absence d'une bonne initialisation (notamment dans la région convexe contenant la meilleure solution si elle est unique), les algorithmes de minimisation convergent vers des minima locaux. Un bon terme de régularisation permet de rendre la fonction coût convexe. Les termes de régularisation déterministes sont à chercher dans l'ensemble des fonctions strictement convexes. Le terme de régularisation qui satisfait cette propriété est le terme de régularisation de Tikhonov qui pénalise la norme de l'écart à l'ébauche (qui peut être éventuellement nulle). Tous les autres termes de régularisation que nous avons étudiés ne définissent pas des fonctions strictement convexes. Ils s'annulent en effet sur les fonctions constantes pour les termes de régularisation du premier ordre et sur les polynômes de degré 1 pour les régularisations du second ordre. Ce sont des formes semi définies positives.

Problèmes liés à l'usage de la régularisation par pénalisation

L'usage de la régularisation restreint théoriquement l'espace de recherche des solutions au sous-espace de l'espace de contrôle sur lequel est défini le terme de régularisation utilisé. En particulier, on restreint l'espace de recherche des solutions au sous-espace de l'espace de contrôle sur lequel sont définies les dérivées premières pour la régularisation au premier ordre et les dérivées secondes pour la régularisation au second ordre. Dans la pratique avec les solveurs itératifs, cette considération n'est pas vérifiée. Comme nous l'avons dit plus haut, la mise à jour de la solution approximative à une itération donnée est faite à partir de deux composantes : une composante d'amélioration de l'écart aux observations et une composante d'amélioration de l'écart à la régularité. Si la composante d'amélioration de l'écart à la régularité permet de rester dans le bon ensemble, ce n'est pas nécessairement le cas pour la composante d'amélioration de l'écart aux observations. L'usage des techniques classiques de régularisation peut ainsi conduire à deux principaux problèmes :

1. l'échec du solveur dû à une solution candidate n'appartenant pas à l'espace approprié. Dans la pratique, ce problème est automatiquement contourné par l'usage des approximations numériques pour le calcul des dérivées sur des domaines discrets ;
2. La restriction de l'espace des solutions à des espaces très réguliers (espaces de Sobolev) tels que \mathcal{H}^1 (pour les régularisations du premier ordre) ou \mathcal{H}^2 (pour les régularisations du second ordre) élimine un certain nombre de solutions candidates qui pourraient être plus satisfaisantes.

Des travaux sur l'étude de ces problèmes qui peuvent être posés par l'usage de la régularisation n'abondent pas dans la littérature. Quand ils sont mentionnés, seules sont faites des hypothèses sur l'appartenance d'une éventuelle solution à l'espace défini par la restriction. Dans le cas de l'estimation de mouvement, Weickert et Schnörr [Weickert and Schnörr(2001)] justifient

ces hypothèses par le fait que les capteurs des caméras créent des images assez régulières et que des lissages par des convolutions gaussiennes doivent y être appliquées tel que prescrit par Barron et ses co-auteurs [Barron et al.(1994)]. Ainsi traitées, les images ne peuvent pas conduire à des solutions discontinues ou irrégulières.

5.3.9 Conclusion partielle

Suite à l'étude que nous avons faite dans cette section, nous pouvons retenir quelques principaux problèmes liés à l'usage de la régularisation par pénalisation :

- Si la régularisation permet d'améliorer la solution, le choix d'un bon paramètre de pondération du terme de régularisation reste très difficile et même impossible à partir des seules connaissances *a priori* ;
- l'amélioration apportée par la régularisation peut être limitée à certains diagnostics de la solution calculée, les autres diagnostics montrant plutôt une détérioration ;
- si l'impact de la régularisation peut être notable et positif en termes de valeurs moyennes sur l'erreur commise, il reste négligeable et même négatif en termes de valeurs maximales.
- le nombre d'itérations doit être minutieusement choisi afin d'espérer l'amélioration sur un maximum de fonctions de diagnostic ;
- l'usage de la régularisation restreint et quelquefois de façon artificielle l'espace de recherche des solutions. Une restriction qui n'est nécessairement pas prise en compte par les solveurs itératifs.

5.4 Régularisation par diffusion généralisée pour des problèmes inverses mal posés

5.4.1 Définition de la nouvelle approche

Proposition 5.4.1 *Soient :*

- $\mathcal{V}_\Phi = \{\mathbf{v} \in D_\Phi / \Phi(\mathbf{v}) = 0\}$ le sous-espace des éléments Φ -réguliers de l'espace de contrôle \mathcal{V} où Φ est une fonction de régularisation telle que définie à la section (5.2) et D_Φ son domaine de définition ;
- \mathcal{P} un projecteur sur \mathcal{V}_Φ (c'est à dire $\mathcal{P} \circ \mathcal{P} = \mathcal{P}$ et $Im\mathcal{P} = \mathcal{V}_\Phi$) défini comme suit :

$$\begin{aligned} \mathcal{P} : \mathcal{V} &\rightarrow \mathcal{V}_\Phi, \\ \mathbf{v} &\mapsto \mathcal{P}(\mathbf{v}). \end{aligned} \tag{5.40}$$

Nous proposons de formuler le terme de régularisation associé à la fonction de régularisation Φ et l'opérateur de projection \mathcal{P} comme :

$$J_r(v) = \frac{1}{2} \|\mathbf{v} - \mathcal{P}(\mathbf{v})\|_{\mathcal{V}}^2. \tag{5.41}$$

Interprétation Le terme de régularisation ainsi défini (équation 5.41) est utilisé pour pénaliser les déviations à la régularité définie par Φ . Avec cette formulation, l'espace de recherche des solutions aux problèmes inverses n'est plus restreint au domaine de définition de la fonction de régularisation.

Le terme de régularisation de l'équation (5.41) peut être réécrit comme :

$$J_r(v) = \frac{1}{2} \|\bar{\mathcal{P}}(\mathbf{v})\|_{\mathcal{V}}^2, \quad (5.42)$$

où $\bar{\mathcal{P}}$ défini par :

$$\begin{aligned} \bar{\mathcal{P}} : \mathcal{V} &\rightarrow \mathcal{V} \setminus \mathcal{V}_\Phi, \\ \mathbf{v} &\mapsto \bar{\mathcal{P}}(\mathbf{v}) = \mathbf{v} - \mathcal{P}(\mathbf{v}) \end{aligned} \quad (5.43)$$

est l'opérateur de projection sur $\mathcal{V} \setminus \mathcal{V}_\Phi$.

La suite consiste à définir l'opérateur de projection \mathcal{P} ou $\bar{\mathcal{P}}$ de façon appropriée.

5.4.2 Des fonctions de régularisation aux opérateurs de projection

Definition 5.4.1 Soient :

- $\mathbf{v}(\mathbf{x}) \in (\mathcal{L}^2(\Omega))^n$ l'espace de contrôle du système d'intérêt avec $\mathbf{x} \in \Omega$ la variable de l'espace physique du système ;
- Φ une fonction de régularisation telle que définie à la section (5.2) ;
- $\varphi(\mathbf{x})$ une fonction scalaire positive décrivant les caractéristiques de la variable d'état \mathbf{v} au point \mathbf{x} . Nous l'appellerons par la suite fonction de confiance. Elle est interprétée comme suit :
 - les petites valeurs sont synonymes de mauvaises connaissances ou peu de connaissances sur le champ \mathbf{v} au point \mathbf{x} ;
 - les grandes valeurs sont synonymes de bonne qualité du champ \mathbf{v} au point \mathbf{x} .

Nous définissons la Φ -projection de \mathbf{v} comme le minimum de la fonctionnelle d'énergie :

$$\varepsilon(\mathbf{u}) = \frac{1}{2} \int_{\Omega} \|\Phi(\mathbf{u}(\mathbf{x}))\|^2 + \varphi(\mathbf{x}) \|\mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x})\|^2 d\mathbf{x}. \quad (5.44)$$

Interprétation La fonctionnelle d'énergie $\varepsilon(\mathbf{u})$ est un équilibre entre la régularité définie par $\|\Phi(\mathbf{u}(\mathbf{x}))\|$ et la pénalisation de l'écart à la variable de contrôle $\|\mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x})\|$. La fonction de confiance permet de contrôler le poids relatif de chacun des termes dans la fonctionnelle d'énergie.

Usage dans la pratique sous les conditions d'existence et d'unicité données à la section (5.1.2), le minimum de cette fonctionnelle d'énergie $\text{Argmin}(\varepsilon)$ satisfait l'équation d'Euler-Lagrange donnée par :

$$\nabla_{\mathbf{u}} \varepsilon(\mathbf{u}) = 0. \quad (5.45)$$

La principale difficulté avec les problèmes non linéaires consiste à exprimer le gradient $\nabla \varepsilon$ de cette fonctionnelle. Si on parvient à exprimer explicitement ce gradient, on peut l'utiliser pour résoudre le problème du minimum grâce à des algorithmes de type descente de gradient. Une façon d'exprimer ce gradient $\nabla \varepsilon$ consiste à rendre explicite la dépendance linéaire de la dérivée de Gâteaux de la fonctionnelle par rapport au gradient. Si nous écrivons la norme $\|\cdot\|$ en termes de produit scalaire associé $\langle \cdot, \cdot \rangle$, la fonctionnelle ε s'écrit alors :

$$\varepsilon(\mathbf{u}) = \frac{1}{2} \int_{\Omega} \langle \Phi(\mathbf{u}(\mathbf{x})), \Phi(\mathbf{u}(\mathbf{x})) \rangle + \varphi(\mathbf{x}) \langle \mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x}), \mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x}) \rangle d\mathbf{x}. \quad (5.46)$$

On peut aussi l'écrire sous la forme :

$$\varepsilon(\mathbf{u}) = \varepsilon^r(\mathbf{u}) + \varepsilon^d(\mathbf{u}), \quad (5.47)$$

avec :

$$\varepsilon^r(\mathbf{u}) = \frac{1}{2} \int_{\Omega} \langle \Phi(\mathbf{u}(\mathbf{x})), \Phi(\mathbf{u}(\mathbf{x})) \rangle d\mathbf{x}; \quad (5.48)$$

$$\varepsilon^d(\mathbf{u}) = \frac{1}{2} \int_{\Omega} \varphi(\mathbf{x}) \langle \mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x}), \mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x}) \rangle d\mathbf{x}, \quad (5.49)$$

où l'exposant r dénote le terme de régularisation et d le terme d'attache aux données. La dérivée de Gâteaux $\hat{\varepsilon}$ de la fonctionnelle ε dans la direction \mathbf{h} (pris dans l'espace approprié selon les cas) se définit par :

$$\hat{\varepsilon}(\mathbf{u}, \mathbf{h}) = \lim_{\alpha \rightarrow 0} \frac{\varepsilon(\mathbf{u} + \alpha \mathbf{h}) - \varepsilon(\mathbf{u})}{\alpha}. \quad (5.50)$$

Si la fonction ε est différentiable en \mathbf{u} (nous supposons que c'est le cas ici), alors :

$$\hat{\varepsilon}(\mathbf{u}, \mathbf{h}) = \langle \nabla \varepsilon(\mathbf{u}), \mathbf{h} \rangle; \quad (5.51)$$

$$\begin{aligned} \frac{\varepsilon^r(\mathbf{u} + \alpha \mathbf{h}) - \varepsilon^r(\mathbf{u})}{\alpha} &= \frac{1}{2} \int_{\Omega} \frac{\langle \Phi(\mathbf{u}(\mathbf{x}) + \alpha \mathbf{h}(\mathbf{x})), \Phi(\mathbf{u}(\mathbf{x}) + \alpha \mathbf{h}(\mathbf{x})) \rangle}{\alpha} d\mathbf{x} \\ &\quad - \frac{1}{2} \int_{\Omega} \frac{\langle \Phi(\mathbf{u}(\mathbf{x})), \Phi(\mathbf{u}(\mathbf{x})) \rangle}{\alpha} d\mathbf{x}. \end{aligned} \quad (5.52)$$

Après développement, la prise en compte de la linéarité de Φ comme mentionnée dans la remarque (5.2.1) et la considération que $\alpha \rightarrow 0$, nous obtenons :

$$\hat{\varepsilon}^r(\mathbf{u}, \mathbf{h}) = \int_{\Omega} \langle \Phi(\mathbf{u}(\mathbf{x})), \Phi(\mathbf{h}(\mathbf{x})) \rangle d\mathbf{x}, \quad (5.53)$$

qui peut être réécrit comme :

$$\hat{\varepsilon}^r(\mathbf{u}, \mathbf{h}) = \int_{\Omega} \langle \Phi^* \circ \Phi(\mathbf{u}(\mathbf{x})), \mathbf{h}(\mathbf{x}) \rangle d\mathbf{x}, \quad (5.54)$$

où Φ^* est l'opérateur adjoint de la fonction de régularisation Φ associée au produit scalaire $\langle \cdot, \cdot \rangle$.

Pour ε^d , nous avons :

$$\begin{aligned} \frac{\varepsilon^d(\mathbf{u} + \alpha \mathbf{h}) - \varepsilon^d(\mathbf{u})}{\alpha} &= \frac{1}{2} \int_{\Omega} \frac{\varphi(\mathbf{x}) \langle \mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x}) + \alpha \mathbf{h}(\mathbf{x}), \mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x}) + \alpha \mathbf{h}(\mathbf{x}) \rangle}{\alpha} d\mathbf{x} \\ &\quad - \frac{1}{2} \int_{\Omega} \frac{\varphi(\mathbf{x}) \langle \mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x}), \mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x}) \rangle}{\alpha} d\mathbf{x}. \end{aligned} \quad (5.55)$$

Après avoir développé et considéré que $\alpha \rightarrow 0$, $\hat{\varepsilon}^d$ peut se réécrire comme :

$$\hat{\varepsilon}^d(\mathbf{u}, \mathbf{h}) = \int_{\Omega} \varphi(\mathbf{x}) \langle \mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x}), \mathbf{h}(\mathbf{x}) \rangle d\mathbf{x}. \quad (5.56)$$

La dérivée de Gâteaux de la fonctionnelle ε peut donc s'écrire :

$$\hat{\varepsilon}(\mathbf{u}, \mathbf{h}) = \int_{\Omega} \langle \Phi^* \circ \Phi(\mathbf{u}(\mathbf{x})) + \varphi(\mathbf{x})(\mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x})), \mathbf{h}(\mathbf{x}) \rangle d\mathbf{x}. \quad (5.57)$$

Par identification à l'équation (5.51), nous obtenons :

$$\nabla \varepsilon(\mathbf{u}) = \Phi^* \circ \Phi(\mathbf{u}(\mathbf{x})) + \varphi(\mathbf{x})(\mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x})). \quad (5.58)$$

5.4.3 Opérateurs de projection pour quelques fonctions de régularisation

Nous allons exprimer dans ce paragraphe, l'opérateur $\nabla\varepsilon$, gradient de la fonctionnelle d'énergie, pour quelques fonctions de régularisation communément utilisées ; plus précisément dans le cas de la pénalisation du gradient (5.18), pénalisation de la divergence (5.20), pénalisation du rotationnel(5.21), et la pénalisation de la divergence et du rotationnel au second ordre de Suter (5.31). L'objectif est de définir l'opérateur $\Phi^* \circ \Phi$ pour ces fonctions de régularisation. Les notations utilisées seront celles introduites à la section (5.2.1). Nous noterons $\nabla_{\mathbf{x}}$ les dérivées partielles par rapport aux variables d'espace physique afin d'éviter les confusions.

Régularisation par pénalisation du gradient

Dans le cas de la régularisation par pénalisation du gradient, le terme de régularisation s'exprime :

$$J_{grad}(\mathbf{v}) = \sum_{i=1}^n \int_{\Omega} \|\nabla_{\mathbf{x}} v_i(\mathbf{x})\|^2 d\mathbf{x}. \quad (5.59)$$

Cette régularisation peut être considérée comme une régularisation à plusieurs termes simples, un terme simple étant :

$$J_{grad}(v_i) = \int_{\Omega} \|\nabla_{\mathbf{x}} v_i(\mathbf{x})\|^2 d\mathbf{x}, 1 < i < n. \quad (5.60)$$

La fonction de régularisation associée à un terme simple s'écrit :

$$\Phi^{grad}(v_i) = \nabla_{\mathbf{x}} v_i(\mathbf{x}), 1 < i < n. \quad (5.61)$$

On peut alors considérer une fonction de confiance unique $\varphi(\mathbf{x})$ pour tous les termes ou une fonction de confiance spécifique $\varphi_i(\mathbf{x})$ pour chaque terme. Sans nuire à la généralité, nous considérerons ici une fonction de confiance unique par soucis de clarté de la présentation. Ceci nous permet également de faire un développement comparable au terme de régularisation par pénalisation du gradient tel que définit par Horn et Schunck. La considération d'une fonction de confiance spécifique à chaque terme ne rend pas les calculs complexes ; en effet, par rapport à l'argument de la fonctionnelle d'énergie ε , la fonction de confiance est une fonction constante.

Avec ces précisions, la fonctionnelle d'énergie s'écrit :

$$\varepsilon_{grad}(u_i) = \frac{1}{2} \int_{\Omega} \|\nabla_{\mathbf{x}} u_i(\mathbf{x})\|^2 + \varphi(\mathbf{x}) \|u_i(\mathbf{x}) - v_i(\mathbf{x})\|^2 d\mathbf{x}, 1 < i < n. \quad (5.62)$$

Le terme de régularisation (équation 5.53) se réécrit comme :

$$\tilde{\varepsilon}_{grad}^r(u_i, h_i) = \int_{\Omega} \langle \nabla_{\mathbf{x}} u_i(\mathbf{x}), \nabla_{\mathbf{x}} h_i(\mathbf{x}) \rangle d\mathbf{x}, 1 < i < n. \quad (5.63)$$

La formule de Green nous permet d'obtenir :

$$\int_{\Omega} \langle \nabla_{\mathbf{x}} u_i(\mathbf{x}), \nabla_{\mathbf{x}} h_i(\mathbf{x}) \rangle d\mathbf{x} = - \int_{\Omega} \langle \nabla_{\mathbf{x}}^2 u_i(\mathbf{x}), h_i(\mathbf{x}) \rangle d\mathbf{x} + \oint_{\partial\Omega} \langle \nabla_{\mathbf{x}} u_i(\mathbf{x}), \nu \rangle h_i(\mathbf{x}) ds, 1 < i < n, \quad (5.64)$$

où $\nabla^2 = \Delta$ est l'opérateur laplacien, $\partial\Omega$ est la frontière de Ω , et ν le vecteur unitaire sortant défini sur la normale à l'unité de surface ds de Ω . Si l'on restreint la recherche aux candidats

\mathbf{u} tels que ∇u_i , $1 < i < n$ soit orthogonal à ν sur $\partial\Omega$, ceci exprime les conditions de Neumann homogène sur les frontières. Dans ces conditions, on obtient :

$$\hat{\varepsilon}^r(u_i, h_i) = - \int_{\Omega} \langle \Delta u_i(\mathbf{x}), h_i(\mathbf{x}) \rangle d\mathbf{x}, 1 < i < n. \quad (5.65)$$

On en déduit une expression de l'opérateur :

$$\Phi_{grad*} \circ \Phi_{grad} = -\Delta, \quad (5.66)$$

et du gradient de la fonctionnelle d'énergie :

$$\nabla \varepsilon_{grad}(u_i) = -\Delta u_i(\mathbf{x}) + \varphi(\mathbf{x})(u_i(\mathbf{x}) - v_i(\mathbf{x})), 1 < i < n. \quad (5.67)$$

Régularisation par pénalisation de la divergence

De l'expression du terme de pénalisation de la divergence exprimé par :

$$J_{div}(\mathbf{v}) = \int_{\Omega} |\nabla_{\mathbf{x}} \cdot v(\mathbf{x})|^2 d\mathbf{x}, \quad (5.68)$$

la fonction de régularisation par pénalisation de la divergence peut se définir comme :

$$\Phi_{div}(\mathbf{u}) = div \mathbf{u}(\mathbf{x}) = \nabla \cdot \mathbf{u}(\mathbf{x}). \quad (5.69)$$

Le terme de régularisation de la fonctionnelle d'énergie (équation 5.53) devient :

$$\hat{\varepsilon}_{div}^r(\mathbf{u}, \mathbf{h}) = \int_{\Omega} \langle div \mathbf{u}(\mathbf{x}), div \mathbf{h}(\mathbf{x}) \rangle d\mathbf{x}. \quad (5.70)$$

De l'égalité :

$$\int_{\Omega} \langle \mathbf{u}, \nabla f \rangle d\mathbf{x} = - \int_{\Omega} f \nabla \cdot \mathbf{u} d\mathbf{x} + \oint_{\partial\Omega} \langle \mathbf{u}, \nu \rangle f ds, \quad (5.71)$$

où f est une fonction scalaire, nous obtenons :

$$\hat{\varepsilon}_{div}^r(\mathbf{u}, \mathbf{h}) = - \int_{\Omega} \langle \mathbf{h}(\mathbf{x}), \nabla_{\mathbf{x}} (\nabla_{\mathbf{x}} \cdot \mathbf{u}(\mathbf{x})) \rangle d\mathbf{x} + \oint_{\partial\Omega} \langle \mathbf{h}, \nu \rangle \nabla \cdot \mathbf{u}(\mathbf{x}) ds. \quad (5.72)$$

Si l'espace de recherche est restreint au sous-ensemble des \mathbf{u} tel que $\nabla \cdot \mathbf{u}(\mathbf{x}) = 0$ sur $\partial\Omega$, (ceci exprime la divergence nulle sur les frontières du domaine physique) alors, l'opérateur $\Phi_{div*} \circ \Phi_{div}$ devient :

$$\Phi_{div*} \circ \Phi_{div} = -\nabla_{\mathbf{x}} \nabla_{\mathbf{x}} \cdot = -\nabla_{\mathbf{x}} div, \quad (5.73)$$

et le gradient de la fonctionnelle d'énergie s'écrit :

$$\nabla \varepsilon_{div}(\mathbf{u}) = -\nabla_{\mathbf{x}} (\nabla_{\mathbf{x}} \cdot \mathbf{u}(\mathbf{x})) + \varphi(\mathbf{x})(\mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x})). \quad (5.74)$$

Régularisation par pénalisation du rotationnel

A partir de l'expression du terme de régularisation par pénalisation du rotationnel exprimé par :

$$J_{curl}(\mathbf{v}) = \int_{\Omega} |\nabla_{\mathbf{x}} \times \mathbf{v}(\mathbf{x})|^2 d\mathbf{x}, \quad (5.75)$$

la fonction de régularisation peut se définir comme :

$$\Phi_{curl}(\mathbf{u}) = curl(\mathbf{u}(\mathbf{x})) = \nabla \times \mathbf{u}(\mathbf{x}). \quad (5.76)$$

Le terme de régularisation de la fonctionnelle d'énergie (équation 5.53) devient :

$$\hat{\varepsilon}_{curl}^r(\mathbf{u}, \mathbf{h}) = \int_{\Omega} \langle \nabla \times \mathbf{u}(\mathbf{x}), \nabla \times \mathbf{h}(\mathbf{x}) \rangle d\mathbf{x}. \quad (5.77)$$

De l'égalité :

$$\int_{\Omega} \langle \mathbf{u}, \nabla \times \mathbf{v} \rangle d\mathbf{x} = \int_{\Omega} \langle \nabla \times \mathbf{u}, \mathbf{v} \rangle d\mathbf{x} + \oint_{\partial\Omega} \langle \mathbf{u} \times \nu, \mathbf{v} \rangle f ds, \quad (5.78)$$

nous obtenons :

$$\hat{\varepsilon}_{curl}^r(\mathbf{u}, \mathbf{h}) = \int_{\Omega} \langle \mathbf{h}(\mathbf{x}), \nabla_{\mathbf{x}} \times (\nabla_{\mathbf{x}} \times \mathbf{u}(\mathbf{x})) \rangle d\mathbf{x} + \oint_{\partial\Omega} \langle (\nabla_{\mathbf{x}} \times \mathbf{u}(\mathbf{x})) \times \nu, \mathbf{h}(\mathbf{x}) \rangle ds. \quad (5.79)$$

Si l'espace de recherche est restreint au sous-ensemble des \mathbf{u} tel que $\langle (\nabla_{\mathbf{x}} \times \mathbf{u}) \times \nu \rangle = 0$ sur $\partial\Omega$, (ce qui se traduit par la rotation uniquement suivant un axe normal à la surface de la frontière quand il s'agit d'un point de la frontière) alors l'opérateur $\Phi_{curl*} \circ \Phi_{curl}$ s'écrit :

$$\Phi_{curl*} \circ \Phi_{curl}(\mathbf{u}) = \nabla_{\mathbf{x}} \times (\nabla_{\mathbf{x}} \times \mathbf{u}) \quad (5.80)$$

et le gradient de la fonctionnelle d'énergie s'écrit :

$$\nabla \varepsilon_{curl}(\mathbf{u}) = \nabla_{\mathbf{x}} \times (\nabla_{\mathbf{x}} \times \mathbf{u}(\mathbf{x})) + \varphi(\mathbf{x})(\mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x})). \quad (5.81)$$

De l'égalité :

$$\nabla_{\mathbf{x}} (\nabla_{\mathbf{x}} \times \mathbf{u}) = \nabla_{\mathbf{x}} (\nabla_{\mathbf{x}} \cdot \mathbf{u}) - \Delta \mathbf{u}, \quad (5.82)$$

$\nabla \varepsilon_{curl}$ devient :

$$\nabla \varepsilon_{curl}(\mathbf{u}) = -\Delta \mathbf{u}(\mathbf{x}) + \nabla_{\mathbf{x}} (\nabla_{\mathbf{x}} \cdot \mathbf{u}(\mathbf{x})) + \varphi(\mathbf{x})(\mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x})). \quad (5.83)$$

Remarque 5.4.1 Pour la régularisation div-curl du premier ordre définie par le terme de régularisation :

$$J_{div-curl}(\mathbf{u}) = \int_{\Omega} \alpha \|\nabla_{\mathbf{x}} \cdot \mathbf{u}(\mathbf{x})\|^2 + \beta \|\nabla_{\mathbf{x}} \times \mathbf{u}(\mathbf{x})\|^2 d\mathbf{x}, \alpha, \beta \in \mathbb{R}, \quad (5.84)$$

où α représente le paramètre de pénalisation du terme de divergence et β celui du terme de rotationnel,

le gradient de la fonctionnelle d'énergie s'écrit :

$$\nabla \varepsilon_{div-curl}(\mathbf{u}) = -\alpha \Delta \mathbf{u}(\mathbf{x}) + (\alpha - \beta) \nabla_{\mathbf{x}} (\nabla_{\mathbf{x}} \cdot \mathbf{u}(\mathbf{x})) + \varphi(\mathbf{x})(\mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x})), \quad (5.85)$$

avec les conditions :

- $\langle (\nabla_{\mathbf{x}} \times \mathbf{u}) \times \nu \rangle = 0;$
- $\nabla \cdot \mathbf{u}(\mathbf{x}) = 0,$

sur les frontières $\partial\Omega$ du domaine physique.

Si l'on pose $\alpha = \beta = 1$, on obtient $\nabla \varepsilon_{div-curl} = \nabla \varepsilon_{grad}$ qui montre l'équivalence de la régularisation par pénalisation du gradient (équation 5.18) et la régularisation par équi-pénalisation de la divergence et du rotationnel au premier ordre (équation 5.18). On retrouve l'énoncé de cette équivalence dans les travaux de Gupta et Prince [Gupta and Prince(Feb 1996)]

Autres cas

Régularisation de Suter

La régularisation par pénalisation au second ordre du gradient de la divergence et du gradient du rotationnel (équation 5.31) peut être analysée comme une régularisation à deux termes avec les fonctions de régularisation :

$$\Phi_{sut1} = \nabla_{\mathbf{x}} div; \quad (5.86)$$

$$\Phi_{sut2} = \nabla_{\mathbf{x}} curl. \quad (5.87)$$

Si nous définissons le changement de variable suivant :

$$\mathbf{u}^1 = div(\mathbf{u}); \quad (5.88)$$

$$\mathbf{u}^2 = curl(\mathbf{u}), \quad (5.89)$$

le développement lié à cette régularisation est équivalent à celui de la section (5.4.3) sur la régularisation par pénalisation du gradient.

régularisation par pénalisation isotropique du gradient.

Le cas de la régularisation isotropique basée sur les données d'Alvarez [Alvarez et al.(1999)] défini par la fonction de régularisation :

$$\Phi_2(\mathbf{v}) = g(\|\nabla f\|^2)\Phi_1(\mathbf{v}), \quad (5.90)$$

ainsi que celui de la régularisation isotropique basée sur le champ de Schnörr [Schnörr(1994)] donné par la fonction :

$$\Phi_4(\mathbf{v}) = \Psi\left(\sum_{i=1}^n \|\nabla v_i\|^2\right) \quad (5.91)$$

se ramènent aisément au cas de la régularisation par pénalisation du gradient.

Résumé de l'expression de gradient de la fonctionnelle d'énergie pour les fonctions usuelles de régularisation

Le tableau 5.3 résume l'expression du gradient de la fonctionnelle d'énergie pour les fonctions de régularisation usuelles. Les expressions données dans ce tableau sont valides sous certaines conditions (exprimées à la dérivation de chaque expression).

TABLE 5.3 – Expression du gradient de la fonction d'énergie pour les fonctions de régularisation usuelles

Régularisation	terme de régularisation	gradient de la fonctionnelle d'énergie $\nabla\varepsilon$
gradient	$\sum_{i=1}^n \int_{\Omega} \ \nabla_{\mathbf{x}} v_i(\mathbf{x})\ ^2 d\mathbf{x}$	$-\Delta u_i(\mathbf{x}) + \varphi(\mathbf{x})(u_i(\mathbf{x}) - v_i(\mathbf{x}))$ $1 < i < n$
divergence	$\int_{\Omega} \ \nabla_{\mathbf{x}} \cdot \mathbf{u}(\mathbf{x})\ ^2 d\mathbf{x}$	$-\nabla_{\mathbf{x}} (\nabla_{\mathbf{x}} \cdot \mathbf{u}(\mathbf{x})) + \varphi(\mathbf{x})(\mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x}))$
Rotationnel	$\int_{\Omega} \ \nabla_{\mathbf{x}} \times \mathbf{u}(\mathbf{x})\ ^2 d\mathbf{x}$	$-\Delta \mathbf{u}(\mathbf{x}) + \nabla_{\mathbf{x}} (\nabla_{\mathbf{x}} \cdot \mathbf{u}(\mathbf{x}))$ $+ \varphi(\mathbf{x})(\mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x}))$
div-curl	$\alpha \int_{\Omega} \ \nabla_{\mathbf{x}} \cdot \mathbf{u}(\mathbf{x})\ ^2 d\mathbf{x}$ $+ \beta \int_{\Omega} \ \nabla_{\mathbf{x}} \times \mathbf{u}(\mathbf{x})\ ^2 d\mathbf{x}$ $\alpha, \beta \in \mathbb{R}$	$-\alpha \Delta \mathbf{u}(\mathbf{x}) + (\alpha - \beta) \nabla_{\mathbf{x}} (\nabla_{\mathbf{x}} \cdot \mathbf{u}(\mathbf{x}))$ $+ \varphi(\mathbf{x})(\mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x}))$

5.4.4 Mise en œuvre

Implémentation classique

Une implémentation classique consisterait à utiliser un algorithme de type descente de gradient pour rechercher le minimum de la fonctionnelle d'énergie (équation 5.44) connaissant une expression de son gradient. Etant donnée une approximation initiale \mathbf{u}^0 et un algorithme de calcul de ce gradient $\nabla\varepsilon(\mathbf{u})$. L'algorithme suivant converge vers le minimum \mathbf{u}^* of ε :

$$\mathbf{u}^k = \mathbf{u}^{k-1} + \rho^k \mathbf{d}^k, \quad (5.92)$$

où \mathbf{d}^k est la direction de descente calculée à partir de $\nabla\varepsilon(\mathbf{u}^{k-1})$, et ρ_k le pas de descente choisi de façon appropriée :

$$\rho_k = \text{ArgMin } \varepsilon \left(\mathbf{u}^{k-1} + \rho \mathbf{d}_k \right). \quad (5.93)$$

Le processus est arrêté quand un critère de convergence prédéfini est atteint. Cet algorithme converge vers la solution sous l'hypothèse qu'elle existe et soit unique telle que définie par les conditions suivantes :

- ε est strictement convexe et semi continue inférieur ;
- $\lim_{\|\mathbf{u}\| \rightarrow +\infty} \varepsilon(\mathbf{u}) \rightarrow +\infty$.

Si ces conditions ne sont pas satisfaites, l'algorithme converge vers un minimum local en fonction des propriétés de ε au voisinage de l'approximation initiale et du minimum local en question.

Implémentation basée sur la diffusion généralisée

Dans ce paragraphe, nous expliquons une mise en œuvre élégante de l'approche que nous avons présentée. Cette mise en œuvre permet d'utiliser l'opérateur à la fois pour la régularisation et pour le préconditionnement. La présentation est faite sur l'exemple de la fonction de régularisation associée à la pénalisation du gradient. L'extension aux autres cas pouvant s'en

déduire. L'objectif est la recherche du minimum de la fonctionnelle d'énergie ε dans le cas où la fonction de régularisation est définie par l'équation (5.44).

Implémentation

Au lieu de résoudre le problème équivalent donné par l'équation d'Euler-Lagrange :

$$\nabla^2 u_i - \varphi(\mathbf{x})(u_i(\mathbf{x}) - v_i(\mathbf{x})) = 0, \quad 1 \leq i \leq n, \quad (5.94)$$

par un algorithme classique de type descente du gradient, nous introduisons une pseudo-variable temporelle et nous considérons u_i comme une fonction de cette variable temps. Nous nous ramenons alors à la résolution du système :

$$\frac{\partial}{\partial t} u_i(\mathbf{x}, t) = \nabla_{\mathbf{x}}^2 u_i(\mathbf{x}, t) - \varphi(\mathbf{x})(u_i(\mathbf{x}, t) - v_i(\mathbf{x})), \quad 1 \leq i \leq n. \quad (5.95)$$

Cet ensemble d'équations (5.95) est connu comme le système d'équations de **diffusion généralisée**. Par soucis de simplicité des notations pour une lecture aisée, nous nous limitons dans ce document à la mise en œuvre sur des domaines physiques rectangulaires en deux dimensions. La variable d'espace \mathbf{x} peut alors s'écrire $\mathbf{x} = (x, y)$. Le domaine physique pourra être défini comme :

$$\Omega = \{(x, y) \mid a \leq x \leq b, c \leq y \leq d; a, b, c, d \in \mathbb{R}\}. \quad (5.96)$$

Le système d'équations (5.95) devient :

$$\frac{\partial}{\partial t} u_1(x, y, t) = \nabla_{\mathbf{x}}^2 u_1(x, y, t) - \varphi(x, y)(u_1(x, y, t) - v_1(x, y)); \quad (5.97)$$

$$\frac{\partial}{\partial t} u_2(x, y, t) = \nabla_{\mathbf{x}}^2 u_2(x, y, t) - \varphi(x, y)(u_2(x, y, t) - v_2(x, y)). \quad (5.98)$$

Afin de séparer les fonctions à variable temps des fonctions à variable d'espace uniquement, le système (5.98) peut se réécrire :

$$\frac{\partial}{\partial t} u_1(x, y, t) = \nabla_{\mathbf{x}}^2 u_1(x, y, t) - \varphi(x, y)u_1(x, y, t) + r_1(x, y); \quad (5.99)$$

$$\frac{\partial}{\partial t} u_2(x, y, t) = \nabla_{\mathbf{x}}^2 u_2(x, y, t) - \varphi(x, y)u_2(x, y, t) + r_2(x, y), \quad (5.100)$$

avec $r_1(x, y) = \varphi(x, y)v_1(x, y)$ et $r_2(x, y) = \varphi(x, y)v_2(x, y)$.

Pour la résolution numérique, nous supposons une discrétisation uniforme à pas fixe Δx en x , Δy en y et Δt pour la dimension temps t . Nous noterons :

- $\mathbf{x}_{i,j} = (x_i, y_j)$ les points de calcul du domaine physique discrétisé ; avec $x_i = i \times \Delta x$ et $y_j = j \times \Delta y$;
- $t_n = n \times \Delta t$ le temps discret ;
- $\varphi_{i,j} = \varphi(x_i, y_j)$ (respectivement $v_{1,i,j} = v_1(x_i, y_j)$, $v_{2,i,j} = v_2(x_i, y_j)$, $r_{1,i,j} = r_1(x_i, y_j)$, $r_{2,i,j} = r_2(x_i, y_j)$) la fonction à variables spatiales v_1 (respectivement v_1, v_2, r_1, r_2) au point (x_i, y_j) ;
- $u_{i,j}^n = u_1(x_i, y_j, t_n)$ (respectivement $u_{i,j}^{2n} = u_2(x_i, y_j, t_n)$) la fonction à variable spatiale et temporelle u_1 (respectivement u_2) au point (x_i, y_j) et au temps t_n .

Les différences centrées sont utilisées pour l'approximation des opérateurs aux dérivées partielles sur les variables d'espace et les différences décentrées à droite pour les opérateurs aux dérivées partielles sur la variable temps.

$$\frac{\partial}{\partial t} u_1(x_i, y_j, t_n) = \frac{1}{\Delta t} (u_{i,j}^{n+1} - u_{i,j}^n); \quad (5.101)$$

$$\frac{\partial}{\partial t} u_2(x_i, y_j, t_n) = \frac{1}{\Delta t} (u_{i,j}^{2n+1} - u_{i,j}^{2n}); \quad (5.102)$$

$$\nabla_{\mathbf{x}}^2 u_1(x_i, y_j, t_n) = \frac{1}{\Delta x \Delta y} (u_{i+1,j}^n + u_{i,j+1}^n + u_{i-1,j}^n + u_{i,j-1}^n - 4u_{i,j}^n); \quad (5.103)$$

$$\nabla_{\mathbf{x}}^2 u_2(x_i, y_j, t_n) = \frac{1}{\Delta x \Delta y} (u_{i+1,j}^{2n} + u_{i,j+1}^{2n} + u_{i-1,j}^{2n} + u_{i,j-1}^{2n} - 4u_{i,j}^{2n}). \quad (5.104)$$

En substituant les différents termes dans le système d'équations (5.98), on obtient le système discret suivant :

$$\begin{aligned} \frac{1}{\Delta t} (u_{i,j}^{n+1} - u_{i,j}^n) &= \frac{1}{\Delta x \Delta y} (u_{i+1,j}^n + u_{i,j+1}^n + u_{i-1,j}^n + u_{i,j-1}^n - 4u_{i,j}^n) \\ &\quad - \varphi_{i,j} u_{i,j}^n + r1_{i,j}; \end{aligned} \quad (5.105)$$

$$\begin{aligned} \frac{1}{\Delta t} (u_{i,j}^{2n+1} - u_{i,j}^{2n}) &= \frac{1}{\Delta x \Delta y} (u_{i+1,j}^{2n} + u_{i,j+1}^{2n} + u_{i-1,j}^{2n} + u_{i,j-1}^{2n} - 4u_{i,j}^{2n}) \\ &\quad - \varphi_{i,j} u_{i,j}^{2n} + r2_{i,j}. \end{aligned} \quad (5.106)$$

On en déduit la solution itérative donnée par l'algorithme suivant :

$$\begin{aligned} u_{i,j}^{n+1} &= (1 - \Delta t \varphi_{i,j}) u_{i,j}^n + \frac{\Delta t}{\Delta x \Delta y} (u_{i+1,j}^n + u_{i,j+1}^n + u_{i-1,j}^n + u_{i,j-1}^n - 4u_{i,j}^n) \\ &\quad + \Delta t r1_{i,j}; \end{aligned} \quad (5.107)$$

$$\begin{aligned} u_{i,j}^{2n+1} &= (1 - \Delta t \varphi_{i,j}) u_{i,j}^{2n} + \frac{\Delta t}{\Delta x \Delta y} (u_{i+1,j}^{2n} + u_{i,j+1}^{2n} + u_{i-1,j}^{2n} + u_{i,j-1}^{2n} - 4u_{i,j}^{2n}) \\ &\quad + \Delta t r2_{i,j}. \end{aligned} \quad (5.108)$$

Conditions nécessaires à l'utilisation de l'implémentation par diffusion généralisée

Si les fonctions φ , $r1$ et $r2$ sont bornées, alors la convergence de ce processus est garantie par un résultat classique de la théorie des méthodes numériques [Ames(1992)]. Ces conditions sont assurées par les propriétés du champ à régulariser ainsi que de celles de la fonction de confiance. Pour ce qui est du champ à régulariser, il est en général borné (dans la pratique). La fonction de confiance quant à elle devra être soigneusement construite.

La stabilité du schéma est assurée si la condition de Courant-Friedrichs-Lewy définie par :

$$\frac{\Delta t}{\Delta x \Delta y} \leq \frac{1}{4}, \quad (5.109)$$

est vérifiée. Cette condition est garantie en choisissant un pas de temps Δt satisfaisant :

$$\Delta t \leq \frac{\Delta x \Delta y}{4}. \quad (5.110)$$

5.4.5 De la diffusion généralisée au préconditionnement

L'approche de résolution basée sur la **diffusion généralisée** a deux principaux atouts : outre l'objectif de régularisation visé, nous obtenons un opérateur construit sur le même principe que les opérateurs basiques de covariance d'erreur tels qu'exposé à la section (2.4.4). Son utilisation pour la régularisation se réduit donc à l'utilisation de la matrice de covariance d'erreur d'ébauche pour la construction du terme d'écart à l'ébauche (voir la section 5.1.2) qui sert de terme de régularisation de type Tikhonov. Aussi, cet opérateur peut être utilisé pour le préconditionnement suivant le schéma proposé dans la section (2.4.4). Nous avons ainsi construit un opérateur qui peut servir à la fois pour la régularisation et pour le préconditionnement du solveur utilisé pour résoudre le problème. L'usage comme opérateur de préconditionnement permet de se libérer de la tâche du choix du paramètre approprié associé au terme de régularisation en se restreignant aux propriétés de préconditionnement. Nous étudierons la possibilité de restriction à l'opérateur de préconditionnement sur un cas académique dans la section (5.5).

5.4.6 Extension : vers la restauration du signal

L'opérateur de projection défini par la minimisation de la fonctionnelle d'énergie (équation 5.44) peut être utilisé pour le débruitage ou la restauration d'un signal. En effet la solution du problème est obtenue en lissant les caractéristiques supposées être lisses du signal (terme de régularisation) et en conservant le signal dans les régions où il est supposé être de bonne qualité telle que défini par la fonction de confiance (terme d'attache aux données) : C'est le principe du débruitage et de la restauration. Une fonction de confiance constante peut être utilisée en l'absence de connaissances *a priori* sur la qualité du signal à restaurer. A la section (5.5.4), nous comparons les effets d'une fonction de confiance constante et d'une fonction de confiance intelligente.

5.5 Application à l'estimation du mouvement

5.5.1 De la régularisation par pénalisation du gradient à l'opérateur de projection

Pour l'application de l'approche de régularisation proposée dans ce document, nous nous limiterons au cas où la projection est définie sur la fonction de régularisation du gradient. Nous décrirons la fonction de confiance unique pour tous les termes et nous donnerons des indications pour obtenir une fonction de confiance spécialisée pour chaque terme.

Fonction de confiance pour l'estimation du mouvement

L'annexe (D) décrit l'estimation du mouvement et le principal problème rencontré en estimation du mouvement : le problème d'ouverture. A cause de ce problème, le mouvement ne peut être déterminé que suivant les normales aux iso contours de luminance dans l'image. Il n'y a pas suffisamment d'informations permettant de déterminer correctement le champ dans les autres régions de l'image. A partir de cette description du problème d'ouverture, nous proposons de définir la fonction de confiance φ pour l'estimation du mouvement comme ayant des grandes valeurs sur les points correspondant aux contours dans l'image et de faible valeur ailleurs.

Carte de contours d'une image

Suivant la formulation de Xu [Xu and Prince(1998b)], nous définissons la carte des contours $c(\mathbf{x})$ d'une image donnée par sa fonction de luminance $f(\mathbf{x})$ comme suit :

$$c^1(\mathbf{x}, f) = \|\nabla_{\mathbf{x}} f(\mathbf{x})\|^2; \quad (5.111)$$

$$c^2(\mathbf{x}, f) = \|\nabla_{\mathbf{x}}(\mathbf{G}_{\sigma}(\mathbf{x}) * f(\mathbf{x}))\|^2. \quad (5.112)$$

Où $\mathbf{G}_{\sigma}(\mathbf{x})$ est une gaussienne d'écart type σ et $*$ l'opérateur de convolution. La convolution avec une gaussienne est utilisée pour le lissage en présence d'images bruitées ; elle peut être remplacée par tout autre opérateur de débruitage.

Fonction de confiance

Nous définissons la fonction de confiance comme la carte des contours de l'image au temps 0 : $\varphi(\mathbf{x}) = c^i(\mathbf{x}, f^0)_{i=1,2}$. Lorsque cette fonction de confiance est utilisée pour déterminer $\text{ArgMin}(\varepsilon)$ (ou ε est la fonction d'énergie de la section 5.44), on obtient une solution \mathbf{u}^* proche de \mathbf{v} sur les contours et Φ -régulier dans les autres régions ; c'est l'effet requis face au problème d'ouverture.

Fonctions de confiance spécialisées

Pour plus de précision, l'on peut être amené à définir une fonction de confiance spécialisée pour chaque sous terme de la régularisation par pénalisation du gradient. En se basant sur la description fine du problème d'ouverture, on peut définir une fonction de confiance directionnelle. On définit la carte de contours directionnelle par la valeur absolue de la dérivée directionnelle de la fonction de luminance.

5.5.2 Résultats expérimentaux

A partir de la description faite ci-dessus, nous avons fait un ensemble d'expériences suivant la description de l'annexe (D) en utilisant l'approche de régularisation introduite ici.

Evolution des diagnostics en fonction du paramètre de pondération

Ici, nous étudions le comportement des fonctions de diagnostic à convergence en fonction du paramètre de pondération. L'objectif est d'avoir une idée de la difficulté relative du choix du paramètre optimal de pondération. La figure (5.10) montre l'évolution des fonctions de diagnostic en fonction du paramètre de pondération. Sur le graphique sont représentées les valeurs atteintes par les fonctions de diagnostic quand la méthode a convergé. En abscisse nous avons le paramètre de pondération. Les courbes sont tracées avec une échelle logarithmique. En se basant sur l'évolution de la fonction coût et du terme d'écart aux observations, on peut diviser la courbe des valeurs moyennes en deux parties :

- une première partie correspondant aux faibles valeurs (jusqu'à 10^{-6} environ) du paramètre de pondération où les courbes de la fonction coût et du terme d'écart aux observations sont presque identiques. Dans cette partie du graphique, l'effet de la régularisation est négligeable et la fonction coût est réduite au terme d'écart aux observations d'où la superposition des deux courbes. Les autres fonctions de diagnostic sont quasi constantes et plus faibles que ce que nous avons observées avec les autres fonctions de régularisation. Nous avons notamment une erreur moyenne de moins de 2% sur le champ de vitesse

- estimé, 5% environ en erreur angulaire et moins de 6% en erreur sur la vorticité. Des valeurs qui sont constantes jusqu'à un paramètre de pondération d'environ 10^{-3} ;
- une deuxième partie correspondant aux grandes valeurs du paramètre de pondération où la fonction coût passe sensiblement au dessus du terme d'écart aux observations. La fonction coût est progressivement dominée par le terme de régularisation. Les autres fonctions de diagnostic restent constantes jusqu'à ce que le paramètre de pondération atteigne une valeur autour de 10^{-3} . Elles croissent ensuite et se rapprochent des diagnostics calculés sur l'ébauche.

Contrairement à ce que nous avons observé avec les autres approches de régularisation, toutes les fonctions de diagnostic étudiées ont exactement le même comportement dans ce cas. Une faible valeur de la fonction coût peut être interprétée comme une meilleure approximation de la variable de contrôle. En ce qui concerne les valeurs maximales, l'amélioration est aussi visible sur toutes les fonctions de diagnostic. Contrairement à l'approche classique de régularisation où l'amélioration n'était visible que sur certaines fonctions de diagnostic et pour quelques termes de régularisation, ici l'amélioration est visible sur tous les diagnostics étudiés. Nous avons notamment pour une grande plage de valeur du paramètre de pondération (de 0 à 10^{-1} environ), une erreur maximale normalisée de l'ordre 40% sur le diagnostic le moins bon. Sur la même plage de valeur nous avons une erreur maximale de 4% seulement sur le champ estimé ; c'est une valeur que nous n'avons pas observée avec l'approche classique de régularisation même en termes de valeurs moyennes. Il est important de mentionner que le processus de minimisation est arrêté avant la convergence.

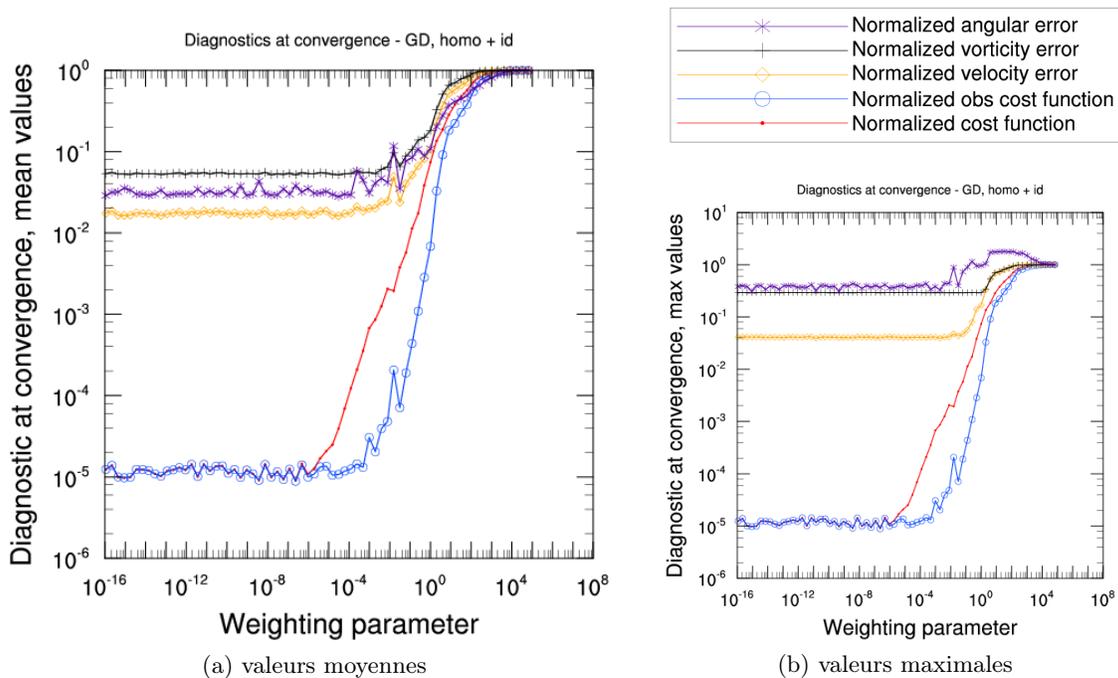


FIGURE 5.10 – Estimation du mouvement avec régularisation et préconditionnement par **diffusion généralisée** : évolution des fonctions de diagnostic à convergence en fonction du paramètre de pondération ; valeurs moyennes (a) et valeurs maximales (b).

Choix du paramètre de pondération

La constance des fonctions de diagnostic sur le champ estimé sur un grand intervalle de valeur du paramètre de pondération montre la robustesse de l'approche et une sensibilité faible à ce paramètre. Une très bonne qualité de la solution même pour de très faibles valeurs du paramètre de pondération nous fait comprendre que l'effet du préconditionnement l'emporte sur l'effet de la pénalisation. Autrement dit, la pénalisation peut être négligée au profit du préconditionnement. Ceci résout le problème de détermination du bon paramètre de pondération qui constitue un problème de première importance avec l'approche classique de régularisation. Nous considérerons la valeur zéro comme la valeur optimale du paramètre de pondération, ce qui prend en compte uniquement l'effet du préconditionnement.

Evolution des fonctions de diagnostic au cours de la minimisation

Ici, nous étudions le comportement des fonctions de diagnostic au cours de la minimisation avec un paramètre de pondération nul (seul l'effet préconditionnement est pris en compte). Cette étude nous permet de nous faire une idée du lien entre la décroissance de la fonction coût et l'amélioration de la solution.

La figure (5.11) montre l'évolution des fonctions de diagnostic au cours des itérations avec l'approche de régularisation que nous avons introduite. Les courbes sont tracées avec une échelle logarithmique en ordonnées. Comme on pouvait s'y attendre, on ne retrouve pas d'oscillations dans l'évolution de la fonction coût. Toutes les fonctions de diagnostic ont le même comportement, elles décroissent avec les itérations, ce qui n'était pas le cas avec l'approche classique. On peut conclure sans trop de risque de se tromper que la décroissance de la fonction coût est synonyme d'amélioration de l'approximation. La décroissance est très forte au début du processus, puis continue avec un rythme moins soutenu. En faisant un nombre assez élevé d'itérations on obtient une erreur de moins de 1% sur le champ estimé avec une fonction coût qui évolue de 5 ordres de grandeurs (Le rapport entre la fonction coût initiale et la fonction coût finale est de 10^{-5}).

5.5.3 Etude comparative avec l'approche classique

Dans ce paragraphe, nous faisons une étude comparative avec l'approche classique de régularisation. Les courbes d'évolution des fonctions de diagnostic avec les itérations sont comparées sur le même graphique. Les figures (5.12, 5.13, 5.14, 5.15) montrent l'évolution comparée respectives du terme d'écart aux observations, de l'erreur sur le champ estimé, de l'erreur angulaire et de l'erreur sur la vorticité avec les régularisations suivantes :

- aucune régularisation ;
- homogène de Horn et Schunck par pénalisation du gradient ;
- second ordre de Suter (gradient de divergence et du rotationnel) ;
- isotropique basée sur les données ;
- isotropique basée sur le champ ;
- préconditionnement par **diffusion généralisée** (approche introduite ici).

Le terme de régularisation étant pondéré dans chaque cas par le paramètre optimal. Dans le cas de la régularisation de Suter, les paramètres utilisés correspondent aux valeurs optimales obtenues avec l'étude sur le terme de pénalisation du gradient du rotationnel et le terme de pénalisation du gradient de divergence.

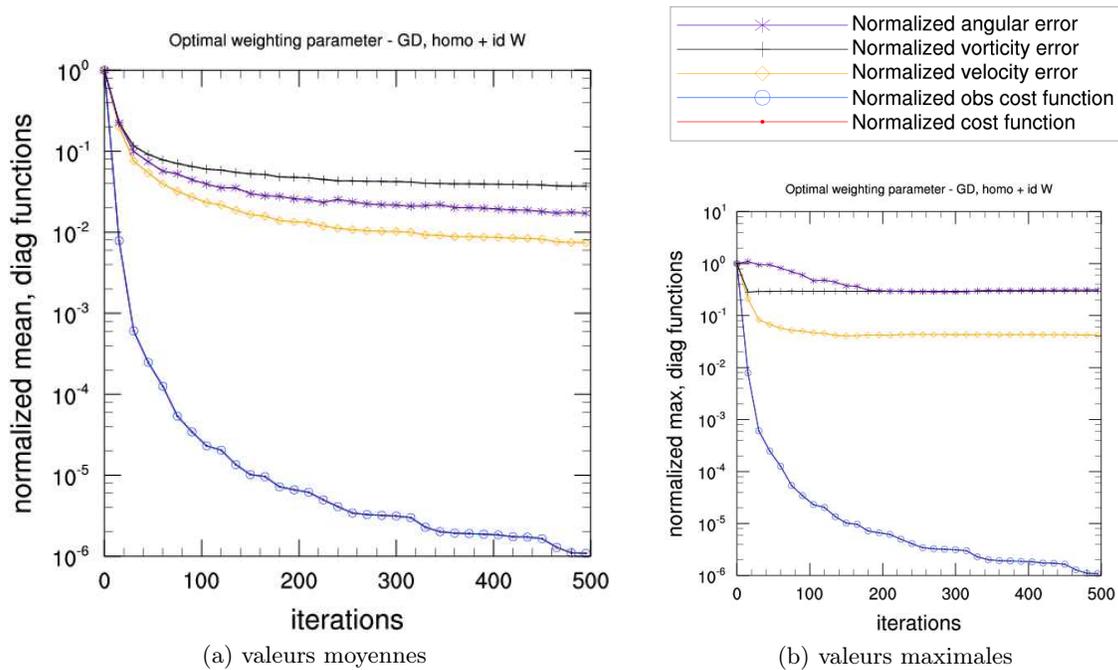


FIGURE 5.11 – Estimation du mouvement avec régularisation et préconditionnement par **diffusion généralisée** : évolution des fonctions de diagnostic normalisées en fonction des itérations ; valeurs moyennes (a) et valeurs maximales (b).

Terme d'écart aux observations de la fonction coût

Sur la figure (5.12), on peut constater que l'évolution du terme d'écart aux observations est comparable dans tous les cas jusqu'au point où les méthodes basées sur l'approche classique stagnent. Ensuite seule la courbe basée sur la **diffusion généralisée** continue de décroître. On peut également noter que seules les courbes basées sur la **diffusion généralisée** et l'expérience sans régularisation n'oscillent pas.

Erreur globale sur le champ estimé

La figure (5.13) montre les courbes comparatives de l'évolution des fonctions de diagnostic. Sur le graphique de gauche (moyennes normalisées), on note une décroissance très rapide de l'approche par **diffusion généralisée** par rapport à l'approche classique. De même que dans le cas du terme d'écart aux observations de la fonction coût, on constate que seule la courbe de l'approche par **diffusion généralisée** continue à décroître au bout d'un certain nombre d'itérations. On note un comportement identique des courbes avec les autres fonctions de diagnostic. L'erreur moyenne normalisée obtenue à l'arrêt du processus de minimisation est de :

- 60% sans régularisation ;
- 30% avec la régularisation isotropique basée sur les données et la régularisation isotropique basée sur le champ ;
- 10% avec la régularisation par pénalisation du gradient et la régularisation au second ordre de Suter ;
- moins de 1% avec l'approche par **diffusion généralisée**.

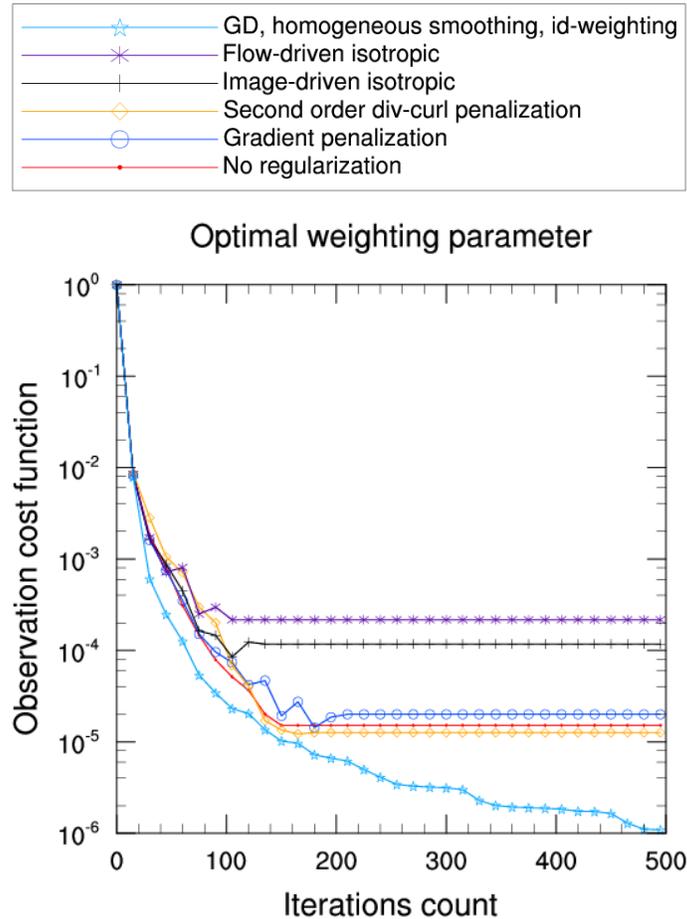


FIGURE 5.12 – Estimation du mouvement avec l’approche classique et l’approche par **diffusion généralisée**, comparaison de l’évolution du terme d’écart aux observations au cours de la minimisation.

En terme de valeurs maximales, graphique de droite de la figure (5.13), on note à l’arrêt du processus de minimisation une erreur sur le champ estimé de l’ordre de :

- 120% sans régularisation ;
- 90% avec la régularisation isotropique basée sur les données et la régularisation isotropique basée sur le champ ;
- 30% avec la régularisation par pénalisation du gradient et la régularisation au second ordre de Suter ;
- 4% avec l’approche par **diffusion généralisée**.

Toutes ces valeurs sont récapitulées dans la table (5.4).

Erreur angulaire sur le champ estimé

En terme d’erreur angulaire (figure 5.14), les remarques sont similaires à celles de l’erreur sur le champ estimé. L’erreur angulaire moyenne normalisée obtenue à l’arrêt du processus de minimisation est de :

- 40% sans régularisation ;

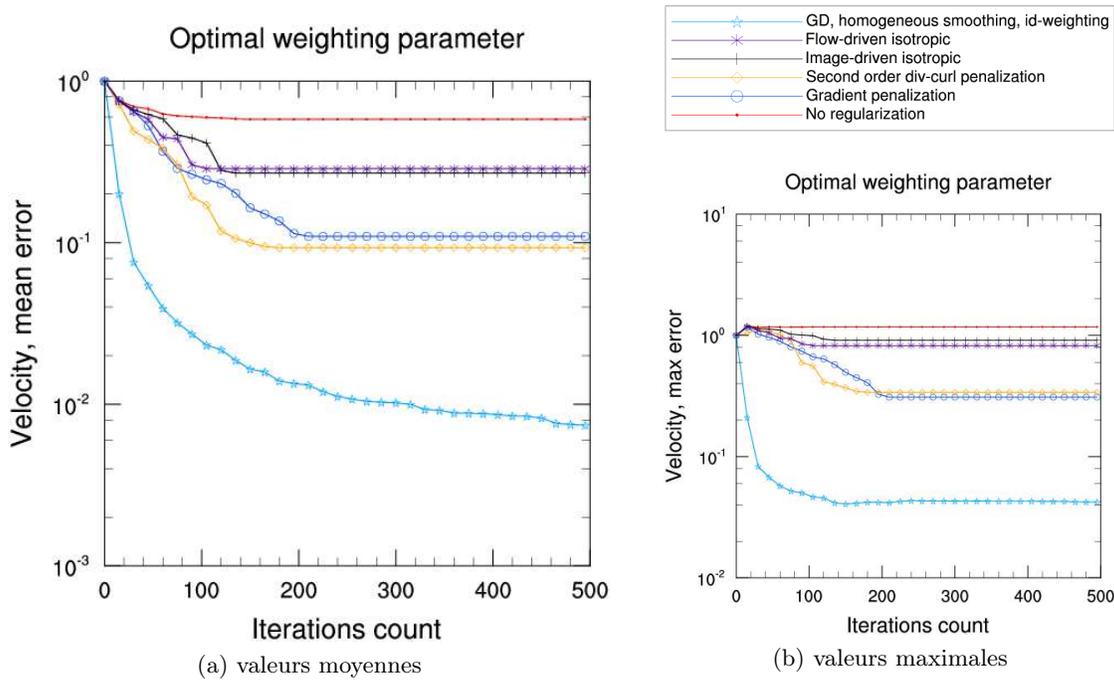


FIGURE 5.13 – Estimation du mouvement avec l’approche classique et l’approche par **diffusion généralisée**, comparaison de l’évolution de l’erreur moyenne normalisée sur le champ estimé au cours de la minimisation ; valeurs moyennes (a) et valeurs maximales (b).

- 20% avec la régularisation isotropique basée sur le champ ;
- 10% avec la régularisation par pénalisation du gradient, la régularisation au second ordre de Suter et la régularisation isotropique basée sur les données ;
- 1% avec l’approche par **diffusion généralisée**.

En termes d’erreur maximale cependant, il n’y a pas de comparaison possible avec l’erreur sur le champ estimé. Seule l’approche par **diffusion généralisée** permet d’obtenir une erreur maximale raisonnable de 25% ; on note une erreur maximale de l’ordre de 90% avec la régularisation par pénalisation du gradient et 150 à 200% dans les autres cas. Toutes ces valeurs sont résumées dans la table (5.4).

Erreur sur la vorticit  du champ estim 

En terme de vorticit  (figure 5.15), seule l’approche par **diffusion g n ralis e** permet d’am liorer de fa on significative l’erreur moyenne. Dans les autres cas l’am lioration reste mitig e. Au d but de la minimisation, l’erreur moyenne sur la vorticit  croit dans tous les cas avant de d croitre dans certains cas. A l’arr t du processus de minimisation, on note une erreur moyenne (graphique de gauche de la figure 5.15) de l’ordre de :

- 400% sans r gularisation ;
- 200% avec la r gularisation isotropique bas e sur le champ et la r gularisation isotropique bas e sur les donn es ;
- 80% avec la r gularisation par p nalisation du gradient et la r gularisation au second ordre de Suter ;
- 4% avec l’approche par **diffusion g n ralis e**.

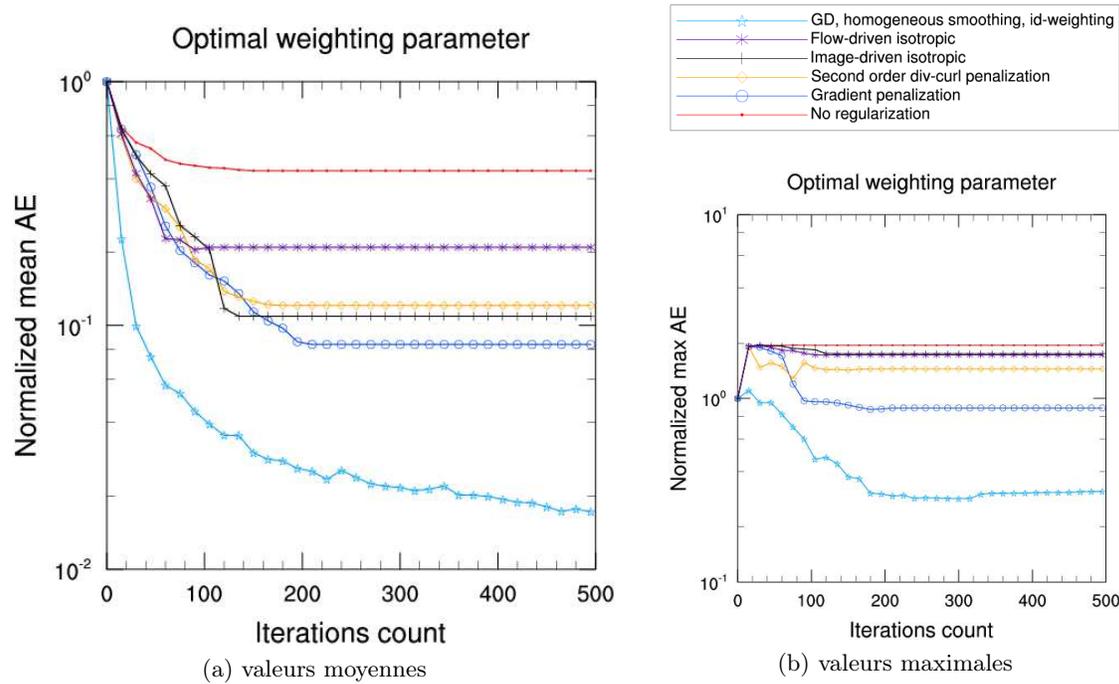


FIGURE 5.14 – Estimation du mouvement avec l'approche classique et l'approche par **diffusion généralisée**, comparaison de l'évolution de l'erreur angulaire moyenne normalisée au cours de la minimisation ; valeurs moyennes (a) et valeurs maximales (b).

En valeurs maximales (graphique de droite de la figure 5.15), seule l'approche par **diffusion généralisée** permet d'améliorer l'estimation, on note à l'arrêt du processus de minimisation une erreur maximale de l'ordre de 15%. Dans les autres cas, on observe une détérioration. Toutes les valeurs sont résumées dans la table (5.4).

Récapitulatif de la comparaison des fonctions de diagnostic

Nous présentons ici un récapitulatif de l'étude comparative sur les fonctions de diagnostic. Le tableau (5.4) récapitule les valeurs des différentes fonctions de diagnostic à convergence avec l'approche par **diffusion généralisée** et avec différents termes de l'approche classique de régularisation. Toutes les valeurs sont normalisées par rapport à leurs équivalents calculés sur l'ébauche sauf dans le cas de l'erreur angulaire où nous donnons les valeurs normalisées (E A normalisée) et les valeurs réelles (E A). La colonne moyenne donne les valeurs moyennes sur le domaine physique et la colonne maxima les valeurs maximales sur le domaine. Toutes les valeurs sont données en pourcentage sauf les valeurs réelles de l'erreur angulaire qui sont données en degrés et le terme d'écart aux observations dont les valeurs relatives sont données directement.

Etude comparative du champ estimé

Ici, nous présentons sur des graphiques le champ estimé avec différentes approches :

- sans régularisation ;

TABLE 5.4 – Estimation du mouvement avec différents approches/termes de régularisation, comparaison des fonctions de diagnostic à l'arrêt du processus de la minimisation

Écart aux observations		
Terme de régularisation	valeur moyenne	valeur maximale
sans régularisation	1.51e-05	-
Horn & Schunck	1.99e-05	-
Suter	1.26e-05	-
Isotropique basée sur les données	1.17e-04	-
Isotropique basée sur le champ	2.16e-04	-
Diffusion généralisée	1.08e-06	-

Diagnostics d'erreur relatives		
Terme de régularisation	valeur moyenne(%)	valeur maximale(%)
Erreur sur le champ de vitesse		
sans régularisation	58	117
Horn & Schunck	10	30
Suter	9	33
Isotropique basée sur les données	26	91
Isotropique basée sur le champ	28	82
Diffusion généralisée	0.7	4
Erreur sur la vortacité		
sans régularisation	411	414
Horn & Schunck	83	169
Suter	68	95
Isotropique basée sur les données	188	321
Isotropique basée sur le champ	184	306
Diffusion généralisée	3.7	29
Erreur angulaire normalisée		
sans régularisation	43	194
Horn & Schunck	8	88
Suter	12	144
Isotropique basée sur les données	10	174
Isotropique basée sur le champ	20	172
Diffusion généralisée	1.7	30
Erreur angulaire en degré		
Terme de régularisation	valeur moyenne(°)	valeur maximale(°)
sans régularisation	25	173
Horn & Schunck	5	79
Suter	7	129
Isotropique basée sur les données	6	156
Isotropique basée sur le champ	12	153
Diffusion généralisée	1	27

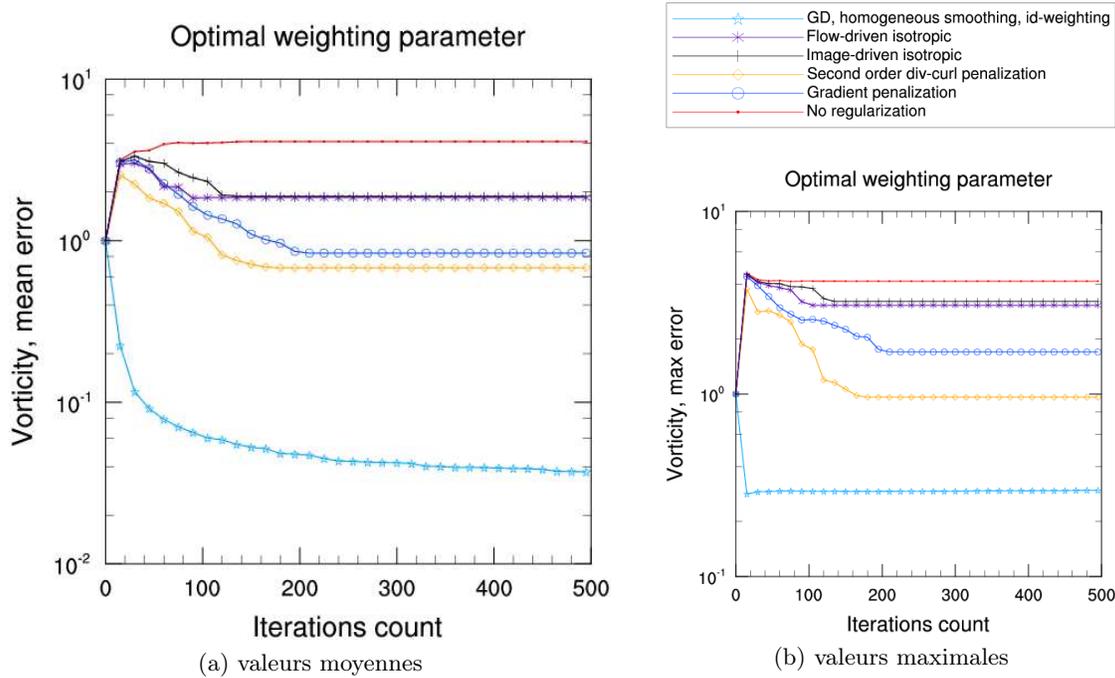


FIGURE 5.15 – Estimation du mouvement avec l’approche classique et l’approche par **diffusion généralisée**, comparaison de l’évolution de l’erreur moyenne normalisée sur la vorticité au cours de la minimisation ; valeurs moyennes (a) et valeurs maximales (b).

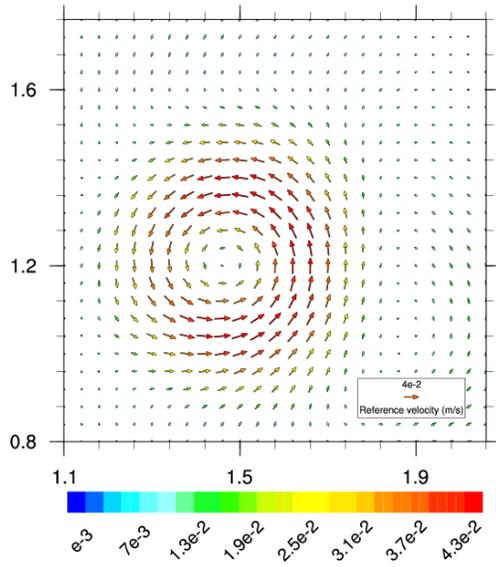
- avec régularisation par pénalisation du gradient ;
- avec l’approche par **diffusion généralisée**.

La figure (5.16) montre le champ de vitesses vrai et le champ estimé dans chaque cas. En l’absence de régularisation, on peut noter l’aspect chaotique du champ estimé comparé au champ réel qui représente à l’échelle de l’expérience un cyclone (le grand tourbillon) et un anticyclone à droite du cyclone. Avec la régularisation par pénalisation du gradient, la différence est bien notable par rapport à l’expérience sans régularisation. Le cyclone est visiblement bien reconstruit à la bonne position et l’anticyclone semble aussi reconstruit (au moins partiellement). La différence se situe au niveau de l’ordre des grandeurs : dans le champ vrai, nous avons des vitesses allant jusqu’à 4.3cm/s , ici on ne parvient qu’à 3.6cm/s . Ce qui explique assez bien le fait qu’on ait une erreur angulaire plus faible que l’erreur sur la vorticité. Dans le cas de la **diffusion généralisée**, la différence avec le champ vrai semble impossible à vue d’œil. Ce qui est bien en concordance avec les graphiques précédents.

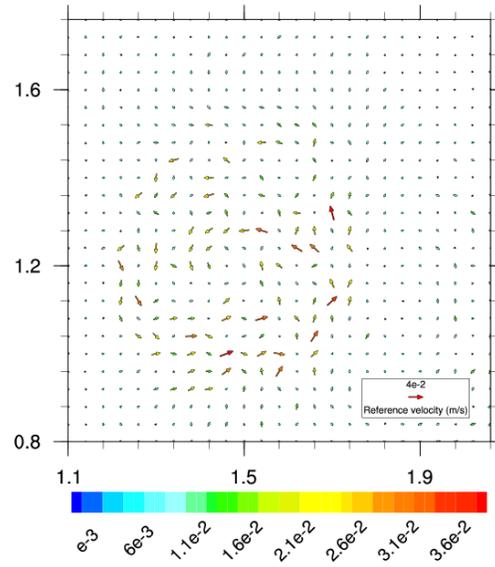
5.5.4 Sensibilité à la fonction de confiance

Afin d’étudier la sensibilité de la nouvelle approche de régularisation par rapport à la fonction de confiance, nous avons fait un ensemble d’expériences avec deux fonctions de confiance différentes :

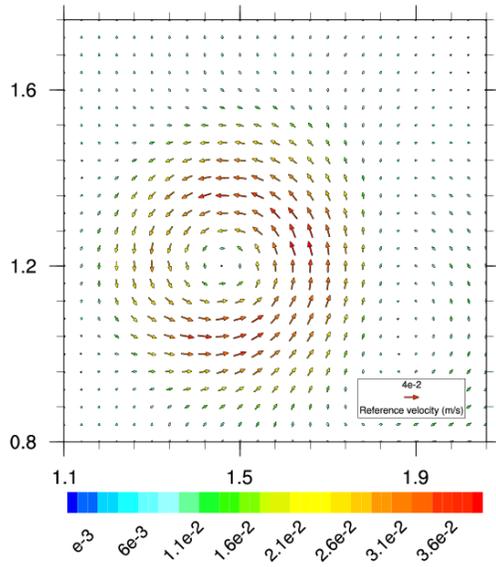
- $\varphi(\mathbf{x}) = 1$, une fonction de confiance constante ;
- $\varphi(\mathbf{x}) = c^1(\mathbf{x})$, c’est la fonction de confiance basée sur la carte de contours, celle que nous avons utilisée jusqu’à présent.



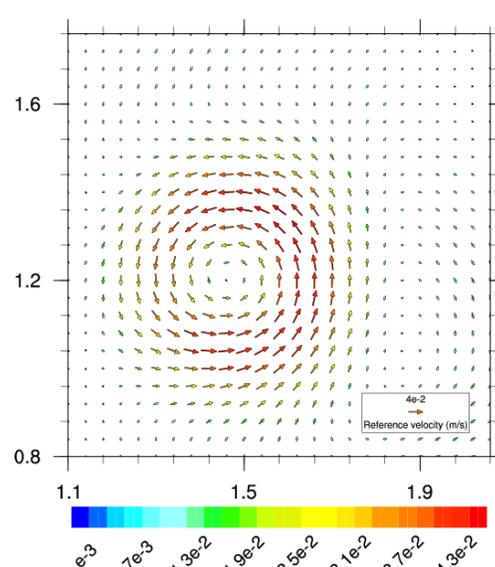
(a) Etat de contrôle vraie



(b) Etat de contrôle analysé : sans régularisation



(c) Etat de contrôle analysé : pénalisation du gradient



(d) Etat de contrôle analysé : diffusion généralisée

FIGURE 5.16 – Champs de vitesse reconstruit par estimation du mouvement avec différentes approches de régularisation

Fonction de confiance constante

Avec une fonction de confiance constante, on donne le même poids à tous les points du domaine. La valeur de la constante détermine le poids relatif entre le terme de lissage et le

terme d'attache aux données. Une telle fonction de confiance définit une pondération homogène du terme d'attache aux données. La **diffusion généralisée** appliquée à la résolution de la fonctionnelle d'énergie (équation 5.44) se comporte de la même façon qu'un opérateur de lissage telle qu'une gaussienne. Son principal atout étant de lisser une caractéristique précise de la donnée traitée telle que le gradient, la divergence ou encore le rotationnel pour ne citer que ceux-ci. Cette fonction de confiance permet notamment d'utiliser la fonction de régularisation pour préconditionner le problème.

Fonction de confiance basée sur la carte des contours

La fonction de confiance basée sur la carte des contours est celle que nous avons présentée à la section (5.5.1). Elle prend en considération les connaissances du problème d'ouverture rencontré en traitement d'images ; c'est une candidate idéale pour les problèmes d'images. On définit ainsi une fonction de confiance basée sur l'image. Une telle fonction de confiance définit une pondération basée sur l'image du terme d'attache aux données.

Sensibilité à la fonction de confiance

A partir des données décrites en annexe (D), nous avons réalisé des expériences d'estimation du mouvement avec l'approche de régularisation introduite dans ce document. Les deux fonctions de confiance (constante et contours) décrites ci-dessus sont utilisées. Pour la fonction de confiance constante, nous avons pris la valeur 1. Nous présentons dans ce paragraphe les courbes comparant les résultats obtenus avec ces deux fonctions de confiance.

Remarque 5.5.1 *De même que nous avons défini une fonction de confiance constante et une fonction de confiance basée sur l'image, on peut définir une fonction de confiance basée sur le champ ; l'utilisation d'une telle fonction de confiance peut se trouver dans le rapport de recherche [Souopgui et al.(2009)] qui a constitué le point de départ de l'approche de régularisation introduite ici.*

Terme d'écart aux observations de la fonction coût

La figure (5.17) représente l'évolution du terme d'écart aux observations de la fonction coût au cours de la minimisation pour une fonction de confiance constante et une fonction de confiance basée sur la carte des contours de l'image. La distinction est très faible.

Erreur globale sur le champ estimé

La figure (5.18) montre les courbes comparatives de l'évolution de l'erreur sur le champ estimé au cours de la minimisation. De même que dans le cas du terme d'écart aux observations, les deux courbes ne permettent pas de dissocier les deux fonctions de confiance, que ce soient en termes de valeurs moyennes (figure 5.18a) ou de valeurs maximales (figure 5.18b).

Erreur angulaire sur le champ estimé

La figure (5.19) montre les courbes comparatives de l'évolution de l'erreur angulaire normalisée sur le champ estimé. Ici, on note des différences significatives dans certaines régions du graphique. L'erreur moyenne (figure 5.19a) avec la fonction de confiance constante est supérieure de 2% à l'erreur angulaire avec la fonction de confiance basée sur la carte des contours de l'image. La différence est aussi visible en termes de valeurs maximales (figure 5.19b), on

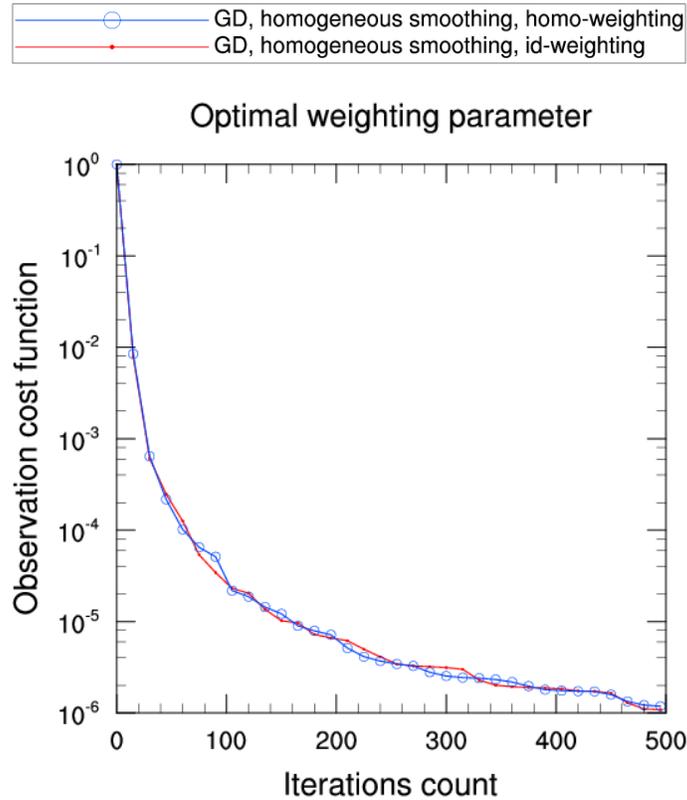


FIGURE 5.17 – Estimation du mouvement : Comparaison de deux fonctions de confiance sur l'évolution du terme d'écart aux observations au cours de la minimisation.

note une différence de presque 10% en faveur de la fonction de confiance basée sur la carte des contours.

Erreur sur la vorticité du champ estimé

La figure (5.18) montre les courbes comparatives de l'évolution de l'erreur sur la vorticité du champ estimé au cours de la minimisation. De même que dans le cas du terme d'écart aux observations, les deux courbes ne permettent pas de dissocier les deux fonctions de confiance, que ce soient en termes de valeurs moyennes (figure 5.20a) ou de valeurs maximales (figure 5.20b).

A partir de cette étude expérimentale, on peut déduire que l'approche de régularisation présentée ici est peu sensible à la fonction de confiance. Ceci confirme d'une certaine façon l'importance de l'aspect préconditionnement dans cette approche d'utilisation de l'information de la régularisation. La différence entre les résultats avec ces deux fonctions de confiance pourrait sans doute être plus visible avec d'autres jeux de données. Les résultats que nous venons de présenter montre que l'usage d'une fonction de confiance (même constante) avec la régularisation et préconditionnement par **diffusion généralisée** produit de meilleurs résultats que l'approche classique.

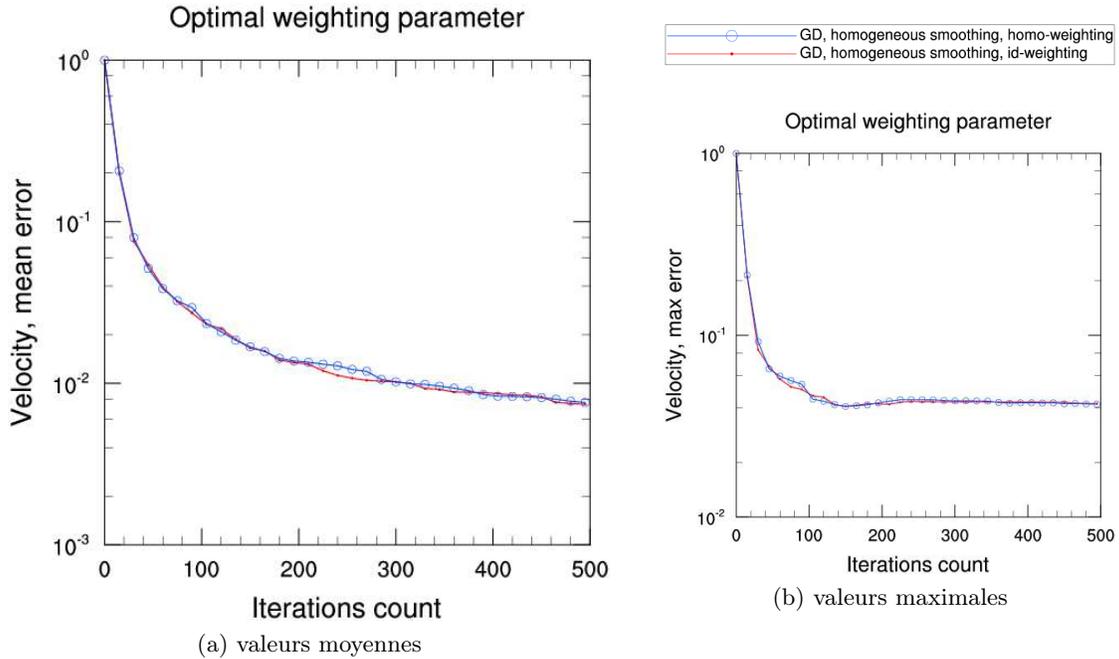


FIGURE 5.18 – Estimation du mouvement : Comparaison de deux fonctions de confiance sur l'évolution de l'erreur moyenne normalisée sur le champ estimé au cours de la minimisation ; valeurs moyennes (a) et valeurs maximales (b).

5.5.5 conclusion partielle

De l'étude faite à la section (5.2) sur l'approche classique de régularisation par pénalisation, nous avons retenu les conclusions suivantes :

- si la régularisation permet d'améliorer la solution, le choix du paramètre de pondération du terme de régularisation reste très difficile et même impossible à partir des seules connaissances *a priori* ;
- l'amélioration apportée par la régularisation peut être limitée à certains diagnostics seulement de la solution calculée, les autres diagnostics montrant plutôt une détérioration ;
- si l'impact de la régularisation peut être notable et positif en termes de valeurs moyennes sur l'erreur commise, il reste négligeable et même négatif en termes de valeur maximales ;
- le nombre d'itérations doit être minutieusement choisi afin d'espérer l'amélioration sur un maximum de fonctions de diagnostic ;
- l'usage de la régularisation restreint, et quelquefois de façon artificielle, l'espace de recherche des solutions. Restriction non prise en compte par les solveurs itératifs en général.

Les problèmes posés par ces conclusions sont quelques uns des problèmes auxquels une nouvelle approche de régularisation devrait s'attaquer. Dans la section (5.4), nous avons proposé une nouvelle approche de régularisation. Nous avons montré ses détails et nous avons proposé une approche de mise en œuvre basée sur la **diffusion généralisée**. Cette approche par **diffusion généralisée** permet d'utiliser l'opérateur de régularisation pour définir un préconditionnement pour la résolution du problème. Dans cette section, nous avons montré les détails de la mise en œuvre de cette nouvelle approche sur le problème d'estimation du mouvement. Nous avons

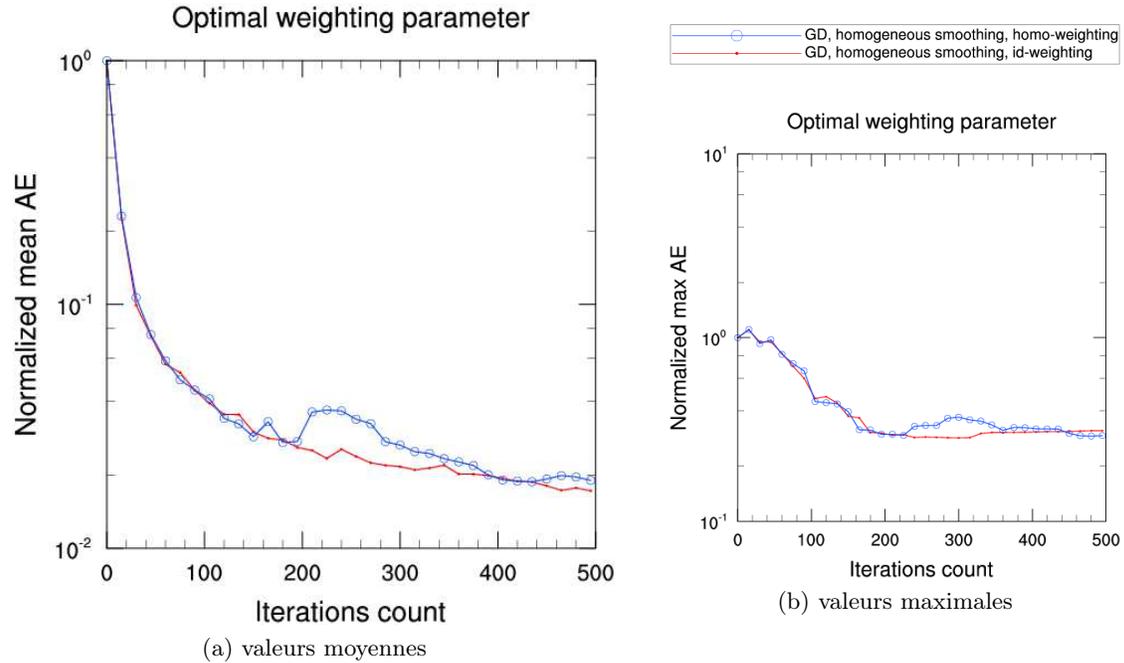


FIGURE 5.19 – Estimation du mouvement : Comparaison de deux fonctions de confiance sur l'évolution de l'erreur angulaire moyenne normalisée au cours de la minimisation ; valeurs moyennes (a) et valeurs maximales (b).

présenté et comparé les résultats avec l'approche classique. Il en ressort notamment que :

- le problème de choix du paramètre de pondération du terme de pénalisation dont nous avons montré la difficulté dans le cas classique peut être esquivé avec la nouvelle approche. Il suffit de n'utiliser que le préconditionnement dérivé de la nouvelle approche ;
- la limitation de l'amélioration à certaines fonctions de diagnostics n'est pas constatée ici. Tous les diagnostics étudiés montrent une amélioration significative de la solution d'un problème inverse par la nouvelle approche ;
- avec l'approche classique, les améliorations ne s'observaient que sur les valeurs moyennes, en termes de valeurs maximales, on n'observait que des détériorations. Avec cette nouvelle approche, l'amélioration est assez spectaculaire même sur les valeurs maximales des fonctions de diagnostic. On relève notamment ici des valeurs maximales de meilleure qualité que les valeurs moyennes dans la plupart des cas avec l'approche classique ;
- le choix du point d'arrêt de la minimisation ne se pose pas non plus ; en effet, on observe une amélioration dès la première itération et sur toutes les fonctions de diagnostic considérées ;
- avec cette approche, le problème de la restriction de l'espace de recherche observé avec l'approche classique est résolu par l'utilisation d'un opérateur de projection.

Il en découle donc une approche élégante et prometteuse pour l'étude des problèmes inverses.

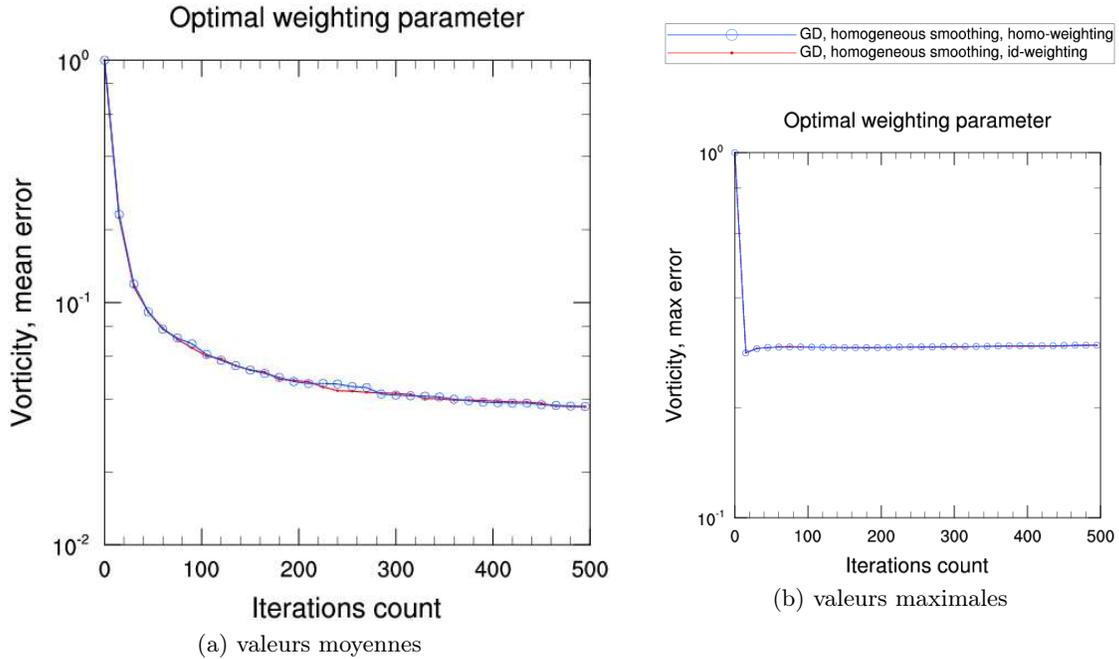


FIGURE 5.20 – Estimation du mouvement : Comparaison de deux fonctions de confiance sur l'évolution de l'erreur moyenne normalisée sur la vorticité au cours de la minimisation ; valeurs moyennes (a) et valeurs maximales (b).

5.6 Conclusion

Nous avons étudié l'utilisation de la régularisation par pénalisation pour la résolution des problèmes inverses par des méthodes itératives et quelques problèmes qui en découlent. Nous avons proposé une nouvelle approche pour la régularisation des problèmes inverses mal posés. Nous avons présenté ses détails et proposé une implémentation basée sur la **diffusion généralisée** qui permet de transformer l'opérateur de régularisation en opérateur de préconditionnement. Nous avons présenté les résultats expérimentaux et une étude comparative avec quelques techniques de régularisation par pénalisation. Ces premiers résultats montrent une méthode performante et robuste qui résout la plupart des problèmes rencontrés avec l'approche par pénalisation. Pour y arriver, la nouvelle approche utilise notamment un opérateur de projection qui permet de se passer de certaines restrictions rencontrées dans les autres cas. Ajouté au préconditionnement qui découle de l'implémentation par **diffusion généralisée**, on obtient une méthode pour laquelle le problème du choix du paramètre de pondération (problème très difficile) est résolu. Les premiers résultats montrent une approche très prometteuse pour l'étude et la résolution des problèmes inverses mal posés.

Chapitre 6

Conclusion générale et perspectives

6.1 Conclusion

La prévision du comportement d'un système physique nécessite l'utilisation de toutes les sources d'information disponibles sur ce système. Ces sources d'information, hétérogènes en nature, en qualité et en quantité incluent les modèles mathématiques, les mesures physiques ou observations, les statistiques d'erreurs, etc. Les méthodes d'assimilation de données permettent d'utiliser toutes ces informations de façon optimale pour réaliser la meilleure prévision. Ces méthodes ont déjà montré leur utilité pour des problèmes aussi complexes que la prévision¹ météorologique et océanographique. Elles sont de plus en plus utilisées dans des domaines aussi variés que la médecine, l'agriculture et la glaciologie sans être exhaustif. Chaque nouvelle source d'observation nécessite la définition d'un opérateur d'observation, d'un opérateur de distance ainsi que des opérateurs de régularisation appropriés. Les observations de type image représentent une importante source d'observation. Leur utilité justifie l'importance des moyens mobilisés pour leur acquisition. La construction, la mise en orbite et l'entretien de satellites d'observations de la terre illustrent l'importance des moyens qui peuvent être déployés pour l'acquisition des images. Les séquences d'images satellite permettent entre autres de localiser et de suivre certains phénomènes tels que les ouragans, les tempêtes, la pollution atmosphérique et océanique. L'information contenue dans les images fournit ainsi une idée de la dynamique des phénomènes observés. De plus, les images constituent la seule source d'observation dans certains domaines. C'est le cas de l'astrophysique où les observations conventionnelles sont presque inexistantes. Un autre exemple, c'est la médecine où on ne peut pas se permettre certaines manipulations. Cependant, les images sont jusqu'à présent utilisées essentiellement de façon qualitative à cause de leur complexité : elles permettent aux experts de corriger manuellement les sorties des modèles afin de prendre en compte les phénomènes observés à travers les images. Une approche d'utilisation quantitative des mesures de type image consiste à estimer d'autres variables d'état du système. Ces estimations sont ensuite utilisées comme observations desdites variables : on parle alors de pseudo-observations. Beaucoup d'efforts ont été faits dans ce sens, mais cette approche souffre du manque de consistance physique des pseudo-observations ainsi extraites. De plus, les caractéristiques des pseudo-observations violent en général les hypothèses de base de l'assimilation de données comme nous l'avons expliqué au chapitre (3). Le but de cette thèse était d'étudier les méthodes d'assimilation

1. L'importance de la prévision météorologique et océanographique n'est plus à démontrer aussi bien pour les applications civiles (transport maritime et aérien, sécurité publique, etc.) que militaires.

variationnelle de données et leur extension aux observations de type image.

Nous avons présenté au chapitre (2) les fondements théoriques et pratiques de l'assimilation variationnelle de données. Nous avons ensuite présenté dans le chapitre (3) les sources d'observation dans le cas des systèmes atmosphériques, avec une attention particulière sur les observations de type image. Nous avons présenté les propriétés des images, leur utilisation actuelle ainsi que les ingrédients nécessaires à leur utilisation directe en assimilation variationnelle de données.

Dans le chapitre (4), nous avons proposé un formalisme général d'assimilation des observations de type séquences d'images dans les modèles numériques. Nous avons défini un cadre approprié afin de tester la faisabilité du formalisme proposé. Nous avons montré que l'utilisation des images sous leur forme brute (matrice de pixels) n'est pas appropriée dans un contexte dynamique. Pour faire face à ce problème, nous avons introduit les notions de niveau d'interprétation des images, d'opérateurs image et d'espaces d'images. A partir de ces notions, nous avons proposé un formalisme de construction d'opérateurs d'observation et de distances pour les observations de type image. Nous avons expérimenté le formalisme proposé sur la représentation spectrale des images. Pour cela, nous avons choisi la décomposition multi-échelles et multi-orientations dans la frame des curvelets. L'utilisation d'un seuillage approprié restreint l'image à l'information pertinente qui y est contenue et permet d'obtenir des résultats de qualité supérieure à celle obtenue en utilisant la représentation en pixels. Nous avons défini la notion de seuillage échelle par échelle sur la frame des curvelets. Ce seuillage s'est montré plus adapté pour l'interprétation des images dans un contexte d'assimilation de données.

Le problème de régularisation a été étudié au chapitre (5). Nous avons fait un état de l'art des techniques de régularisation par pénalisation. Ce sont des techniques habituellement utilisées pour la régularisation des problèmes inverses en général et des problèmes de traitement d'images en particulier. Nous avons ensuite proposé une nouvelle approche d'utilisation de l'information de régularisation dans la résolution des problèmes inverses. Les détails de cette approche ont été présentés ainsi qu'une implémentation basée sur la diffusion généralisée. Cette implémentation permet de transformer les opérateurs classiques de régularisation en opérateurs de préconditionnement. Avec cette utilisation, le problème est reformulé en intégrant toute l'information de la régularisation. La convergence des solveurs est ainsi accélérée tout en améliorant la qualité du résultat obtenu. Nous avons présenté les résultats expérimentaux ainsi qu'une étude comparative avec quelques autres techniques de régularisation. Ces résultats montrent une méthode performante et robuste qui résout la plupart des problèmes rencontrés avec l'approche de régularisation par pénalisation. Pour y arriver, la nouvelle approche utilise notamment un opérateur de projection qui permet d'éviter certaines restrictions rencontrées avec la régularisation par pénalisation. Les expériences montrent qu'il est possible d'utiliser uniquement l'aspect préconditionnement découlant de l'implémentation par diffusion généralisée sans dégrader ni la qualité de la solution ni la convergence du solveur. Cette propriété permet ainsi de résoudre le problème du choix du paramètre de pondération du terme de régularisation qui est un problème très difficile, quoique n'étant pas très abordé dans la littérature.

6.2 Perspectives

6.2.1 Opérateurs images

Décomposition en ondelettes

Dans les travaux menés au cours de cette thèse, nous nous sommes limités dans un premier temps à la décomposition en curvelets. Quoique cette décomposition soit mieux adaptée pour le traitement des discontinuités, le temps de calcul et les ressources mémoire nécessaires à son utilisation peuvent constituer une limite dans certaines applications. Dans une telle situation, les décompositions en ondelettes constituent une bonne alternative. Elles sont peu gourmandes en temps de calcul et en ressources mémoire et la littérature est abondante sur le sujet. La bibliothèque d'assimilation d'images construite dans le cadre de cette thèse est adaptée pour prendre en compte d'autres décompositions parmi lesquelles les décompositions en ondelettes. Il suffit d'implémenter la décomposition appropriée avec des interfaces adaptées pour la bibliothèque.

Décompositions tridimensionnelles

Les séquences d'images peuvent être considérées comme des signaux tridimensionnels. Dans ce cas, les décompositions tridimensionnelles sont plus appropriées. Elles permettraient à notre avis de capturer les structures caractéristiques des images ainsi que leur évolution dans le temps. Nous avons commencé une étude avec la décomposition tridimensionnelle en curvelets. Cette étude a constitué le sujet du stage de deuxième année d'école d'ingénieur de Habib Toye² que nous avons co-encadré. Les résultats préliminaires ont montré une meilleure capture des discontinuités dans les images avec la correction de nombreux défauts de reconstruction après seuillage sur la décomposition bidimensionnelle [Toye(2009)]. Les études quantitatives et la comparaison avec les résultats obtenus par la décomposition bidimensionnelle seront faites ultérieurement.

6.2.2 Vers des problèmes réalistes

Une étape importante dans la suite de ce travail consiste à coupler l'assimilation directe de séquences d'images aux modèles réalistes. C'est une des tâches qui est actuellement en exploration avec le modèle océanographique NEMO³. Pour y arriver, de nombreux défis sont à relever parmi lesquels :

1. la gestion des occlusions dues aux nuages dans les images satellites ;
2. la gestion des régions côtières dans les modèles océanographiques ;
3. la gestion de l'altitude des phénomènes observés dans le cas des modèles atmosphériques.

La gestion des occlusions a déjà été abordée dans plusieurs travaux. L'exemple des travaux de Papadakis [Papadakis and Mémin(2008b)] montre que l'assimilation directe d'images est plus appropriée dans ce contexte. La détermination de l'altitude des phénomènes observés dans l'atmosphère reste un grand challenge. [Michel(2008)] montre comment exploiter les images du

2. Étudiant en deuxième année d'école d'ingénieur à l'ENSIMAG pendant le stage qui a travaillé sur l'interfaçage de la décomposition en curvelets 3D dans la bibliothèque BALAISE.

3. NEMO (Nucleus for European Modelling of the Ocean) est le modèle océanographique européen utilisé pour la recherche, la prévision saisonnière en océanographie opérationnelle et l'étude du climat. Le lecteur trouvera plus de détails sur le site web du projet : <http://www.nemo-ocean.eu/>

canal vapeur d'eau pour la classification des phénomènes observés dans le cadre de l'estimation du tourbillon potentiel dans les images satellite. La gestion des régions côtières quant à elle demeure un problème très difficile. Les côtes constituent des frontières entre l'océan qui est le système étudié en océanographie et la terre ferme. Ce sont des frontières non régulières alors que les opérateurs mathématiques de traitement d'images sont en général définis sur des domaines rectangulaires.

6.2.3 Distance entre images

Le calcul de l'écart aux observations est une étape primordiale pour l'utilisation d'une classe d'observation quelconque en assimilation de données. Cet écart mesure la distance entre les sorties du modèle et les mesures. Il est ainsi indispensable pour chaque classe d'observation de définir une distance appropriée. Nous avons expérimenté plusieurs distances pour le calcul de l'écart aux observations de type image. Chacune de ces distances étant équivalente à la distance induite par la norme \mathcal{L}^2 et restreinte à une partie du signal. Les expériences montrent qu'un seuillage approprié permet d'obtenir de meilleurs résultats. C'est le cas du seuillage échelle par échelle sur la frame des curvelets. Cependant, avec une telle distance, la décroissance du terme d'écart aux observations n'est pas forcément un bon critère de quantification de l'amélioration du résultat. C'est ce que nous avons montré en comparant les résultats obtenus avec différents seuillages au chapitre (4). Nous l'avons également illustré au chapitre (5) en comparant les résultats obtenus avec différents termes de régularisation par pénalisation. Ce problème peut être résolu, au moins en partie avec un opérateur de préconditionnement approprié. Nous l'avons montré au chapitre (5) avec l'opérateur de préconditionnement issu de la régularisation par diffusion généralisée.

D'autres études ont déjà montré que la norme \mathcal{L}^2 n'était pas adaptée dans le contexte des images. Dans la littérature, on rencontrera notamment l'utilisation des normes de type \mathcal{L}^1 telle que la variation totale. Le problème principal avec les normes du type \mathcal{L}^1 est la non-différentiabilité des opérateurs associés. Cette non-différentiabilité constitue un obstacle pour l'assimilation variationnelle de données.

A notre avis, le problème posé par la norme \mathcal{L}^2 est une dérivée du problème d'ouverture. En effet, si une image f_2 est dérivée d'une image f_1 par une transformation simple telle qu'une translation, il est difficile de quantifier la translation sur la seule base de la norme \mathcal{L}^2 de la différence $f_2 - f_1$ entre les deux images. Pour illustrer ce problème, considérons une image définie par la fonction f_1 sur le domaine Ω . Nous supposons que f_1 est à support limité Ω_1 et constante sur ce support :

$$\begin{cases} f_1(\mathbf{x}) &= 1, \text{ si } \mathbf{x} \in \Omega_1, \\ f_1(\mathbf{x}) &= 0, \text{ sinon.} \end{cases} \quad (6.1)$$

La figure (6.1) illustre une telle fonction. Soit f_2 , la fonction définissant une seconde image dérivée de la première par translation de son support : les figures (6.2) et (6.3) illustrent la distance induite par la norme \mathcal{L}^2 dans le cas d'un support rectangulaire en dimension 2. Dans ces figures, nous avons représenté le support de chacune des fonctions par une zone hachurée.

Lorsque le support initial et le support final se recouvrent comme le montre la figure (6.2), la distance donnée par la norme \mathcal{L}^2 de la différence entre les deux images dépend de la direction de la translation. Si on déplace de δd selon la longueur (sous figure (6.2b)), on obtient une distance de $2l\delta d$ où l représente la largeur du support. Par contre, si on déplace suivant la largeur (sous figure (6.2a)), on obtient une distance de $2L\delta d$ où L est la longueur du support. Avec des déplacements d'amplitudes identiques, les distances sont différentes.

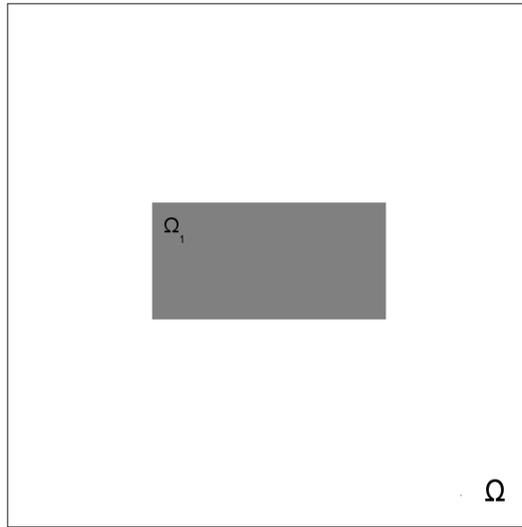


FIGURE 6.1 – Illustration d’une image définie par une fonction constante sur un support limité. Ω représente le domaine de définition de l’image et Ω_1 le support de la fonction associée

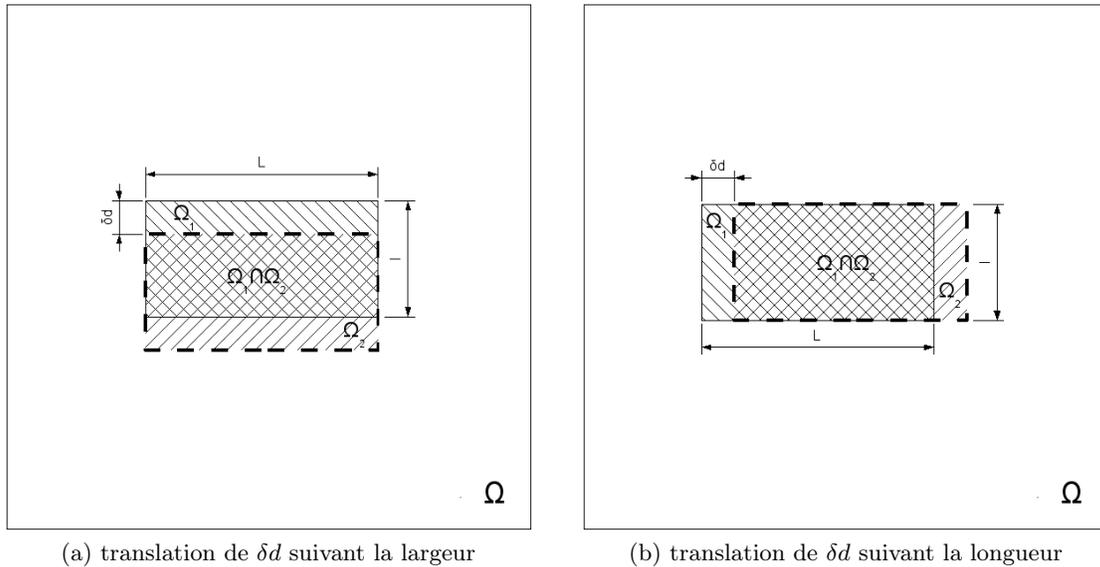


FIGURE 6.2 – Illustration de la distance définie par la norme \mathcal{L}^2 sur des translations de faible amplitude (Ω_1 représente le support de la fonction image avant la translation et Ω_1 le support après translation). (a) montre un déplacement de δd dans le sens de la largeur et (b) un déplacement de δd dans le sens de la longueur.

La figure (6.3) illustre une configuration dans laquelle le support initial et le support final sont disjoints. Dans ce cas, la distance donnée par la norme \mathcal{L}^2 est indépendante de l’amplitude de la translation tant que le support de la fonction image reste dans le domaine d’étude. Elle est de deux fois la surface du support ($2Ll$). Cet exemple sur la translation d’une forme élémentaire dans l’espace illustre les lacunes de la norme \mathcal{L}^2 pour le calcul de la distance entre deux images dans un cadre dynamique.

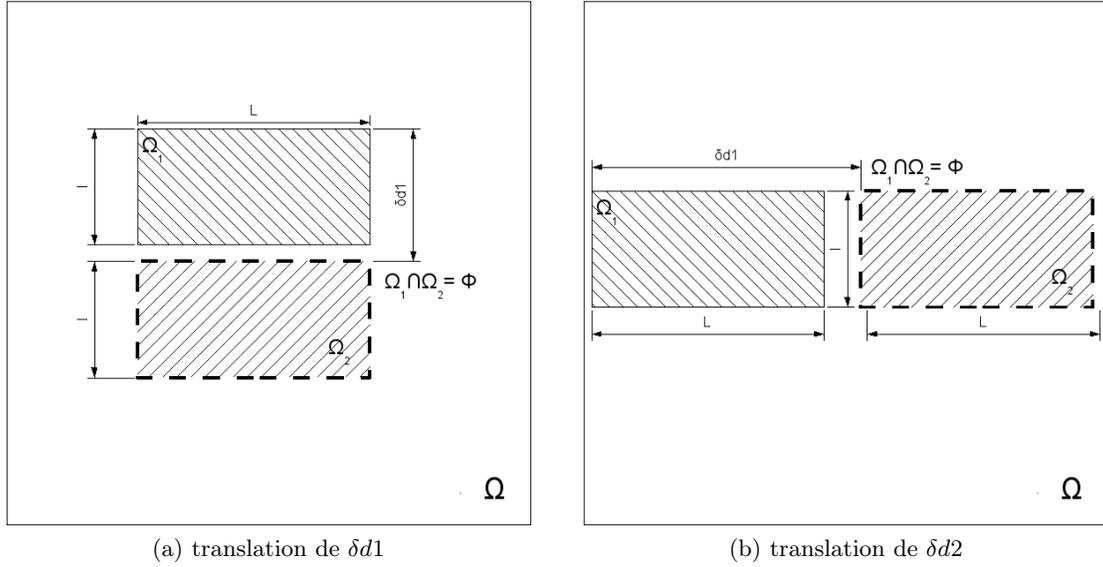


FIGURE 6.3 – Illustration de la distance définie par la norme \mathcal{L}^2 sur des translations de grandes amplitudes (Ω_1 représente le support de la fonction image avant la translation et Ω_2 , le support après translation). (a) montre un déplacement de $\delta d1$ dans le sens de la largeur et (b) un déplacement de $\delta d2$ dans le sens de la longueur.

Proposition 6.2.1 *Une distance appropriée pour le calcul de la distance entre deux images dans un cadre dynamique peut se définir par l'intégrale de la fonction d'intensité du mouvement sur le domaine étudié. Cette distance est donnée par :*

$$d(f_1(x, y), f_2(x, y)) = \int_{\Omega} \|m(f(x, y), f_2(x, y))\|, \quad (6.2)$$

avec d la fonction de distance, Ω le domaine étudié et m la fonction de mouvement entre les deux images f_1 et f_2 ; elle représente le champ de vecteur définissant le mouvement en chaque point du domaine d'étude Ω .

Les différentes techniques d'estimation du mouvement peuvent être utilisées pour le calcul de la fonction de mouvement m . Elles permettent d'estimer des mouvements plus complexes que de simples translations. Des études sur de telles distances sont actuellement menées au Laboratoire Jean Kuntzmann dans le cadre de la thèse de Vincent Chabot.

Dans le cas où la dynamique ne peut pas être réduite au mouvement des points de l'image, il faudrait prendre en compte dans la définition de la distance l'équivalent du phénomène qui est à l'origine de la transformation entre les images. Les principales difficultés demeurent le calcul du phénomène sous-jacent ou son équivalent ainsi sa différentiabilité pour l'assimilation variationnelle de données. Ce calcul peut être aussi difficile que le problème initial d'assimilation de séquences d'images.

Annexe A

Approches séquentielles de l'assimilation de données

Comme définie au chapitre 1, l'assimilation de utilise toutes les sources d'information disponibles sur un système afin déterminer de façon optimale les paramètres de ce système. L'approche séquentielle, encore appelée approche stochastique repose sur une analyse statistique de l'état analysé du système. L'objectif est de déterminer un état analysé optimal du système qui prenne en compte les observations. L'optimalité est basée sur une analyse statistique des erreurs d'analyse.

A.1 Interpolation optimale

Supposons qu'on connaisse une ébauche \mathbf{x}^b de l'état d'un système dont la vérité est \mathbf{x}^t , l'erreur d'ébauche ε^b est de moyenne nulle ($E(\varepsilon^b) = \langle \varepsilon^b \rangle = 0$) et de covariance $\mathbf{B} = \langle \varepsilon^b \varepsilon^{bT} \rangle$. On dispose en plus d'une observation \mathbf{y}^o d'erreur ε^o de moyenne nulle et de variance $\langle \varepsilon^o \varepsilon^{oT} \rangle$. ; la contrepartie modèle de l'observation est défini à travers l'opérateur d'observation linéaire $\mathbf{H}pary = \mathbf{H}\mathbf{x}$.

Le meilleur estimateur linéaire sans biais[Phillips(1982)], noté BLUE en abbréviation de son expression en anglais (Best Linear Unbiased Estimator) définit l'état analysé \mathbf{x}^a suivant l'équation :

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b). \quad (\text{A.1})$$

L'optimalité est donnée par le minimum de la trace de la matrice de covariance d'erreur sur l'analyse. Le but est de définir \mathbf{K} encore appelé matrice de gain ; $\mathbf{y}^o - \mathbf{H}\mathbf{x}^b$ est appelé vecteur innovation.

On minimise la trace $Tr(\mathbf{P}^a)$ pour obtenir le \mathbf{K} optimal, avec $\mathbf{P}^a = \langle \varepsilon^a \varepsilon^{aT} \rangle$. D'après l'équation (A.1), on peut déduire que :

$$\begin{aligned} \varepsilon^a &= \mathbf{x}^a - \mathbf{x}^t \\ &= \mathbf{x}^b - \mathbf{x}^t + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^t - \mathbf{H}\mathbf{x}^b + \mathbf{H}\mathbf{x}^t) \\ &= \varepsilon^b + \mathbf{K}(\varepsilon^o - \mathbf{H}\varepsilon^b). \end{aligned} \quad (\text{A.2})$$

On peut alors écrire :

$$\begin{aligned}
 \varepsilon^a \varepsilon^{aT} &= \left(\varepsilon^b + \mathbf{K}(\varepsilon^o - \mathbf{H}\varepsilon^b) \right) \left(\varepsilon^b + \mathbf{K}(\varepsilon^o - \mathbf{H}\varepsilon^b) \right)^T \\
 &= \varepsilon^b \varepsilon^{bT} + \varepsilon^b (\varepsilon^{oT} - \varepsilon^{bT} \mathbf{H}^T) \mathbf{K}^T + \mathbf{K}(\varepsilon^o - \mathbf{H}\varepsilon^b) \varepsilon^{bT} \\
 &\quad + \mathbf{K}(\varepsilon^o - \mathbf{H}\varepsilon^b) (\varepsilon^{oT} - \varepsilon^{bT} \mathbf{H}^T) \mathbf{K}^T \\
 &= \varepsilon^b \varepsilon^{bT} - \varepsilon^b \varepsilon^{bT} \mathbf{H}^T \mathbf{K}^T - \mathbf{K} \mathbf{H} \varepsilon^b \varepsilon^{bT} + \mathbf{K} \mathbf{H} \varepsilon^b \varepsilon^{bT} \mathbf{H}^T \mathbf{K}^T \\
 &\quad + \varepsilon^b \varepsilon^{oT} - \mathbf{K} \mathbf{H} \varepsilon^b \varepsilon^{oT} \mathbf{K}^T \\
 &\quad + \mathbf{K} \varepsilon^o \varepsilon^{bT} - \mathbf{K} \varepsilon^o \varepsilon^{bT} \mathbf{H}^T \mathbf{K}^T \\
 &\quad + \mathbf{K} \varepsilon^o \varepsilon^{oT} \mathbf{K}^T \\
 &= \varepsilon^b \varepsilon^{bT} - \varepsilon^b \varepsilon^{bT} \mathbf{H}^T \mathbf{K}^T - \mathbf{K} \mathbf{H} \varepsilon^b \varepsilon^{bT} + \mathbf{K} \left(\mathbf{H} \varepsilon^b \varepsilon^{bT} \mathbf{H}^T + \varepsilon^o \varepsilon^{oT} \right) \mathbf{K}^T \\
 &\quad + \varepsilon^b \varepsilon^{oT} + \mathbf{K} \varepsilon^o \varepsilon^{bT} - \mathbf{K} \left(\mathbf{H} \varepsilon^b \varepsilon^{oT} + \varepsilon^o \varepsilon^{bT} \mathbf{H}^T \right) \mathbf{K}^T. \tag{A.3}
 \end{aligned}$$

Ainsi :

$$\begin{aligned}
 \mathbf{P}^a &= \langle \varepsilon^a \varepsilon^{aT} \rangle \\
 &= \langle \varepsilon^b \varepsilon^{bT} \rangle - \langle \varepsilon^b \varepsilon^{bT} \rangle \mathbf{H}^T \mathbf{K}^T - \mathbf{K} \mathbf{H} \langle \varepsilon^b \varepsilon^{bT} \rangle + \mathbf{K} \left(\mathbf{H} \langle \varepsilon^b \varepsilon^{bT} \rangle \mathbf{H}^T + \langle \varepsilon^o \varepsilon^{oT} \rangle \right) \mathbf{K}^T \\
 &= + \langle \varepsilon^b \varepsilon^{oT} \rangle + \mathbf{K} \langle \varepsilon^o \varepsilon^{bT} \rangle - \mathbf{K} \left(\mathbf{H} \langle \varepsilon^b \varepsilon^{oT} \rangle + \langle \varepsilon^o \varepsilon^{bT} \rangle \mathbf{H}^T \right) \mathbf{K}. \tag{A.4}
 \end{aligned}$$

Sous l'hypothèse d'indépendance statistique entre les erreurs d'ébauche et les erreurs d'observation ($\langle \varepsilon^o \varepsilon^{bT} \rangle = 0$), on obtient :

$$\begin{aligned}
 \mathbf{P}^a &= \langle \varepsilon^b \varepsilon^{bT} \rangle - \langle \varepsilon^b \varepsilon^{bT} \rangle \mathbf{H}^T \mathbf{K}^T - \mathbf{K} \mathbf{H} \langle \varepsilon^b \varepsilon^{bT} \rangle + \mathbf{K} \left(\mathbf{H} \langle \varepsilon^b \varepsilon^{bT} \rangle \mathbf{H}^T + \langle \varepsilon^o \varepsilon^{oT} \rangle \right) \mathbf{K}^T \\
 &= \mathbf{B} - \mathbf{B} \mathbf{H}^T \mathbf{K}^T - \mathbf{K} \mathbf{H} \mathbf{B} + \mathbf{K} (\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T) \mathbf{K}^T. \tag{A.5}
 \end{aligned}$$

Moyennant de bonnes propriétés de convexité et de continuité, $Tr(\mathbf{P}^a)$ atteint son minimum quand sa dérivée s'annule :

$$\begin{aligned}
 \frac{dTr(\mathbf{P}^a)}{d} \mathbf{K} &= -\mathbf{B} \mathbf{H}^T - \mathbf{B}^T \mathbf{H}^T + 2\mathbf{K} (\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T) \\
 &= -2\mathbf{B} \mathbf{H}^T + 2\mathbf{K} (\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T). \quad \text{par symétrie de } \mathbf{B}. \tag{A.6}
 \end{aligned}$$

$$\begin{aligned}
 \frac{dTr(\mathbf{P}^a)}{d\mathbf{K}} &= 0 \\
 \Leftrightarrow \mathbf{K} &= \mathbf{B} \mathbf{H}^T (\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T)^{-1}. \tag{A.7}
 \end{aligned}$$

L'optimalité est donc définie par :

$$\mathbf{K} = \mathbf{B} \mathbf{H}^T (\mathbf{R} + \mathbf{H} \mathbf{B} \mathbf{H}^T)^{-1}. \tag{A.8}$$

On en déduit que :

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K} \mathbf{H}) \mathbf{B}. \tag{A.9}$$

A.2 Le filtre de Kalman

Dans cette section, nous nous intéressons à un système dynamique dont on connaît :

- le modèle, dont une approximation linéaire de la version discrète est donnée par :

$$\begin{cases} \mathbf{x}_{i+1} &= \mathbf{M}_i \mathbf{x}_i + \eta_i \\ \mathbf{x}_0 &\text{donné} \end{cases} \quad (\text{A.10})$$

où η_i est une quantification de l'erreur modèle ;

- une ébauche au temps initial de la variable d'état \mathbf{x}_0^b de matrice de covariance d'erreur \mathbf{B} ;
- des observations \mathbf{y}_k^o associées aux opérateurs d'observation linéaires \mathbf{H}_k et de matrice de covariances d'erreur \mathbf{R}_k aux pas de temps $0..N$.

Le filtre de Kalman [Gelb(1974), Jazwinski(1970)] fourni de façon séquentielle à chaque nouvelle observation \mathbf{y}_k^o un état analysé \mathbf{x}_k^a optimal au sens statistique prenant en compte l'ébauche initiale et toutes les observations $\mathbf{y}_i^o; 0 \leq i \leq k$.

Hypothèses : les erreurs d'ébauche ε_0^b , d'observations ε_k^o et modèles η_k sont supposées :

- gaussiennes ;
- statistiquement indépendantes ou non corrélées, ce qui se traduit par :

$$\begin{aligned} \langle \varepsilon_0^b \varepsilon_k^{oT} \rangle_{0 \leq k \leq N} &= \langle \eta_k \varepsilon_k^{oT} \rangle_{0 \leq k \leq N} \\ &= \langle \eta_k \varepsilon_0^{bT} \rangle_{0 \leq k \leq N} \\ &= \langle \varepsilon_k^o \varepsilon_l^{oT} \rangle_{0 \leq k, l \leq N} \\ &= \langle \eta_k \eta_l^T \rangle_{0 \leq k, l \leq N} \\ &= 0; \end{aligned} \quad (\text{A.11})$$

- non biaisées : $\langle \varepsilon_0^b \rangle = \langle \varepsilon_k^o \rangle_{0 \leq k \leq N} = \langle \eta_k \rangle_{0 \leq k \leq N} = 0$.

Etape d'analyse : soit \mathbf{x}_k^f (f pour filtré, voir paragraphe suivant) entaché de l'erreur ε_k^f de covariance \mathbf{P}_k^f , le principe est similaire à celui du BLUE pour chaque nouvelles observations :

$$\begin{cases} \mathbf{K}_k &= \mathbf{P}_k^f \mathbf{H}_k^T (\mathbf{R}_k + \mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T)^{-1} \\ \mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k (\mathbf{y}_k^o - \mathbf{H}_k \mathbf{x}_k^f) \\ \mathbf{P}_k^a &= (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_k^f \end{cases} \quad (\text{A.12})$$

Etape de prévision : étant données :

- $\mathbf{x}_0^f = \mathbf{x}_0^b$, $\mathbf{P}_0^f = \mathbf{B}$;
- \mathbf{y}_i^o , \mathbf{R}_k , \mathbf{H}_k ; $0 \leq i \leq k$;
- \mathbf{M}_i , \mathbf{Q}_k ; $0 \leq i \leq k$,

l'algorithme de Kalman définit la prévision à l'étape $k + 1$ comme :

$$\begin{cases} \mathbf{x}_{k+1}^f &= \mathbf{M}_k \mathbf{x}_k^a \\ \mathbf{P}_{k+1}^f &= \mathbf{M}_k \mathbf{P}_k^a \mathbf{M}_k^T + \mathbf{Q}_k \end{cases} \quad (\text{A.13})$$

Le Filtre de Kalman a été introduit en météorologie pour la détermination des conditions initiales par Jones [Jones(1965)].

A.3 Problèmes non linéaires : le filtre de Kalman étendu

Si on a un système dynamique dont le modèle d'évolution est non linéaire, le filtre de Kalman ne peut pas s'appliquer tel quel; on utilise une version évoluée appelée filtre de Kalman étendu (Extended Kalman Filter) [Jazwinski(1970)]. Soient :

$$\begin{aligned}
 \text{Modèle : } & \mathbf{x}_{k+1}^t = \mathcal{M}(\mathbf{x}_k^t) + \eta_k \\
 \text{Prévision : } & \mathbf{x}_{k+1}^f = \mathcal{M}(\mathbf{x}_k^a) \\
 \text{Erreur : } & \varepsilon_{k+1}^f = \mathbf{x}_{k+1}^f - \mathbf{x}_{k+1}^t = \mathcal{M}(\mathbf{x}_k^t) + \eta_k - \mathcal{M}(\mathbf{x}_k^a)
 \end{aligned} \tag{A.14}$$

Sous l'hypothèse que le modèle ne soit pas fortement non linéaire, On définit approximation linéaire locale (des détails sur le modèle linéaire tangent seront donnés plus loin) par :

$$\mathcal{M}(\mathbf{x}_k^t) = \mathcal{M}(\mathbf{x}_k^a) + \mathbf{M}(\mathbf{x}_k^t - \mathbf{x}_k^a) + O(\mathbf{x}_k^t - \mathbf{x}_k^a) \tag{A.15}$$

où \mathbf{M} représente le modèle linéaire tangent au point \mathbf{x}_k^a Ainsi :

$$\begin{aligned}
 \varepsilon_{k+1}^f &= \mathbf{M}(\varepsilon_k^a) + \eta_k \\
 \mathbf{P}_{k+1}^f &= \mathbf{M}\mathbf{P}_k^a\mathbf{M}^T + \mathbf{Q}_k
 \end{aligned} \tag{A.16}$$

On fait de même pour l'opérateur d'observation.

Remarques Dans la pratique, le Filtre de Kalman peut être limité par la dimension du problème. La construction de la matrice de covariance d'erreurs \mathbf{P}_{k+1}^f nécessite $2n$ applications du modèle à un vecteur (où n représente la taille du vecteur d'état); pour les problèmes météorologiques actuels, n est de l'ordre de 10^8 . De même, plus il y a d'observations, plus la taille de la matrice \mathbf{R} est grande, ce qui rend la construction de la matrice de gain lourde. Pour pallier à ce problème, on peut faire recours à la réduction d'ordre dont les principales stratégies sont :

- le Filtre de Kalman d'ensemble (EnKF pour Ensemble Kalman Filter); proposé par [Evensen(1997)], utilise une méthode de Monte Carlo pour une estimation efficace des matrices de covariance d'erreurs;
- le filtre SEEK (Singular Evolutive Extended Kalman); proposé par [Pham et al.(1996)], réduit l'ordre du système en se limitant aux modes dominants. Des exemples d'applications du filtre SEEK peuvent être trouvés dans [Ballabrera-Poy et al.(2001), Hoteit(2001)].

Annexe B

Méthodes de descente pour la minimisation

B.1 Principe général des méthodes de descente pour la minimisation

Le principe général des méthodes que nous allons décrire dans cette section peut se résumer par l'algorithme suivant :

Algorithm 2: Principe général des algorithmes déterministes de minimisation

Entrées: approximation initiale \mathbf{x}_0 pour $k = 0$

répéter

 Calculer la direction de recherche \mathbf{d}_k ;

 Déterminer le pas de progression dans la direction de recherche $\alpha_k > 0$;

 Calculer l'approximation suivante $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$

jusqu'à convergence ;

Les critères de convergence sont définis selon le problème et les ressources temps et matériels disponibles. Il est courant de se baser sur le résidu ou la progression de l'approximation au cours des itérations. La direction de recherche est le vecteur indiquant la direction dans laquelle il faut poursuivre la recherche et α_k indique la distance à parcourir dans cette direction.

Direction de recherche

La valeur de la fonction f à minimiser, habituellement appelée fonction objectif, doit diminuer au fur et à mesure des itérations : $f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) < f(\mathbf{x}_k)$. On dit aussi que la direction de recherche est une direction de descente.

De l'approximation tronquée de Taylor :

$$f(\mathbf{x} + \alpha \mathbf{d}) \approx f(\mathbf{x}) + \alpha (\nabla f(\mathbf{x}))^T \mathbf{d}, \quad (\text{B.1})$$

En imposant des pas de descente positifs ($\alpha > 0$), on déduit que les directions suivant lesquelles la dérivée directionnelle est négative ($\langle \nabla f(\mathbf{x}), \mathbf{d} \rangle < 0$) sont de bons candidats. Les méthodes itératives déterministes se distinguent par le choix de la direction de descente. On les divise en trois grandes classes :

- les méthodes de relaxation qui parcourent toutes les directions de l'espace pour rechercher la bonne ; ces méthodes sont très simples à mettre en œuvre mais très coûteuses en calcul quand la dimension du problème est grand ;
- les méthodes du gradient (basée le gradient) ;
- les méthodes de Newton qui utilisent le Hessien.

Nous allons nous limiter aux deux dernières.

Pas de recherche

Les méthodes de détermination du pas de descente ne seront pas abordés dans ce mémoire, le lecteur peut les se retrouver dans les ouvrages classiques d'optimisation tels que [Gould and Leyffer(2003), Dennis and Schnabel(1983), Luenberger(1984), Celis et al.(1985), Byrd et al.(1987)].

B.2 Méthodes du gradient

On retrouve dans cette classe, la méthode de descente de la plus grande pente (steepest descent) et le gradient conjugué.

B.2.1 Méthode de descente de la plus grande pente

La direction de recherche est choisie comme celle dans laquelle le gradient est négatif :

$$\mathbf{d}_k = -\nabla f, \quad (\text{B.2})$$

c'est le choix le plus simple pour avoir une dérivée directionnelle négative.

C'est une des méthodes les plus anciennes et une des plus simples à mettre en œuvre. Elle nécessite peu d'espace mémoire, $\mathcal{O}(n)$, où n est le nombre de paramètre à déterminer. Sa convergence est linéaire lorsqu'elle est appliquée à des problèmes quadratiques.

Quand la courbure des iso-surfaces de la fonction objectif n'est pas constante, ce qui est en général le cas quand on se rapproche de la solution, la progression est très lente comme l'illustre la figure B.1. Dans le premier cas, la direction de descente est orientée vers le minimum, ce qui n'est pas nécessairement vrai dans le second. On préfère alors des méthodes plus robustes telles que le gradient conjugué.

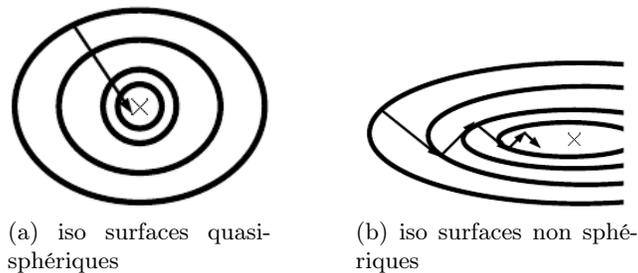


FIGURE B.1 – Comportement de la descente de plus grande pente

B.2.2 Méthode du gradient conjugué

Initialement conçu pour la minimisation de fonctions quadratiques et convexes, le gradient conjugué a été étendu au cas général [Powell(1984), Fletcher and Reeves(1964)]. Avec cette méthode, la direction de recherche est calculée comme :

$$\begin{cases} \mathbf{d}_k &= \nabla f(\mathbf{x}_k) + \beta_k \mathbf{d}_{k-1} \\ \mathbf{d}_0 &= \nabla f(\mathbf{x}_0). \end{cases} \quad (\text{B.3})$$

Le paramètre β_k pouvant être choisi selon une des formules proposées par Fletcher-Reeves (FR), Polak-Ribière (PR) ou encore Hestenes-Stiefel (HS) du nom de leurs auteurs :

$$\beta_k^{FR} = \langle \nabla f(\mathbf{x}_k), \nabla f(\mathbf{x}_k) \rangle / \langle \nabla f(\mathbf{x}_{k-1}), \nabla f(\mathbf{x}_{k-1}) \rangle \quad (\text{B.4})$$

$$\beta_k^{PR} = \langle \nabla f(\mathbf{x}_k), \nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}_{k-1}) \rangle / \langle \nabla f(\mathbf{x}_{k-1}), \nabla f(\mathbf{x}_{k-1}) \rangle \quad (\text{B.5})$$

$$\beta_k^{HS} = \langle \nabla f(\mathbf{x}_k), \nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}_{k-1}) \rangle / \langle \mathbf{d}_{k-1}, \nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}_{k-1}) \rangle \quad (\text{B.6})$$

Les directions de recherche ainsi obtenus sont dites conjuguées ; deux directions successives ne peuvent être identiques.

Le choix de directions conjuguées permet de corriger le problème rencontré par la méthode de descente de plus grande pente en présence d'iso surfaces non sphériques. Le préconditionnement permet d'éviter la sensibilité au problème d'arrondi et accélère la convergence [Golub(1983), Evans(1967), Concus et al.(1976)]. Le gradient conjugué a une consommation faible en espace mémoire ($\mathcal{O}(n)$), et de plus bonnes propriétés de convergence comparé à la descente de plus grande pente. C'est une des méthodes les plus utilisées dans la résolution de la plupart des problèmes.

B.3 Méthodes de Newton

B.3.1 Principe des méthodes de Newton

Ces méthodes s'inspirent de l'approximation quadratique de la fonction objectif au point de recherche et dans la direction de recherche :

$$f(\mathbf{x}_k + \mathbf{d}) = f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{d} \rangle + \frac{1}{2} \langle \mathbf{d}, \mathbf{H}_k \mathbf{d} \rangle \quad (\text{B.7})$$

où \mathbf{H}_k est le Hessien de f au point \mathbf{x}_k .

Cette approximation atteint son minimum en \mathbf{d}_k , minimum de la fonction quadratique :

$$q_{\mathbf{H}_k}(\mathbf{d}) = \langle \nabla f(\mathbf{x}_k), \mathbf{d} \rangle + \frac{1}{2} \langle \mathbf{d}, \mathbf{H}_k \mathbf{d} \rangle = \langle \mathbf{d}, \nabla f(\mathbf{x}_k) + \mathbf{H}_k \mathbf{d} \rangle \quad (\text{B.8})$$

Ce minimum est atteint au point \mathbf{d} vérifiant l'équation :

$$\mathbf{H}_k \mathbf{d} = -\nabla f(\mathbf{x}_k) \quad (\text{B.9})$$

connue sous le nom équation de Newton.

Cette méthode a une convergence quadratique quand \mathbf{x}_0 est suffisamment proche de la solution . Une description détaillée de la méthode peut se trouver dans [Schnabel et al.(1985), Luenberger(1984)]. Cette méthode utilise l'information du Hessian, et donc de la courbure locale des iso surfaces, ce qui lui confère un avantage par rapport aux autres méthode de

type descente du gradient. Tel que décrit, cet algorithme serait difficile à mettre en œuvre dans un cas pratique d'assimilation de données météorologique où la variable d'état est de l'ordre de 10^8 pour les modèles globaux actuellement ; il nécessite l'inverse de la Hessienne au point \mathbf{x}_k où la résolution d'un système avec la hessienne comme matrice à chaque étape, alors que le stockage de cette hessienne n'est déjà pas évident vu qu'elle est de l'ordre de 10^{16} éléments. Dans la pratique, on utilise des approximations de type quasi-newton que nous décrivons ci-dessous.

B.3.2 Algorithmes de type quasi-Newton

Principe

L'idée est de remplacer la Hessienne \mathbf{H} (ou son inverse \mathbf{W}) par une suite d'approximations symétriques définies positives que l'on met à jour à chaque étape ; le coût de ce procédé est relativement faible. Pour passer de l'étape k à l'étape $k + 1$, on pose :

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \gamma \mathbf{u}\mathbf{u}^T + \beta \mathbf{v}\mathbf{v}^T \quad (\text{B.10})$$

où γ et β sont des scalaires et \mathbf{u}, \mathbf{v} des vecteurs.

Le principe peut se résumer comme suit :

- à l'étape k , si on a une approximation \mathbf{W}_k de $[\mathbf{H}_k]^{-1}$, on calcule la direction de descente $\mathbf{d}_k = -\mathbf{W}_k \nabla f(\mathbf{x}_k)$;
- on cherche le pas de descente α_k par l'une des méthodes citées plus haut,
- on pose $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$;
- puis on calcule une approximation \mathbf{W}_{k+1} de $[\mathbf{H}_{k+1}]^{-1}$ avec la correction de rang 2 présentée à l'équation (B.10) en utilisant uniquement $\mathbf{x}_k, \mathbf{x}_{k+1}, \nabla f(\mathbf{x}_k)$ et $\nabla f(\mathbf{x}_{k+1})$.

Le travail consiste maintenant à trouver une bonne formule de mise à jour.

Algorithme BFGS

Du nom de ses auteurs Broyden, Fletcher, Goldfarb et Shanno, BFGS est un algorithme de type quasi-Newton qui met à jour l'approximation de la Hessienne inverse par :

$$\mathbf{W}_{k+1} = \mathbf{U}(\mathbf{W}_k, \mathbf{s}_k, \mathbf{y}_k) = \left(\mathbf{I} - \frac{\mathbf{s}_k \otimes \mathbf{y}_k}{\langle \mathbf{y}_k, \mathbf{s}_k \rangle} \right) \mathbf{W}_k \left(\mathbf{I} - \frac{\mathbf{y}_k \otimes \mathbf{s}_k}{\langle \mathbf{y}_k, \mathbf{s}_k \rangle} \right) + \frac{\mathbf{s}_k \otimes \mathbf{s}_k}{\langle \mathbf{s}_k, \mathbf{s}_k \rangle} \quad (\text{B.11})$$

Avec $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$, $\mathbf{y}_k = \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$ et \otimes est le produit tensoriel associé au produit scalaire $\langle \cdot, \cdot \rangle$ et est défini par $[\mathbf{u} \otimes \mathbf{v}] \mathbf{d} = \langle \mathbf{v}, \mathbf{d} \rangle \mathbf{u}$. Cet algorithme nécessite le stockage des \mathbf{W}_k qui requiert beaucoup d'espace de stockage, on est amené dans un but pratique à ne stocker que \mathbf{W}_0 et les différentes paires $\{\mathbf{s}_k, \mathbf{y}_k\}$. Ce qui nécessite tout de même trop d'espace de stockage dans les cas où la dimension du problème est important comme en assimilation de données appliquée à la météorologie, de l'ordre de 10^8 .

Algorithme L-BFGS

de Limited Memory BFGS, L-BFGS est dérivé de l'algorithme BFGS en limitant le nombre

de paires $\{\mathbf{s}, \mathbf{y}\}$ stockée; le nombre M de vecteurs réellement stockés est fixé à l'avance et on ne conserve que les M dernières paires $\{\mathbf{s}, \mathbf{y}\}$. W est alors mis à jour comme suit :

$$\left\{ \begin{array}{l} 0 \leq k \leq M - 1 : \quad \mathbf{W}_{k+1} = \mathbf{U}(\mathbf{W}_k, \mathbf{s}_k, \mathbf{y}_k), \\ k \geq M + 1 : \quad \left\{ \begin{array}{l} \mathbf{W}_k^0 = \mathbf{D}_k \\ \mathbf{W}_k^{j+1} = \mathbf{U}(\mathbf{W}_k^j, \mathbf{s}_{k-M+j}, \mathbf{y}_{k-M+j}), 0 \leq j \leq M - 1 \\ \mathbf{W}_k = \mathbf{W}_k^M \end{array} \right. \end{array} \right. \quad (\text{B.12})$$

Où \mathbf{D}_k est une matrice diagonale.

Mise à jour de la matrice diagonale Il existe plusieurs formules de mise à jour de la matrice diagonale D_k permettant une bonne initialisation de l'algorithme L-BFGS [Auroux(2003)]. Nous reprenons la formule dite BFGS inverse qui met à jour la i^e composante comme suit :

$$\mathbf{D}_+^{(i)} = \mathbf{D}^{(i)} + \left(\frac{1}{(\mathbf{y}, \mathbf{s})} + \frac{(\mathbf{D}\mathbf{y}, \mathbf{y})}{(\mathbf{y}, \mathbf{s})^2} \right) (\mathbf{s}, \mathbf{e}_i)^2 - \frac{2\mathbf{D}^{(i)}(\mathbf{y}, \mathbf{e}_i)(\mathbf{s}, \mathbf{e}_i)}{(\mathbf{y}, \mathbf{s})} \quad (\text{B.13})$$

où $\mathbf{e}_i, 1 \leq i \leq n$ désigne une base orthonormée pour le produit scalaire $\langle \cdot, \cdot \rangle$

Annexe C

Modèles d'atmosphère et observations conventionnelles

C.1 Grandeurs physiques caractéristiques du système

Les principales variables d'un modèle mathématique de l'atmosphère sont le vent (vitesse et direction), la température, l'humidité et la pression. Ce sont les grandeurs physiques nécessaires à la description mathématique de l'état de l'atmosphère. Leur choix est guidé par les lois de la physique qui régissent le comportement du système. À ces grandeurs physiques caractéristiques du système s'ajoutent d'autres variables permettant une description plus fine du système ainsi que d'autres diagnostics dont les plus usuels sont les mesures de précipitation, de couverture nuageuse, de visibilité, d'ozone et de rayonnements divers.

C.1.1 Le vent

Le vent est la principale variable caractéristique qui nous intéressera. En effet, le travail présenté dans ce mémoire a été motivé par l'importance de l'information dynamique observable à travers les séquences d'images. En météorologie, on utilise l'expression **vent** pour qualifier la vitesse du déplacement horizontal de l'air par rapport à la surface de la Terre. Le déplacement vertical, en moyenne beaucoup plus faible comparé au déplacement horizontal est appelé vitesse verticale. Ainsi défini, le vent peut être décrit par un vecteur avec son module (intensité) et sa direction. L'intensité du vent que l'on appelle couramment vitesse ou force du vent est un ordre de grandeur physique mesurable au sol à l'aide d'un anémomètre. elle est exprimé en mètre par seconde ($m.s^{-1}$) dans le système international. Il est courant en météorologie de l'exprimer en kilomètre par heure ($km.h^{-1}$) ou en nœud (kt). Il est important de distinguer la vitesse instantanée qui peut être très fluctuante de la vitesse moyenne qui est définie par la moyenne arithmétique de la vitesse instantanée sur un intervalle de temps caractéristique relativement long par rapport à la fréquence de mesure des valeurs instantanées. Par exemple les principales stations d'observation du vent effectuent des mesures instantanées toutes les demi-secondes et les moyennes sont calculées sur des périodes de l'ordre de la dizaine de minutes. La direction du vent (mesurable au sol à l'aide d'une girouette) désigne la direction angulaire d'où vient le vent. Celle-ci est repérée par rapport aux directions cardinales.

Il est courant de décomposer le vent suivant les axes d'un repère. En météorologie, on le décompose dans le repère géographique. On parlera ainsi du vent méridien et du vent zonal.

Le vent méridien désigne la composante ouest-est et le vent zonal représente la composante sud-nord. Ce sont les composantes dans un repère d'étude local du vent dans le voisinage de la verticale d'un point fixe \mathbf{P} de la surface terrestre. Le point \mathbf{P} étant considéré comme l'origine du repère, la tangente en \mathbf{P} au parallèle passant par \mathbf{P} et orientée vers l'est est assimilé à l'axe $\mathbf{x}'\mathbf{x}$ des abscisses, la tangente en \mathbf{P} au méridien passant par \mathbf{P} et orientée vers le nord est considéré comme l'axe $\mathbf{y}'\mathbf{y}$ des ordonnées et la verticale en \mathbf{P} orientée vers le zénith est associé à l'axe $\mathbf{z}'\mathbf{z}$. On assimile localement la sphère terrestre à un plan horizontal passant par \mathbf{P} . Ce qui permet d'étudier le vent dans le voisinage de \mathbf{P} avec ses coordonnées dans le repère orthonormé précédemment décrit en considérant ses coordonnées \mathbf{u} (vent méridien) et \mathbf{v} (vent zonal) suivant les axes $\mathbf{x}'\mathbf{x}$ et $\mathbf{y}'\mathbf{y}$. Ces deux systèmes de représentation nous seront utiles pour les analyses d'incertitudes et la définition de l'erreur sur le vent en terme de :

- coordonnées dans le repère local ;
- vitesse et son orientation ; l'incertitude en orientation définissant l'erreur angulaire de Barron [Barron et al.(1994)] que nous décrirons dans les chapitres suivants.

En météorologie, il est courant de décomposer le vent en composantes **géostrophique** et **agéostrophique**. Le vent géostrophique se définit comme la résultante de l'équilibre géostrophique entre la force de Coriolis et la force du gradient de pression atmosphérique agissant sur une parcelle d'air. Il souffle parallèlement aux isobares dans l'atmosphère. C'est la composante principale de l'écoulement atmosphérique au-dessus de la couche limite en dehors des tropiques, ce qui en fait une approximation communément utilisée en météorologie. Le vent agéostrophique quant à lui désigne la différence entre le vent total et le vent géostrophique. A l'échelle synoptique, il est relativement faible et est souvent négligé en première approximation dans un contexte de diagnostic. Le vent agéostrophique est cependant d'une importance capitale. C'est cette composante du vent qui définit l'évolution future de l'écoulement. Il dénote en effet la présence de forçages qui sont à l'origine de la convergence ou de la divergence dans l'écoulement, d'où le mouvement vertical, source de plusieurs phénomènes météorologiques tels que les dépressions dans les latitudes moyennes.

C.1.2 La température

La température est une grandeur physique qui caractérise l'agitation moléculaire moyenne. Elle est proportionnelle à l'énergie cinétique moyenne associée aux mouvements désordonnés des molécules dans la matière. De façon qualitative, la température désigne la sensation relative de type chaud ou froid ressentie. L'unité du système international de mesure de la température est le kelvin (K). Dans cette unité, la température est toujours positive, la référence zéro correspondant à un état de repos de la matière. Il est courant en météorologie d'utiliser le degré Celsius ($^{\circ}\text{C}$) comme unité de mesure de la température. La variation d'une unité étant identique dans les deux unités.

C.1.3 L'humidité

L'humidité désigne l'évaluation quantitative de la présence de vapeur d'eau dans l'air d'une région de l'atmosphère.

C.1.4 La pression atmosphérique

En terme scientifique, la pression en un point d'un fluide caractérise les contraintes internes qui s'appliquent, par unité de surface, dans la direction normale à un élément de surface

quelconque passant par ce point. Dans l'atmosphère au repos, cette grandeur se confond pratiquement avec le poids de la colonne d'air de section horizontale égale à l'unité de surface et située au dessus du point de mesure. L'unité de mesure de la pression dans le système international est le pascal (Pa) qui correspond à une pression d'un newton par mètre carré. La pression est une grandeur physique d'une importance capitale dans la caractérisation des phénomènes météorologiques.

C.1.5 Les précipitations

Les précipitations désignent en météorologie, un ensemble organisé de particules d'eau liquide ou solide tombant en chute libre au sein de l'atmosphère. On y recense notamment la pluie, la bruine, la neige, la grêle et le grésil. Les précipitations se mesurent par la hauteur du liquide tombée sur une unité de surface. Elles s'expriment en millimètre (mm).

C.1.6 Les nuages

Un nuage est un amas de minuscules gouttelettes d'eau et/ou de cristaux de glace en suspension dans l'air et dont la frontière basse (base du nuage) ne touche pas le sol. Ceci exclut donc la brume et le brouillard. Les nuages sont le résultat de processus complexes au niveau de la (micro)physique des particules d'eau qui les composent.

C.2 Observations conventionnelles

C.2.1 Réseau d'observations *in situ*

On désigne par mesures *in situ* les observations réalisées par des instruments géographiquement situés dans la région où la mesure est effectuée. Ces instruments permettent d'échantillonner un volume d'air très petit (comparé à celui des phénomènes météorologiques décrits). Les observations conventionnelles sont en général des mesures directes des variables du modèle et peuvent donc être utilisées telles quelles dans un système d'assimilation de données (après éventuellement une interpolation). La présence de l'instrument dans la région échantillonnée permet d'obtenir des observations bien localisées avec une grande précision.

Le réseau synoptique

Ainsi désigne-t-on le réseau mondial soutenu par l'OMM. Il permet de décrire assez fidèlement les phénomènes météorologiques à l'échelle synoptique.

La figure (C.1) montre la carte des mesures opérationnelles du réseau synoptique et des observations par bateau le 24 février 2010 à 00 TU. Cette carte montre l'inégale répartition des mesures sur le globe terrestre. La couverture est très faible dans de nombreuses régions à l'intérieur des tropiques, les zones désertiques, les océans et les régions polaires. Ces observations du réseau synoptiques sont complétées par d'autres observations telles que ceux des réseaux de bouées et les réseaux de mesure en altitude.

Le réseau d'observation par les bouées en mer, voir la figure (C.2) est plus équitablement réparti sur les océans mais reste faible en densité. Les réseaux de mesure par avions de ligne ont une répartition similaire à celle du réseau synoptique. Dans cette classe de mesures, on distingue les réseaux ACARS (automatic air craft reports) et AMDAR (Aircraft Meteorological Data Relay), voir la figure (C.3). Le réseau de mesure en altitude par ballons sondes est encore

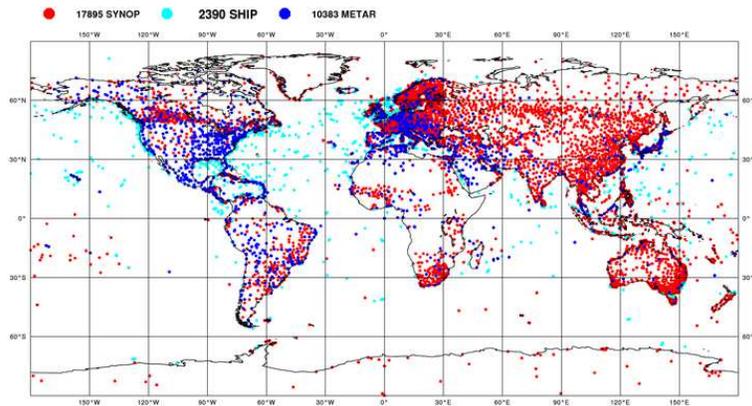


FIGURE C.1 – Observations du réseau synoptique et bateau de ligne. Répartition/couverture des données utilisées au centre européen pour les prévisions à moyen terme (données du 24 février 2010 à 00 UTC). Le nombre total d'observations est de 30668 : en rouge, données du réseau synoptique ; en cyan, données des bateaux, en bleu, observation du réseau METAR. Source : <http://www.ecmwf.int/>.

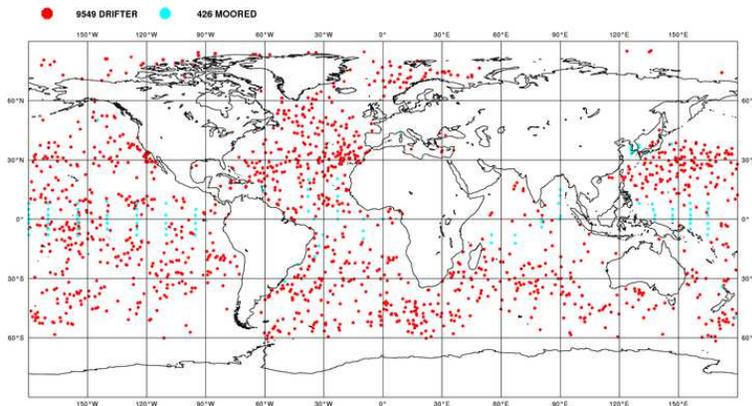


FIGURE C.2 – Observations par bouées. Répartition des données de bouées utilisées au centre européen pour les prévisions à moyen terme (données du 24 février 2010 à 00 UTC). Le nombre total d'observations est de 9975 : en rouge, bouées dérivantes, en cyan, bouées fixes. Source : <http://www.ecmwf.int/>.

plus faible et plus localisé. Le coût est en effet assez élevé pour ce type de mesure. Cependant, l'atmosphère étant moins perturbé en altitude qu'au niveau du sol où elle interagit avec la surface, la quantité d'informations nécessaire en altitude pour une égale description au niveau de la surface est moins importante.

Réseaux régionaux

Certains pays mettent en place des réseaux régionaux capables de décrire l'évolution des phénomènes d'échelle plus petite. Ces réseaux se construisent par installation de postes d'observation supplémentaires à ceux du réseau synoptique.

Réseaux adaptatifs

Afin de prendre en compte certains phénomènes particuliers à des échelles relativement fines, certains pays mettent en place des réseaux dits **adaptatifs**. C'est le cas des États-Unis ou

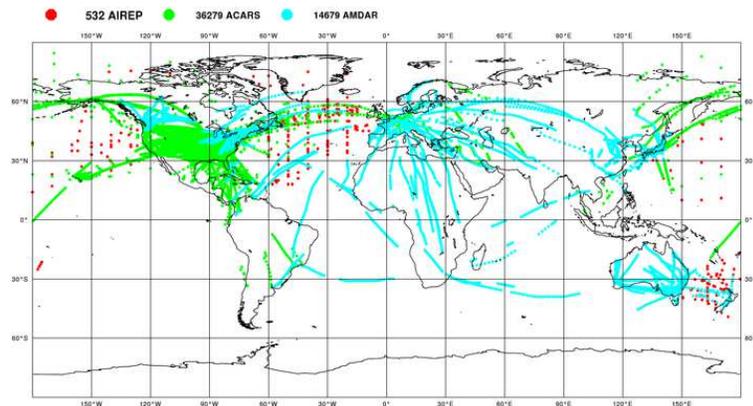


FIGURE C.3 – Observations par avions de ligne (réseaux AIREP, ACARS et AMDAR) ; Répartition des mesures utilisées au centre européen pour les prévisions à moyen terme (données du 24 février 2010 à 00 UTC). Le nombre total d’observations est de 51490 : en rouge, réseau AIREP ; en vert clair, réseau ACARS ; en cyan, réseau AMDAR. Source : <http://www.ecmwf.int/>.

encore de la France. Le réseau est adapté en fonction de la nature des phénomènes météorologiques qui prennent naissance ou ont souvent lieu dans la région. Les stations mobiles sont conçues dans l’idéologie de l’adaptativité des réseaux. Elles sont déployées en fonction de l’évolution des phénomènes.

C.2.2 Télémétries

Télémétries par radar

Le radar (RADio Detection And Ranging) est un système émetteur-récepteur d’un rayonnement électromagnétique de grandes longueurs d’onde. Ces grandes longueurs d’onde interagissent très peu avec les gaz atmosphériques, les aérosols et les nuages ; ce qui permet au rayonnement de traverser aisément l’atmosphère libre. Cette gamme de longueur d’onde associe au rayonnement radar une grande diffusivité au contact des grosses particules telles que les gouttes de pluie, les flocons de neige et la grêle. Le système envoie des impulsions puis mesure le rayonnement réfléchi dans sa direction. De ce fait on range les mesures par radar dans la classe des télémétries actives. Ils sont utilisés notamment pour la localisation et la mesure de l’intensité des précipitations et dans certains cas pour la mesure des vents par effet DOPPLER. Le temps de parcours du chemin aller-retour entre le radar et la cible permet de déterminer la distance entre les deux et l’intensité du rayonnement réfléchi donne des informations sur l’intensité des précipitations. Des légers décalages dans la longueur d’onde du rayonnement réfléchi en direction du radar apparaissent lorsque la particule cible est en mouvement (c’est ce qu’on l’on appelle effet Doppler). Les radars équipés de cette technologie permettent ainsi d’estimer la composante radiale de la vitesse du vent.

Télémétries par satellites

Les deux dernières décennies ont vu croître les observations satellitaires en quantité, en qualité et en diversité. Les centres météorologiques opérationnels reçoivent par jour des données de plusieurs satellites. Chaque satellite étant équipé de plus d’une technologie de mesure, les sources de données sont ainsi très nombreuses. Par exemple, le centre européen de prévision

à moyen terme exploitait déjà en 2003, 28 sources de données de 15 satellites différents. On distingue deux principaux types de satellites d'observations selon leurs orbites autour de la Terre : les satellites à orbite quasi polaire et les satellites géostationnaires.

Satellites à orbite polaire

Les satellites à orbite polaire ou quasi polaire tournent autour de la Terre à une altitude comprise entre 400 et 800 km. Cette altitude basse leur permet de couvrir tout le spectre du rayonnement (bande des micro-ondes comprise) contrairement aux satellites géostationnaires que nous décriront par la suite. Cependant, à cette altitude, la couverture géographique est assez étroite ; il faut plusieurs jours (sur certaines orbites) pour que le satellite passe au dessus du même point. Ils ont besoins de plusieurs orbites pour couvrir le globe terrestre en entier.

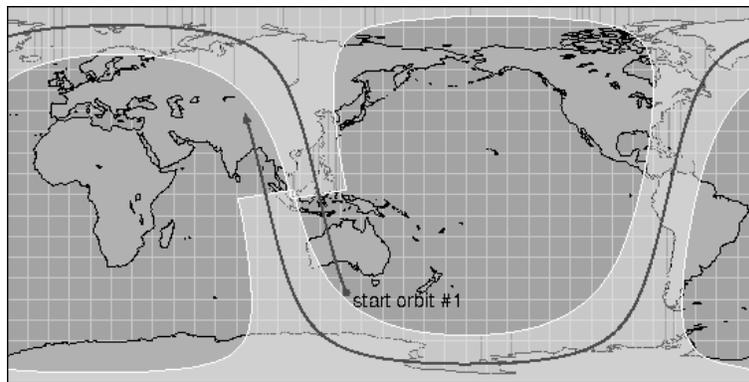


FIGURE C.4 – Trace d'un satellite à orbite quasi polaire lors d'une révolution autour de la Terre

La figure (C.4) montre le tracé de la couverture au sol d'un satellite à orbite quasi polaire lors d'une révolution autour de la Terre. Il faut une constellation de plusieurs satellites pour avoir un échantillonnage temporel raisonnable. Cependant leur proximité du sol permet d'avoir une résolution plus fine au point sous satellite. Malgré les passages peu réguliers en temps au dessus du même point, à chaque révolution autour de la Terre, ces satellites passent au dessus des pôles et permettent ainsi d'obtenir plus fréquemment des informations dans ces zones peu couvertes par les satellites géostationnaires et presque inaccessible par les instruments conventionnels.

La figure (C.5) montre les observations par le réseau des satellites équipés de la technologie AMSUA (Advanced Microwave Sounding Unit A en anglais). AMSU-A est un sondeur mesurant les profils de température de l'atmosphère et fournit de l'information sur l'humidité (eau sous toutes ses formes exception faite des particules de glace) de l'atmosphère. Cette technologie équipe les satellites NOAA (15 et 19), AQUA, et METOP-A fournit des données utilisées dans les systèmes d'assimilation de données des principaux centres de prévision météorologiques opérationnels.

Les satellites géostationnaires

Les satellites géostationnaires sont situés sur une orbite à environ 36000 km dans le plan équatorial au dessus de la Terre. A cette altitude, ils tournent à la même vitesse que la Terre, leur mouvement est ainsi imperceptible de la Terre, d'où l'appellation géostationnaire. Les satellites géostationnaires ont une large couverture géographique et une fréquence d'acquisition assez élevée : La plupart des satellites opérationnels actuels fournissent une observation

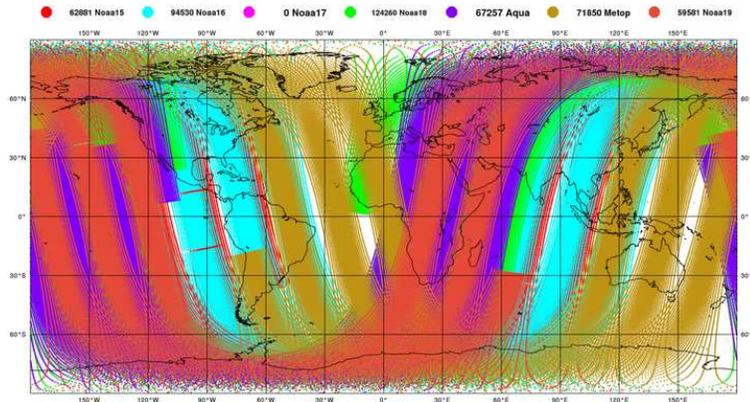


FIGURE C.5 – Observations par le sondeur AMSU-A (Advanced Microwave Sounding Unit A) à bord des satellites à orbite quasi polaire (NOAA, AQUA et METOP). Répartition des mesures utilisées au centre européen pour les prévisions à moyen terme (données du 24 février 2010 à 00 UTC). Le nombre total d'observations est 480359 : en rouge, NOAA15 (62881) ; en cyan, NOAA16 (94530) ; en vert clair, NOAA18 (124260) ; en violet, AQUA (67257) ; en jaune, METOP (71850), en rouge ocre, NOAA19 (59581). Source : <http://www.ecmwf.int/>.

complète de la zone couverte toutes les 15 minutes. C'est le cas des satellites européens METEOSAT 8 et 9. Leur positionnement au dessus de l'équateur ne leur permet pas d'observer les régions polaires et leur altitude ne leur permet pas de faire des acquisitions dans la bande des micro-ondes.

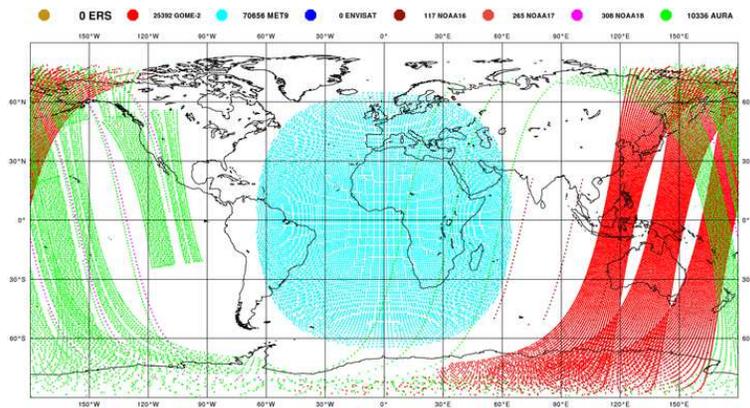


FIGURE C.6 – Observations de concentration d'ozone par les sondeurs à bords des satellites géostationnaires et à orbite quasi polaire. Répartition des mesures utilisées au centre européen pour les prévisions à moyen terme (données du 24 février 2010 à 00 UTC). Le nombre total d'observations est de 107074 : en rouge vif, GOME-2 (25392) ; en cyan, METEOSAT 9 (70656) ; en rouge foncé, NOAA16 (117) ; en rouge ocre, NOAA17 (265) ; en magenta, NOAA18 (308) et en vert AURA (10336) Source : <http://www.ecmwf.int/>.

La figure (C.6) montre les mesures d'ozone (le 24 février 2010) par le satellite géostationnaire européen METEOSAT 9 et les satellites à orbite quasi polaires GOME-2, NOAA (16, 17 et 18) et AURA. Cela illustre la complémentarité (spatiale) des stations d'observation.

Annexe D

Estimation du mouvement

D.1 Description de l'estimation du mouvement

L'estimation du mouvement désigne l'ensemble des techniques qui permettent de déterminer les vecteurs du mouvement décrivant la transformation d'une image en à une autre. La technique la plus utilisée pour l'estimation du mouvement est la technique du flot optique de Horn et Schunck [Horn and Schunck(1981)] basée sur l'hypothèse de la conservation de la luminance : on suppose que l'intensité de gris (pour les images en niveau de gris) ou la couleur (pour les images en couleur) des points d'une image ne varie pas au cours du déplacement. La présentation sera limitée ici aux images en deux dimensions. La variable d'espace sera notée $\mathbf{x} = (x, y)$, la variable temps t , la fonction image (luminance) sera notée, $f = f(x, y, t)$ et le champ de vitesse ou champ de vecteurs du mouvement $\mathbf{v} = \mathbf{v}(\mathbf{x}) = (v_1(\mathbf{x}), v_2(\mathbf{x}))$.

Avec ces notations, l'hypothèse de conservation de luminance s'écrit :

$$\frac{df}{dt} = 0, \tag{D.1}$$

et se traduit par :

$$\frac{\partial f}{\partial x} v_1 + \frac{\partial f}{\partial y} v_2 + \frac{\partial f}{\partial t} = 0. \tag{D.2}$$

Dans le cas des fluides géophysiques, un modèle plus réaliste prend en compte la conservation de la masse [Fitzpatrick(1985), Fitzpatrick(1988), Béréziat et al.(2000)], l'hypothèse de conservation s'écrit :

$$\frac{\partial f}{\partial t} + f \left(\frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y} \right) = 0. \tag{D.3}$$

Etant donnée l'image au temps zéro par sa fonction de luminance $f(x, y, 0) = f^0(x, y)$, la solution de l'équation (D.1) ou de l'équation (D.3) définit l'image au temps t comme une fonction du champ de mouvement statique \mathbf{v} et de l'image f^0 . Si au lieu de disposer du champ de mouvement \mathbf{v} , on a plutôt l'image au temps t , on définit alors l'estimation du mouvement comme le problème inverse de détermination de \mathbf{v} à partir de l'image au temps t et de l'image au temps zéro f^0 .

D.2 Modèle expérimental d'estimation du mouvement

Pour les expériences dans ce document, nous utilisons une équation de transport (D.4) d'un traceur passif, elle est similaire à l'équation de conservation de luminance :

$$\frac{\partial f}{\partial x}v_1 + \frac{\partial f}{\partial y}v_2 + \frac{\partial f}{\partial t} = 0, \quad (\text{D.4})$$

où $f = f(x, y, t)$ représente la concentration du traceur passif pouvant être assimilé à la luminance dans les images. Etant donnée la concentration du traceur $f(x, y, t = 0) = f^0(x, y) = f^0$ au temps 0 et le champ de vitesse du mouvement $\mathbf{v} = \mathbf{v}(x, y)$, la concentration du traceur $f = f(t) = f(x, y, t)$ au temps t s'obtient par intégration de l'équation du transport (D.4). Cette intégration définit ainsi le modèle de transport \mathcal{M} ou modèle direct du mouvement comme :

$$\begin{aligned} \mathcal{M}: \mathcal{V} &\rightarrow \mathcal{F}, \\ \mathbf{v} &\mapsto f = \mathcal{M}(\mathbf{v}) \end{aligned} \quad (\text{D.5})$$

pour le temps expérimental t .

\mathcal{V} est l'espace des champs du mouvement que nous assimilerons à l'espace de contrôle (voir section 5.1.2). $\mathbf{v} = (v_1, v_2) \in \mathcal{V}$ est un champ de mouvement donné, c'est l'équivalent de la variable de contrôle de la section (5.1.2). \mathcal{F} est l'espace des fonctions de concentration (luminance) admissibles, nous l'assimilerons à l'espace d'état. f est la concentration au temps expérimental t , c'est la variable d'état du système. Le domaine physique $\Omega \subset \mathbb{R}^2$ est une région rectangulaire du plan.

Ce modèle nous permet de définir l'image f au temps t à partir du champ de vitesse du mouvement et de l'image au temps zéro f^0 supposée pris en compte implicitement dans l'opérateur modèle \mathcal{M} . Dans la réalité, on dispose en général de l'image au temps t et non du champ de mouvement. On en déduit le champ de vitesse décrivant le mouvement par résolution d'un problème inverse. Etant donnée l'image au temps t , si nous définissons l'opérateur d'observation comme l'identité, l'estimation du mouvement entre l'image au temps zéro et l'image au temps t se ramène à un problème inverse décrit sous la forme incrémentale comme suit :

$$\text{trouver } \delta\mathbf{v}^* = \text{ArgMin}(J(\delta\mathbf{v})), \delta\mathbf{v} \in \mathcal{V}, \quad (\text{D.6})$$

avec $\mathbf{v} = \mathbf{v}^b + \delta\mathbf{v}$ et :

$$J(\delta\mathbf{v}) = \frac{1}{2}\|\mathcal{M}(\mathbf{v}^b + \delta\mathbf{v}) - \mathbf{y}^o\|_{\mathcal{O}}^2 + \alpha_b \frac{1}{2}\|\delta\mathbf{v}\|_{\mathcal{V}}^2. \quad (\text{D.7})$$

Dans les expériences ici, nous prenons une ébauche nulle ($\mathbf{v}^b = \mathbf{0}$). L'estimation du mouvement ainsi défini est affectée par le problème d'ouverture : seule la composante du mouvement suivant la normale aux iso-contours de l'image peut ainsi être déterminée. Le mouvement suivant les tangentes aux iso-contours et le mouvement dans les régions homogènes ne peut être déterminé. C'est un problème mal posé. Les travaux de Bertero [Bertero et al.(1988)] donnent plus d'informations sur les problèmes inverses mal posés en traitement d'images. Pour faire face à ce problème, on utilise des techniques de régularisation (voir la section 5.2). Pour la minimisation, nous utilisons l'algorithme M1QN3 de la bibliothèque LIBOPT [Gilbert and Jonsson(2007)]. Le résultat de la minimisation (état analysé) est noté \mathbf{v}^a .

D.3 Analyse d'incertitudes

Pour la comparaison des différentes méthodes de régularisation utilisées, nous définissons :

- l'erreur absolue sur le champ de vitesse estimé (erreur sur la variable de contrôle analysée) par l'équation :

$$e^{\mathbf{v}}(x, y) = \|\mathbf{v}^{\mathbf{a}}(x, y) - \mathbf{v}^{\mathbf{t}}(x, y)\|; \quad (\text{D.8})$$

- l'erreur absolue sur la vorticit   du champ estim   par :

$$e^{\mathbf{v}^{\mathbf{r}}}(x, y) = \|\nabla_x \times \mathbf{v}^{\mathbf{a}}(x, y) - \nabla_x \times \mathbf{v}^{\mathbf{t}}(x, y)\|. \quad (\text{D.9})$$

Etant donn   l'erreur absolue e , nous d  finissons l'erreur moyenne \bar{e} sur le domaine physique Ω par :

$$\bar{e} = \frac{\int_{\Omega} e(\mathbf{x}) d\mathbf{x}}{\int_{\Omega} 1 d\mathbf{x}}, \quad (\text{D.10})$$

et l'erreur moyenne normalis  e par :

$$\bar{\bar{e}} = \frac{\bar{e}}{\bar{e}_b}, \quad (\text{D.11})$$

o   \bar{e}_b repr  sente l'erreur sur l'  bauche. Nous d  finissons   galement la fonction co  t normalis  e par :

$$\bar{\bar{J}} = \frac{\bar{J}}{\bar{J}_b}, \quad (\text{D.12})$$

o   J_b est la fonction co  t calcul  e sur l'  bauche.

D.4 Donn  es exp  rimentales pour l'estimation du mouvement

Les images utilis  es pour les exp  riences d'estimation du mouvement dans ce document sont issues des exp  riences r  alis  es au Laboratoire des   coulements G  ophysiques et Industriels (LEGI) de Grenoble. L'  volution d'un vortex dans l'atmosph  re est simul  e gr  ce    la plateforme CORIOLIS (grande table tournante de 14 m  tre de diam  tre permettant d'  tudier les   coulements g  ophysiques). Le vortex est cr  e en enlevant brusquement un certain volume de fluide ou en agitant une portion de fluide circonscrit par un cylindre creux. Le vortex est rendu visible par ajout d'un traceur passif (fluorescine) et une cam  ra fix  e    l'ensemble permet de prendre des photographies de la vue de dessus. Plus de d  tails sur cette exp  rience peuvent   tre retrouv  s dans les travaux des exp  rimentateurs [Fl  r and Eames(2002)]. Les images d'origine sont de dimensions 1024×1024 pixels couvrant un domaine physique de $2.525m \times 2.525m$. Ce qui correspond    une r  solution spatiale d'environ $\sim 4 \text{ pixels/cm}$. Pour nos calculs, la r  solution a   t   d  grad  e    $\sim 0.5 \text{ pixels/cm}$ pour des images de 128×128 pixels sur le m  me domaine spatial. Apr  s avoir choisit une paire d'images $\{f_1^*, f_2^*\}$ de la s  quence d'origine (avec f_1^* l'image au temps $t = 0s$ et f_2^* l'image au temps $t = 0.5s$), la composante zonale $v_1 = v_1(x, y)$ et la composante m  ridionale $v_2 = v_2(x, y)$ du champ de vitesse de d  placement entre les deux images de la paire sont d  termin  s par assimilation directe de s  quences d'images [Titaud et al.(2009)]. Ce champ de vitesse de r  f  rence est consid  r   comme champ de vitesse vrai. Il est alors utilis   pour calculer une image f_1 gr  ce    un mod  le de

transport passif (D.4) à partir de la première image (f_1^*) de la paire d'images d'origine. Nous définissons ainsi une nouvelle séquence $\{f_0, f_1\}$ (avec $f_0 = f_1^*$) d'images que nous appellerons séquence CORIOLIS et qui sera utilisée pour les tests. La figure (D.1) montre deux images de la séquence d'origine et le champ de vitesse de référence calculé par assimilation directe de séquences d'images.

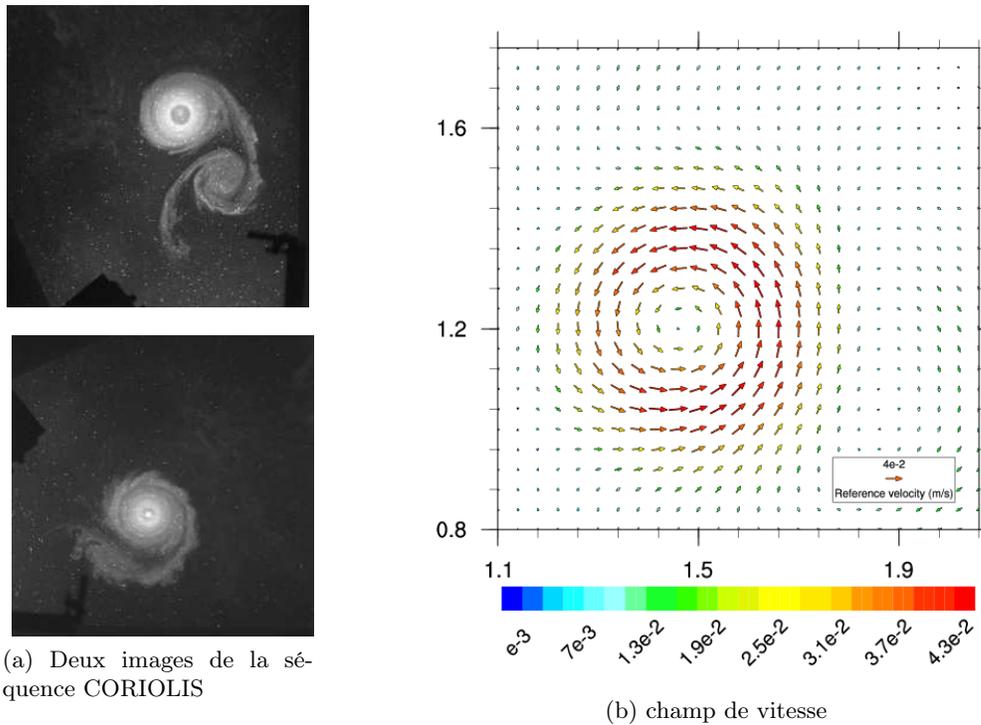


FIGURE D.1 – champ de vitesse vrai pour les expériences

Annexe E

Illustration du comportement des termes de régularisation classiques

E.1 Expérience sans régularisation

La figure (E.1) montre l'évolution au cours des itérations des fonctions de diagnostic normalisées pour une expérience sans régularisation.

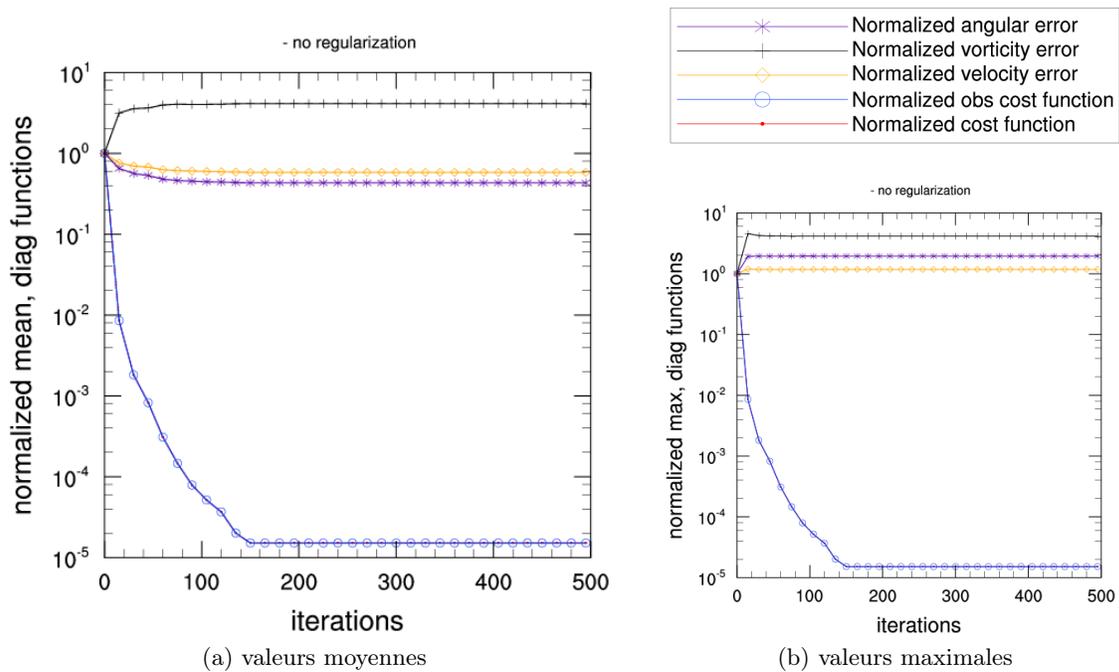


FIGURE E.1 – Estimation du mouvement sans régularisation : évolution des fonctions de diagnostic normalisées en fonction des itérations ; valeurs moyennes (a) et valeurs maximales (b).

E.2 Evolution des diagnostics en fonction du paramètre de pondération

Ici, nous étudions le comportement des fonctions de diagnostic à convergence en fonction du paramètre de pondération pour différents termes de régularisation. L'objectif est d'avoir une idée de la difficulté relative du choix du paramètre optimal de pondération. Les courbes sont tracées en coordonnées logarithmiques pour mieux lire les détails sur les faibles valeurs de la fonction coût. Les autres fonctions de diagnostic présentent moins de détails mais suffisamment pour voir l'évolution des courbes. La figure (E.2) montre l'évolution des fonctions de diagnostic en fonction du paramètre de pondération avec régularisation de Tikhonov. Sur le graphique sont représentées les valeurs atteintes par les fonctions de diagnostic quand la méthode a convergé. En abscisse nous avons le paramètre de pondération. En se basant sur l'évolution de la fonction coût et du terme d'écart aux observations, on peut diviser la courbe en deux parties :

- la première partie correspond aux faibles valeurs (jusqu'à $\sim 10^{-5}$) du paramètre de pondération où les courbes de la fonction coût et du terme d'écart aux observations sont presque identiques. Dans cette partie du graphique, l'effet de la régularisation est négligeable et la fonction coût est dominée par le terme d'écart aux observations d'où la superposition des deux courbes. Les autres fonctions de diagnostic sont quasi constantes et coïncident avec les valeurs à convergence obtenues sans régularisation, voir figure (E.1) ;
- la seconde partie correspond aux grandes valeurs du paramètre de pondération où la fonction coût passe sensiblement au dessus du terme d'écart aux observations. La fonction coût est progressivement dominée par le terme de régularisation. Les fonctions de diagnostic notamment l'erreur sur le champ estimé et l'erreur angulaire croissent et se rapprochent des diagnostics calculés sur l'ébauche.

A la jonction de ces deux parties du graphique, on a les valeurs pour lesquelles le terme d'écart aux observations et le terme de régularisation s'équilibrent. Ce comportement est en accord avec ce que l'on attend de la pénalisation.

Les figures (E.3, E.4, E.5, E.6, E.7, E.8, E.9) montrent respectivement l'évolution des fonctions de diagnostic (en fonction du paramètre de pondération) avec les régularisations respectives :

- homogène de Horn et Schunck par pénalisation du gradient du champ ;
- pénalisation du rotationnel du champ ;
- pénalisation de la divergence du champ ;
- pénalisation du gradient du rotationnel du champ ;
- pénalisation du gradient de la divergence du champ ;
- isotropique basée sur les données ;
- isotropique basée sur le champ.

De la même façon que dans le cas de la régularisation de Tikhonov, les graphiques représentent les valeurs atteintes par les fonctions de diagnostic quand la méthode a convergé. En abscisse nous avons le paramètre de pondération. Les courbes sont tracées avec une échelle logarithmique. Tous ces graphiques montrent des comportements presque identiques pour ce qui est de l'erreur commise sur le champ estimé et l'erreur angulaire.

En se basant sur l'évolution de la fonction coût et du terme d'écart aux observations, on peut diviser la courbe en deux parties comme dans le cas de la régularisation de Tikhonov. Mais

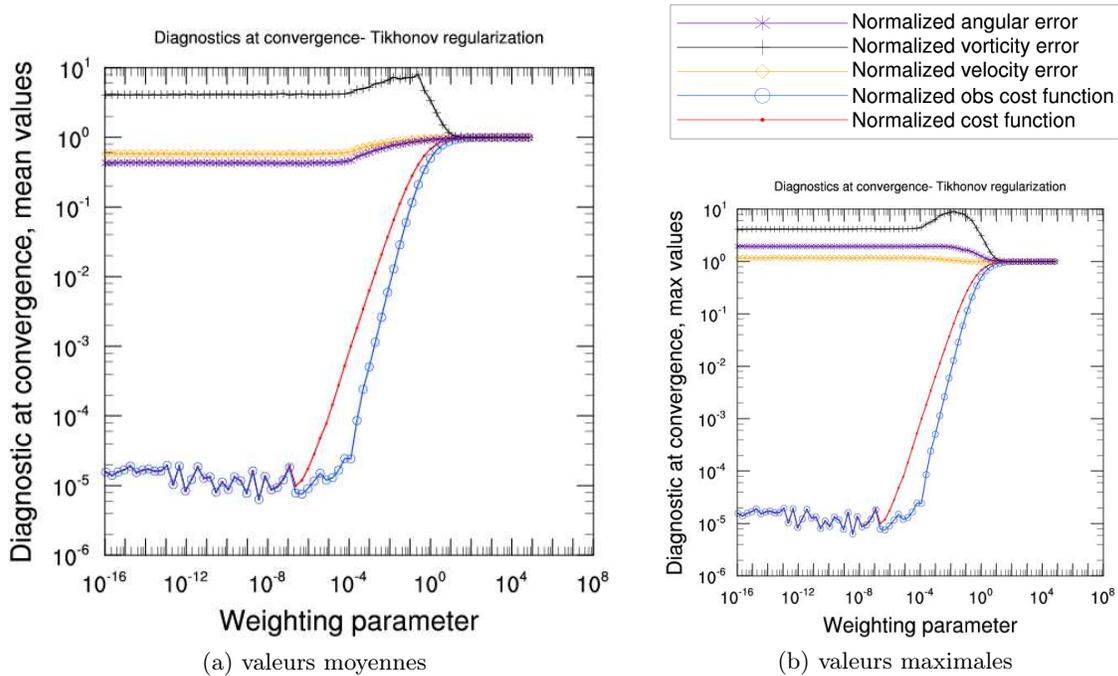


FIGURE E.2 – Estimation du mouvement avec régularisation de Tikhonov : évolution des fonctions de diagnostic à convergence en fonction du paramètre de pondération ; valeurs moyennes (a) et valeurs maximales (b).

on ne peut pas se limiter au comportement de la fonction coût. En effet les autres fonctions de diagnostic notamment l'erreur sur le champ estimé et l'erreur angulaire suivent une évolution très différente de celui du cas de la régularisation de Tikhonov. En prenant en compte le comportement de l'erreur sur le champ estimé, tous les graphiques peuvent alors se diviser en trois principales parties (les limites de chaque partie dépendant du type de régularisation) :

- la première partie correspond aux faibles valeurs du paramètre de pondération où on note un comportement similaire à celui observé avec la régularisation de Tikhonov. Les courbes de la fonction coût et du terme d'écart aux observations sont presque identiques. Dans cette partie du graphique, l'effet de la régularisation est négligeable et la fonction coût est dominée par le terme d'écart aux observations d'où la superposition des deux courbes. Les autres fonctions de diagnostic sont quasi constantes. La constante correspond dans chaque cas à la valeur obtenue sans régularisation lorsque l'algorithme de minimisation a convergé, (voir figure (E.1)) ;
- la deuxième partie avec les valeurs intermédiaires du paramètre de régularisation où l'erreur sur le champ de vitesse estimé passe en deçà des valeurs optimales obtenues sans régularisation. Dans cette région du graphique, la courbe de la fonction coût se détache sensiblement de la courbe du terme d'écart aux observations. La fonction coût est donnée par un équilibre entre le terme d'écart aux observations et le terme de régularisation. Les courbes des fonctions de diagnostic décrivent une concavité. Il est évident que le paramètre optimal se trouve au minimum de cette vallée. Cette dernière est plus ou moins mise en évidence selon la régularisation utilisée. Dans tous les cas, la concavité est moins marquée sur la courbe de l'erreur sur la vorticit . L'effet de la r gularisation

- y est moins accentué que sur l'erreur angulaire et l'erreur sur le champ estimé ;
- la troisième partie correspond aux grandes valeurs du paramètre de pondération du terme de régularisation. On note ici un comportement similaire à celui de la deuxième partie de la figure (E.2) de la régularisation de Tikhonov. Les fonctions de diagnostic notamment l'erreur moyenne sur le champ estimé et l'erreur angulaire moyenne croissent avec le paramètre de pondération et tendent vers une borne correspondant aux diagnostics sur l'ébauche. Pour ces valeurs du paramètre de pondération, la fonction coût est essentiellement dominée par le terme de régularisation qui limite la minimisation.

La deuxième partie des graphiques montre une différence significative par rapport à la régularisation de Tikhonov. C'est ce qui fait la différence par rapport à ce terme qui sert essentiellement à rendre le problème bien posé. En plus d'agir sur le caractère mal posé du problème, les autres types de régularisation améliorent de façon significative le résultat moyennant le choix d'un bon paramètre de pondération.

Les courbes moyennes montrent dans tous les cas une amélioration par rapport à la régularisation de Tikhonov avec un bon choix du paramètre de pondération. Dans le cas des valeurs maximales on note plutôt une détérioration ou une amélioration peu significative sauf dans le cas de la régularisation homogène de Horn et Schunck par pénalisation du gradient où l'effet est bien visible sur la courbe d'erreur sur le champ estimé.

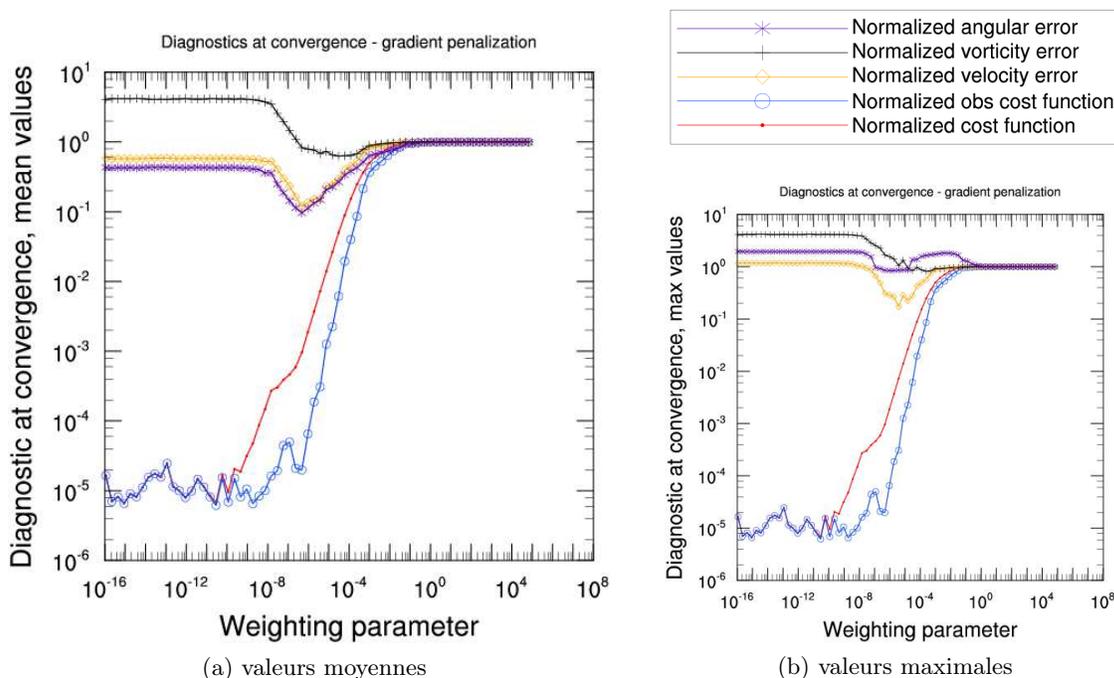


FIGURE E.3 – Estimation du mouvement avec pénalisation du gradient : évolution des fonctions de diagnostic à convergence en fonction du paramètre de pondération ; valeurs moyennes (a) et valeurs maximales (b).

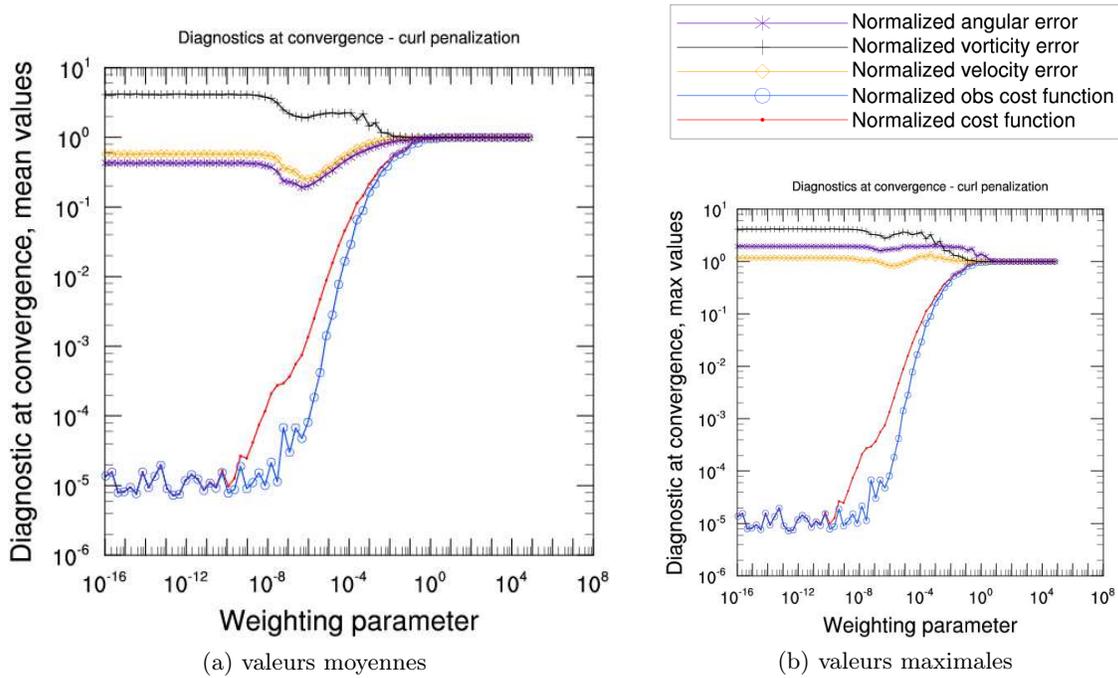


FIGURE E.4 – Estimation du mouvement avec pénalisation du rotationnel : évolution des fonctions de diagnostic à convergence en fonction du paramètre de pondération ; valeurs moyennes (a) et valeurs maximales (b).

E.3 Evolution des fonctions de diagnostic au cours de la minimisation

Les figures (E.10, E.12, E.13, E.14, E.15, E.16, E.17) montrent l'évolution des fonctions de diagnostics au cours des itérations avec les régularisations respectives :

- homogène de Horn et Schunck par pénalisation du gradient du champ ;
- pénalisation du rotationnel du champ ;
- pénalisation de la divergence du champ ;
- pénalisation du gradient du rotationnel du champ ;
- pénalisation du gradient de la divergence du champ ;
- isotropique basée sur les données ;
- isotropique basée sur le champ.

chaque terme de régularisation étant associé au paramètre optimal. Ce meilleur paramètre est déterminé à partir de l'étude précédente. Les courbes sont tracées avec une échelle logarithmique en ordonnée. Dans tous les cas, on peut noter des oscillations sur la courbe d'évolution du terme d'écart aux observations de la fonction coût. Pour expliquer ces oscillations qui ne sont pas observées sur la fonction coût globale, il faut se rappeler que la fonction coût globale est constituée de plusieurs termes dont un terme d'écart aux observations et un ou plusieurs termes de régularisation. L'algorithme itératif de minimisation calcule une suite d'approximations convergeant vers la solution minimisant la fonction coût. A chaque itération, la nouvelle approximation est la somme d'une composante améliorant le terme d'écart aux observations et d'une composante améliorant le(s) terme(s) de régularisation. Les deux composantes pouvant

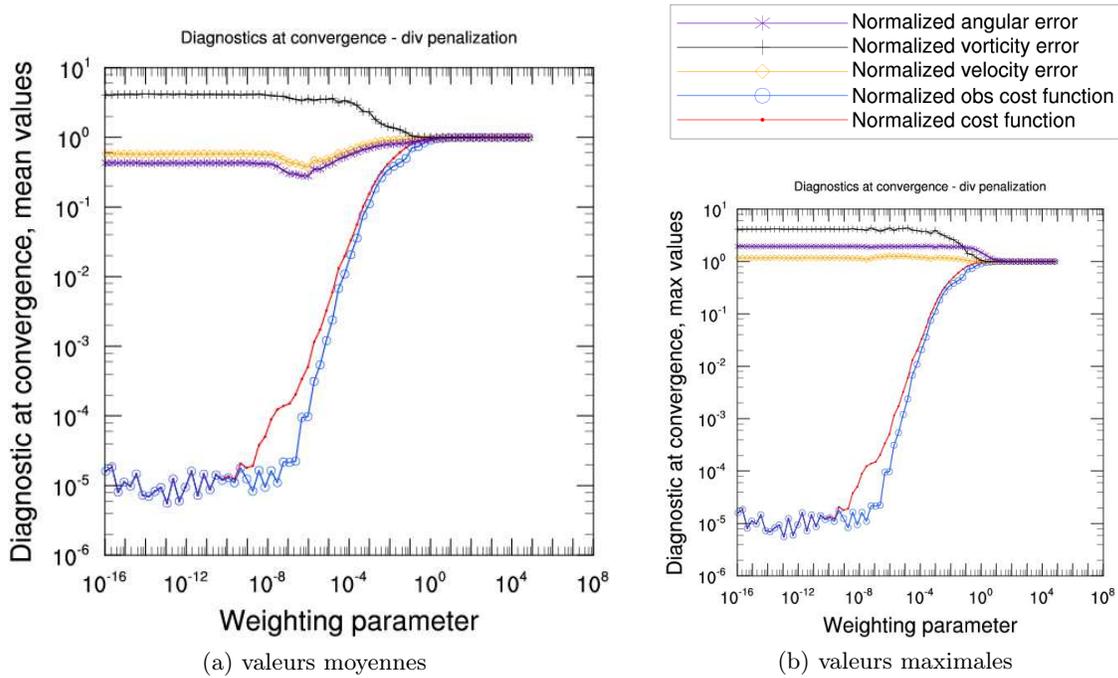


FIGURE E.5 – Estimation du mouvement avec pénalisation de la divergence : évolution des fonctions de diagnostic à convergence en fonction du paramètre de pondération; valeurs moyennes (a) et valeurs maximales (b).

être contradictoires. Ainsi la détérioration apportée par la composante de régularisation sur le terme d'écart aux observations peut dominer sur l'amélioration apportée par le terme associé. Ce qui crée des oscillations sur le terme d'écart aux observations. Un comportement similaire est observé avec les autres termes de régularisation.

Les figures (E.18, E.19, E.20, E.21) montrent l'évolution comparée du terme d'écart aux observations, de l'erreur sur le champ estimé, de l'erreur angulaire et de l'erreur sur la vorticité avec les contraintes suivantes :

- aucune régularisation ;
- régularisation homogène de Horn et Schunk par pénalisation du gradient ;
- régularisation second ordre de Suter (gradient de divergence et du rotationnel) ;
- régularisation isotropique basée sur les données ;
- régularisation isotropique basée sur le champ.

Le terme de régularisation étant pondéré dans chaque cas par le paramètre optimal. Dans le cas de la régularisation de Suter, les paramètres optimaux obtenus avec l'étude sur le terme de pénalisation du gradient du rotationnel et le terme de pénalisation du gradient de divergence.

Sur la figure (E.18), on peut constater que l'évolution du terme d'écart aux observations est comparable pour tous les termes de régularisation. On note des oscillations sauf dans le cas de l'expérience sans régularisation ; ce qui témoigne des effets contradictoires du terme d'écart aux observations et du terme de régularisation. La décroissance avec la régularisation isotropique basée sur les données et la régularisation isotropique basée sur le champ s'arrête très vite par rapport aux autres cas. Le terme de régularisation dans ces cas domine la fonction coût et la minimisation est arrêtée comme le confirme l'évolution des autres fonctions de

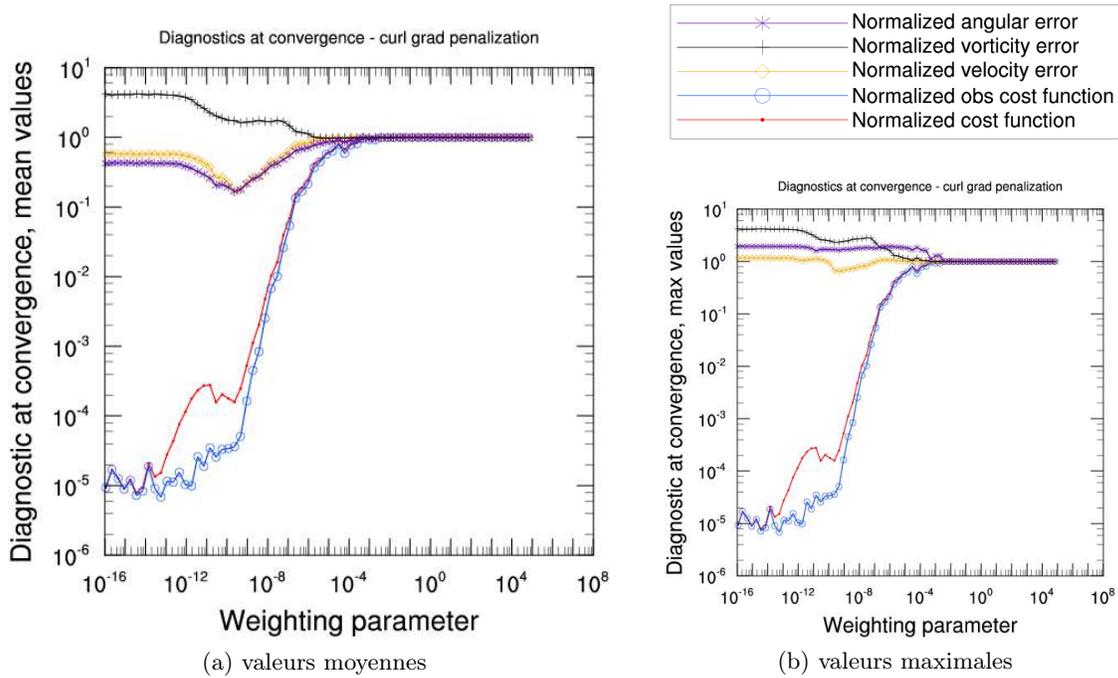


FIGURE E.6 – Estimation du mouvement avec pénalisation du gradient du rotationnel : évolution des fonctions de diagnostic à convergence en fonction du paramètre de pondération ; valeurs moyennes (a) et valeurs maximales (b).

diagnostic (Figures E.19, E.20 et E.21). Dans les autres cas (pénalisation du gradient et second ordre de Suter) le terme de l'écart aux observations atteint la valeur minimale obtenue sans régularisation. L'impression de l'arrêt très brutale de la minimisation est due à l'échelle logarithmique utilisée en ordonnée.

Sur le graphique de gauche (moyennes normalisées) de la figure (E.19), la progression de l'erreur sur le champ estimé est comparable dans les différents cas. A convergence, on a une erreur sur le champ estimé de l'ordre de :

- 60% sans régularisation ;
- 30% avec la régularisation isotropique basée sur les données et la régularisation isotropique basée sur le champ ;
- 10% avec la régularisation par pénalisation du gradient et la régularisation au second ordre de Suter.

En terme de valeurs maximales, sur le graphique de droite de la figure (E.19), on note à la convergence une erreur sur le champ estimé de l'ordre de :

- 120% sans régularisation ;
- 90% avec la régularisation isotropique basée sur les données et la régularisation isotropique basée sur le champ ;
- 30% avec la régularisation par pénalisation du gradient et la régularisation au second ordre de Suter.

Toutes ces valeurs sont récapitulées dans la table (5.2).

En terme d'erreur angulaire (figure E.20), l'évolution des courbes est similaire à ceux du cas de l'erreur sur le champ estimé avec des valeurs moyennes relatives légèrement plus faibles

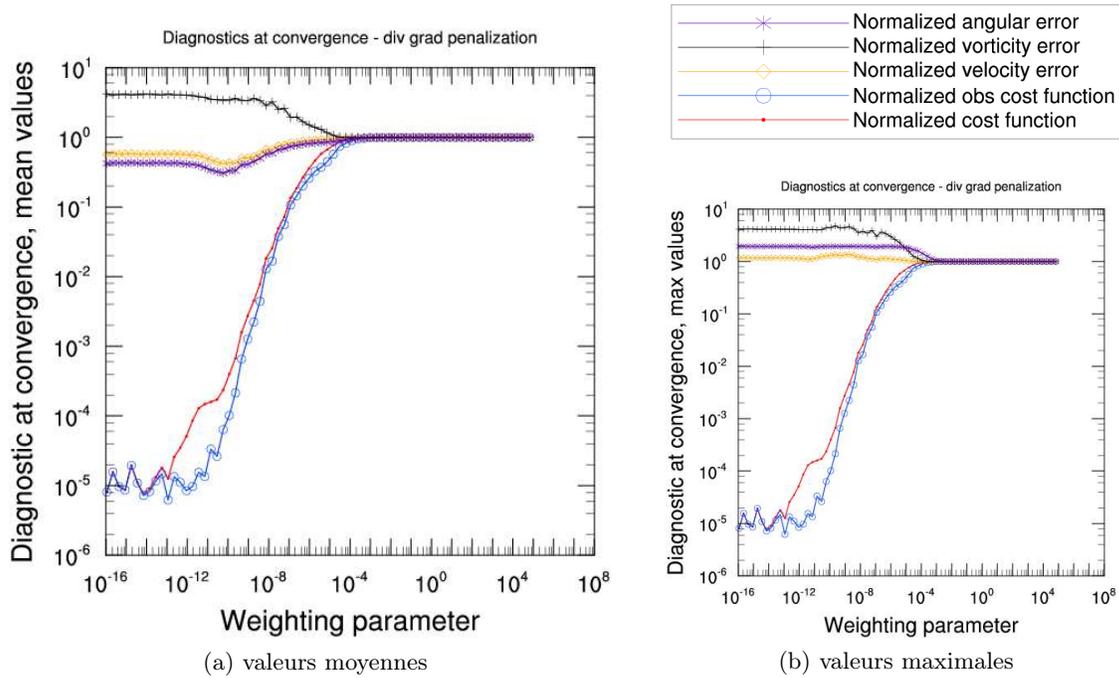


FIGURE E.7 – Estimation du mouvement avec pénalisation du gradient de la divergence : évolution des fonctions de diagnostic à convergence en fonction du paramètre de pondération ; valeurs moyennes (a) et valeurs maximales (b).

à convergence. On note à convergence une erreur moyenne de l'ordre de :

- 40% sans régularisation ;
- 20% avec la régularisation isotropique basée sur le champ ;
- 10% avec la régularisation par pénalisation du gradient, la régularisation au second ordre de Suter et la régularisation isotropique basée sur les données.

En termes d'erreur maximale cependant, il n'y a pas de comparaison possible avec l'erreur sur le champ estimé. A convergence, on note une erreur maximale de l'ordre de 90% avec la régularisation par pénalisation du gradient et de l'ordre de 150 à 200% dans les autres cas. Toutes ces valeurs sont résumées dans la table (5.2).

En terme de vorticit  (figure E.21), l'am lioration est discutable m me sur les valeurs moyennes. Au d but de la minimisation, l'erreur moyenne sur la vorticit  croit dans tous les cas avant de d croitre dans certains cas. A convergence, on note une erreur moyenne (graphique de gauche de la figure E.21) de l'ordre de :

- 400% sans r gularisation ;
- 200% avec la r gularisation isotropique bas e sur le champ et la r gularisation isotropique bas e sur les donn es ;
- 80% avec la r gularisation par p nalisation du gradient et la r gularisation au second ordre de Suter.

En valeurs maximales (graphique de droite de la figure E.21), seule la r gularisation au second ordre de Suter converge avec une erreur l g rement inf rieure   100%. Toutes ces valeurs sont r sum es dans la table (5.2).

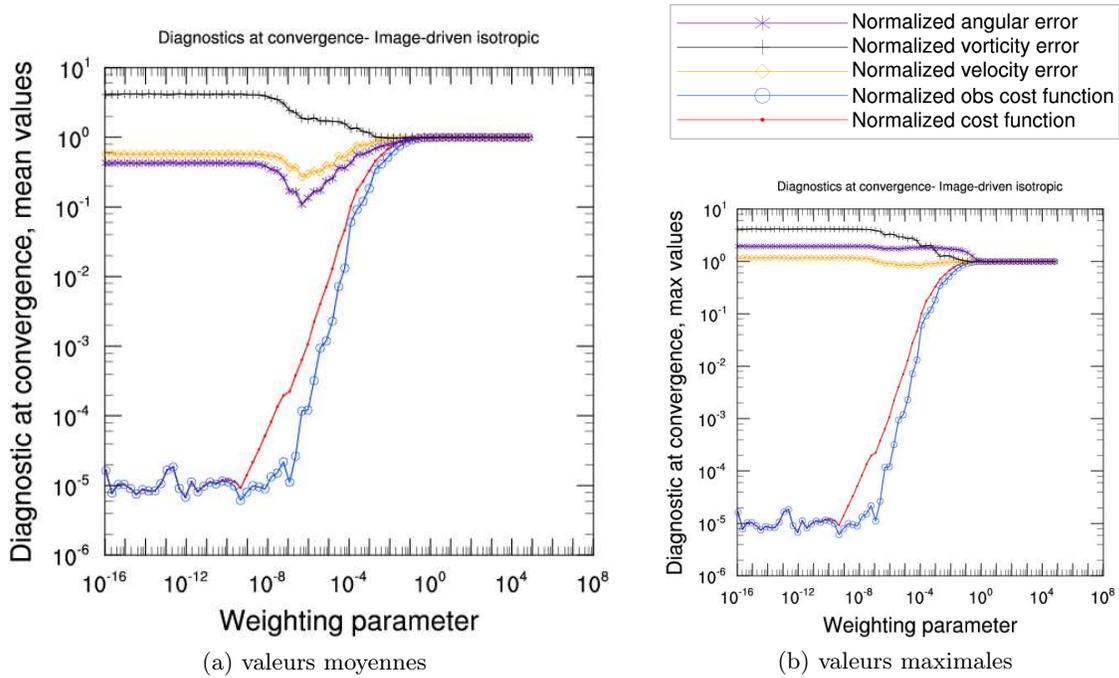


FIGURE E.8 – Estimation du mouvement avec régularisation isotropique basée sur les données : évolution des fonctions de diagnostic à convergence en fonction du paramètre de pondération ; valeurs moyennes (a) et valeurs maximales (b).

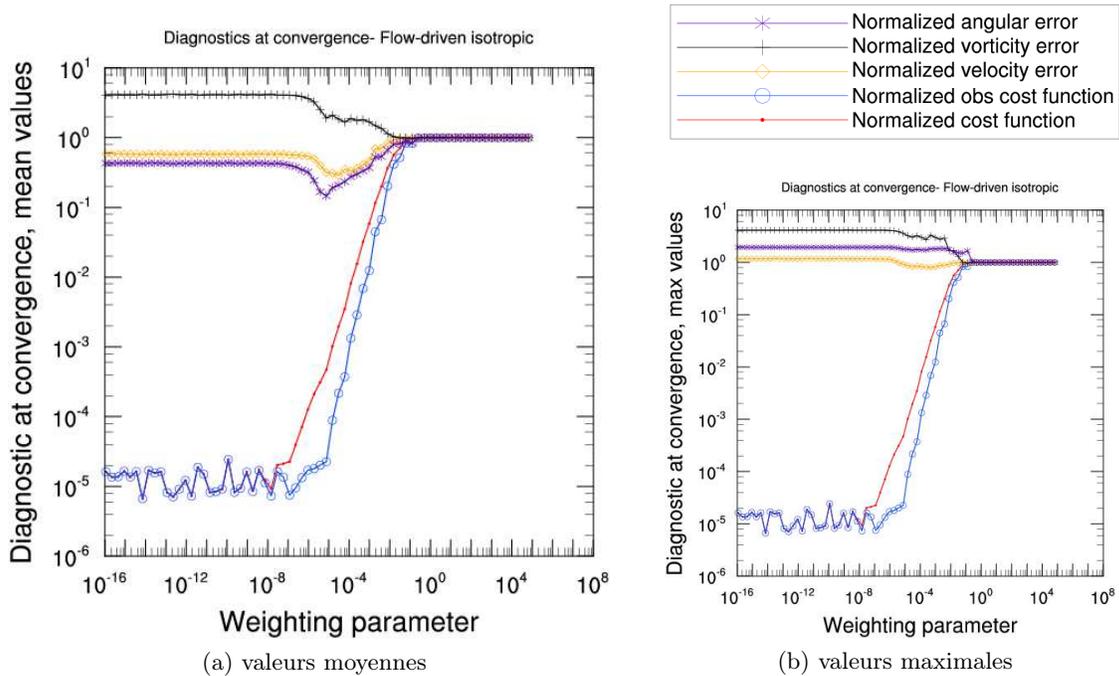


FIGURE E.9 – Estimation du mouvement avec régularisation isotropique basée sur le champ : évolution des fonctions de diagnostic à convergence en fonction du paramètre de pondération ; valeurs moyennes (a) et valeurs maximales (b).

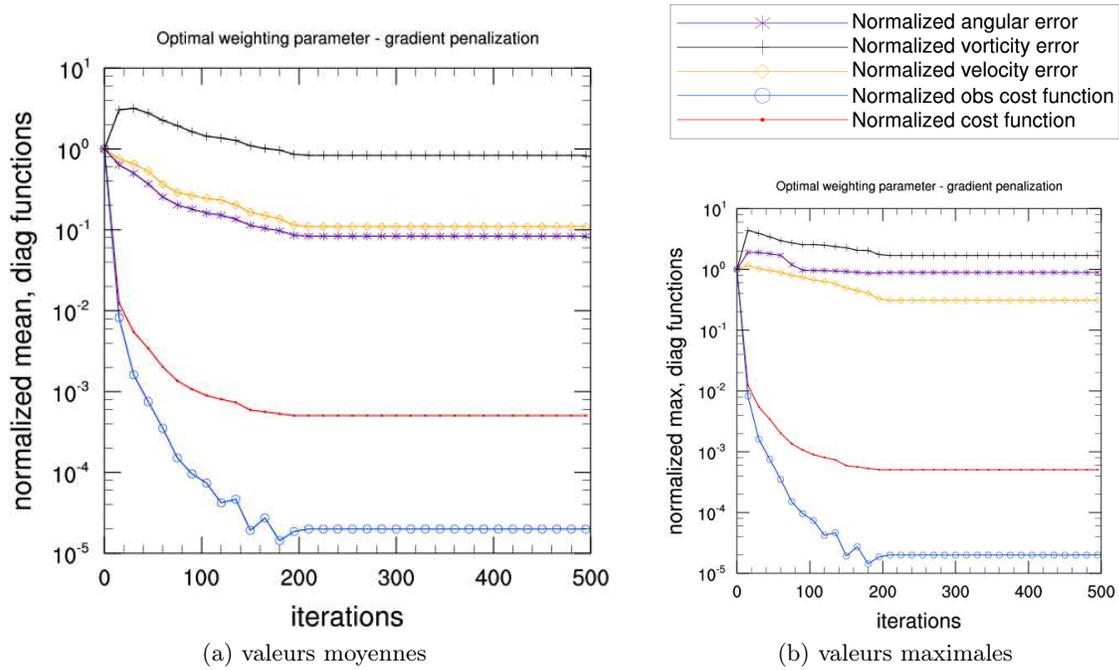


FIGURE E.10 – Estimation du mouvement, régularisation par pénalisation du gradient avec le meilleur paramètre de pondération : évolution des fonctions de diagnostic normalisées en fonction des itérations ; valeurs moyennes (a) et valeurs maximales (b).

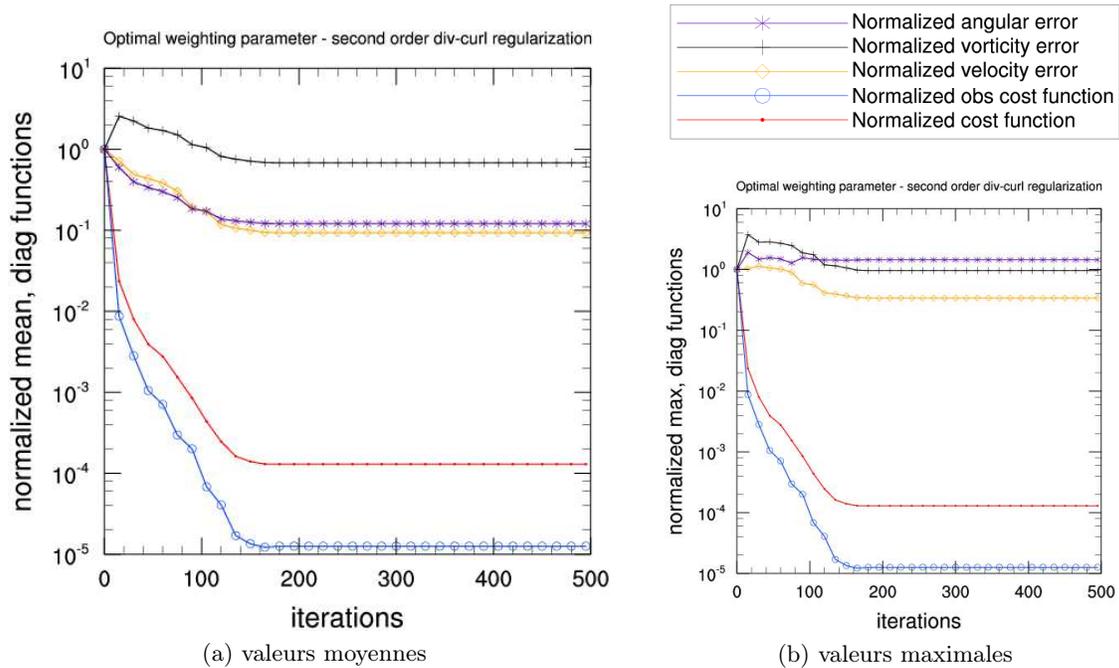


FIGURE E.11 – Estimation du mouvement, régularisation div-curl du second ordre de Suter avec les meilleurs paramètres de pondération : évolution des fonctions de diagnostic normalisées en fonction des itérations ; valeurs moyennes (a) et valeurs maximales (b).

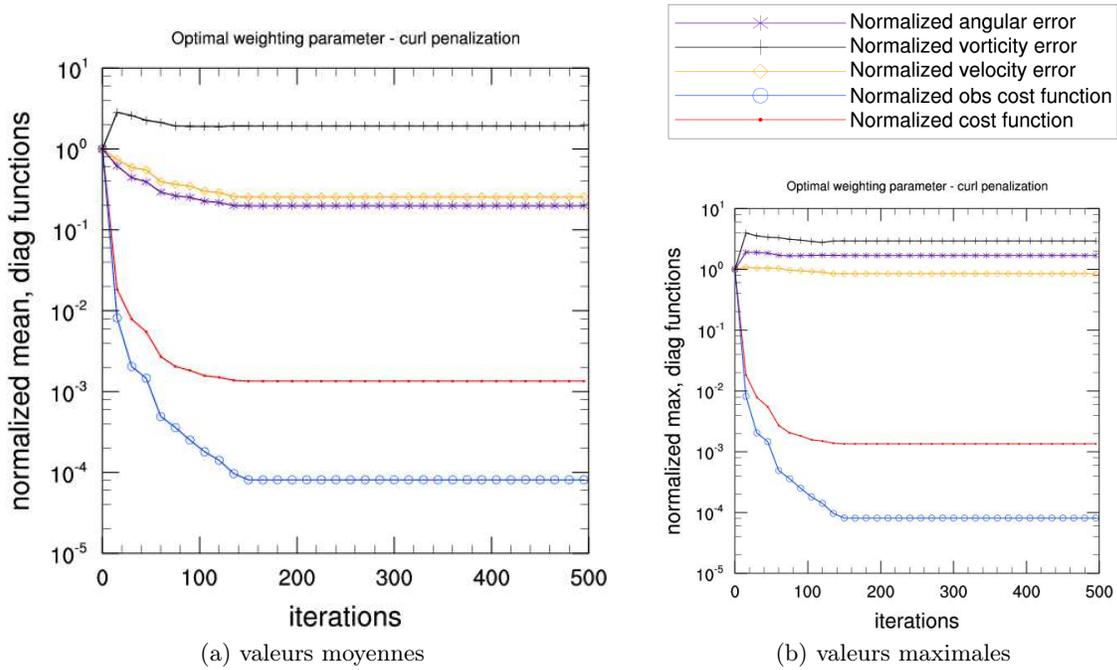


FIGURE E.12 – Estimation du mouvement, régularisation par pénalisation du rotationnel avec le meilleur paramètre de pondération : évolution des fonctions de diagnostic normalisées en fonction des itérations ; valeurs moyennes (a) et valeurs maximales (b).

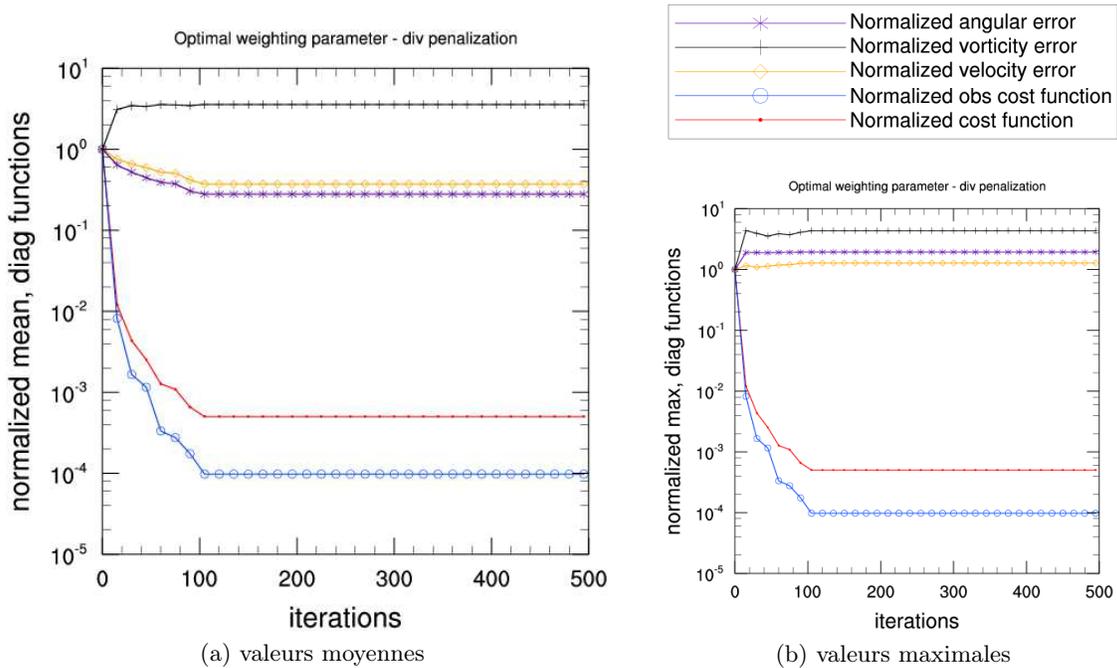


FIGURE E.13 – Estimation du mouvement, régularisation par pénalisation de la divergence avec le meilleur paramètre de pondération : évolution des fonctions de diagnostic normalisées en fonction des itérations ; valeurs moyennes (a) et valeurs maximales (b).

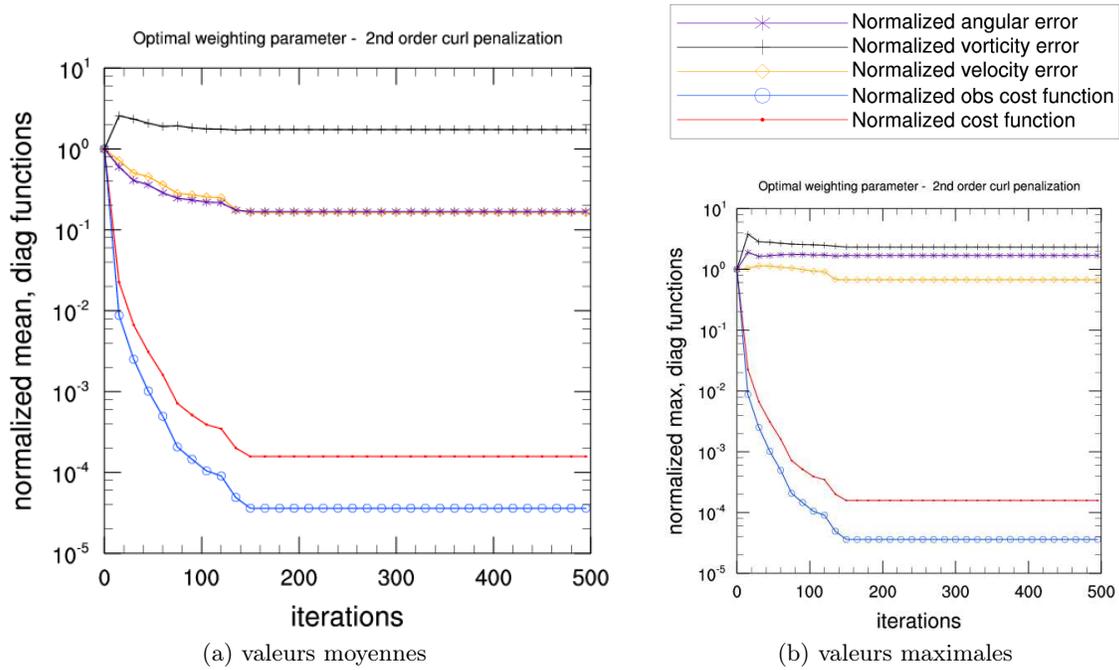


FIGURE E.14 – Estimation du mouvement, régularisation par pénalisation du gradient du rotationnel avec le meilleur paramètre de pondération : évolution des fonctions de diagnostic normalisées en fonction des itérations ; valeurs moyennes (a) et valeurs maximales (b).

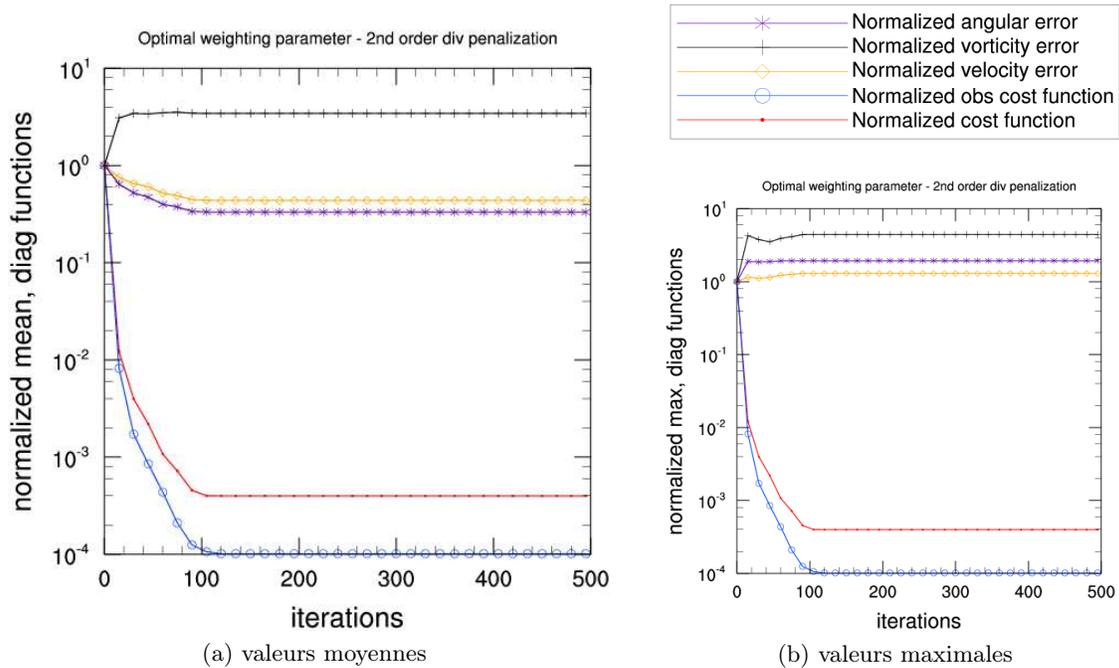


FIGURE E.15 – Estimation du mouvement, régularisation par pénalisation du gradient de la divergence avec le meilleur paramètre de pondération : évolution des fonctions de diagnostic normalisées en fonction des itérations ; valeurs moyennes (a) et valeurs maximales (b).

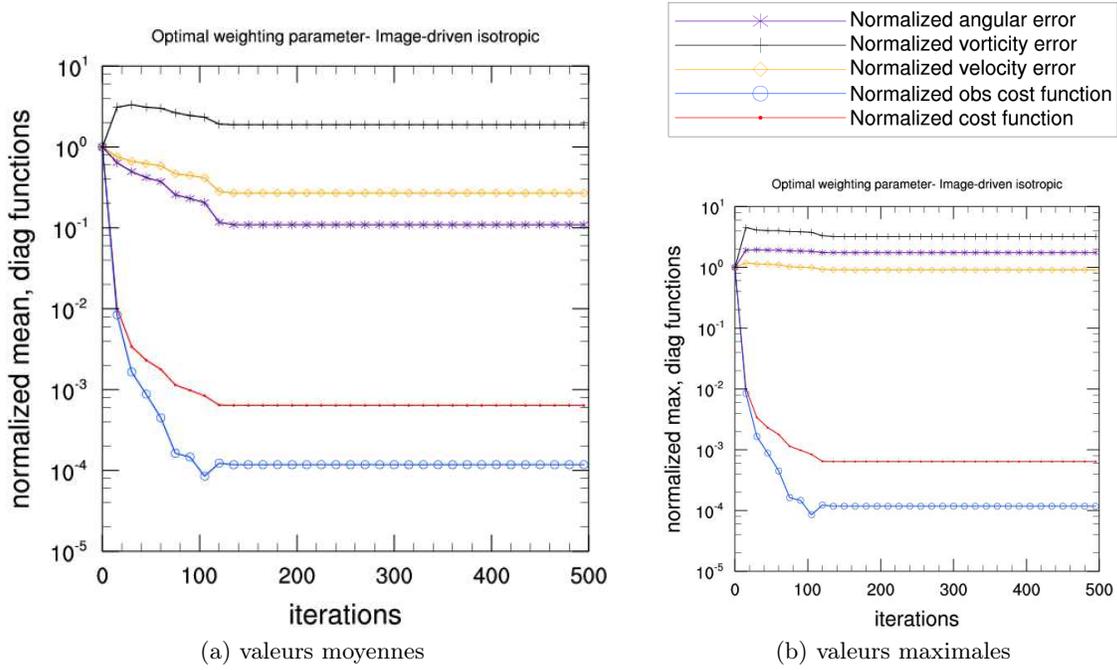


FIGURE E.16 – Estimation du mouvement, régularisation isotropique basée sur les données avec le meilleur paramètre de pondération : évolution des fonctions de diagnostic normalisées en fonction des itérations ; valeurs moyennes (a) et valeurs maximales (b).

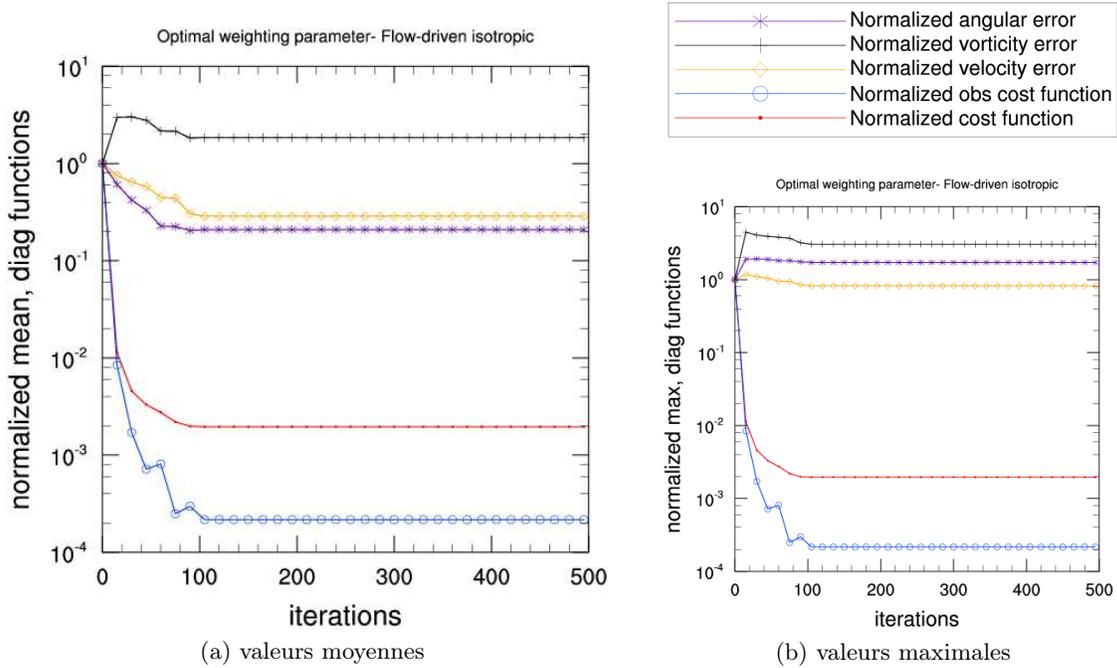


FIGURE E.17 – Estimation du mouvement, régularisation isotropique basée sur le champ avec le meilleur paramètre de pondération : évolution des fonctions de diagnostic normalisées en fonction des itérations ; valeurs moyennes (a) et valeurs maximales (b).

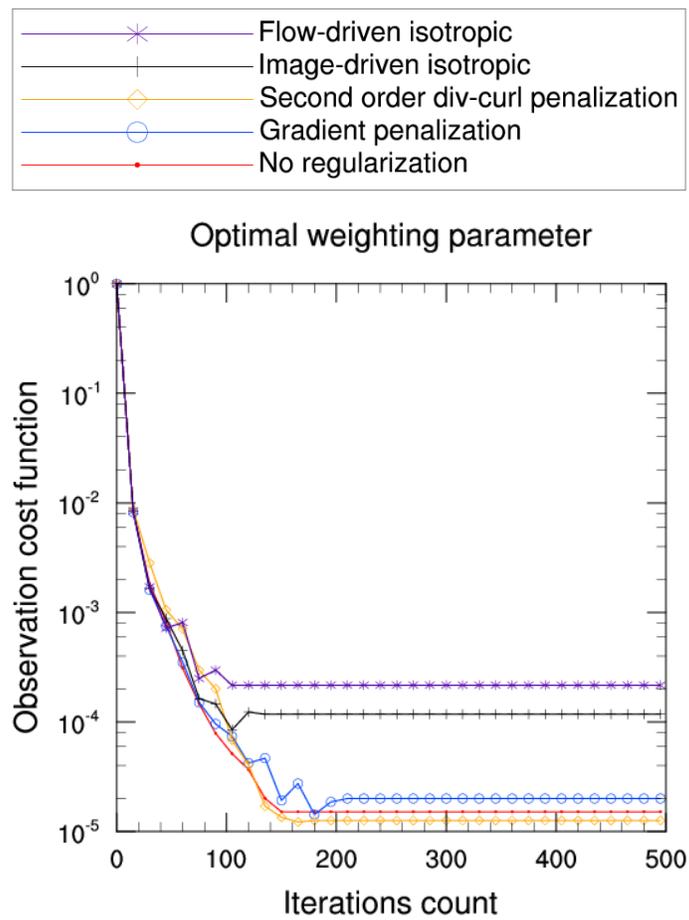


FIGURE E.18 – Estimation du mouvement avec différents terme de régularisation, comparaison de l'évolution du terme d'écart aux observations au cours de la minimisation

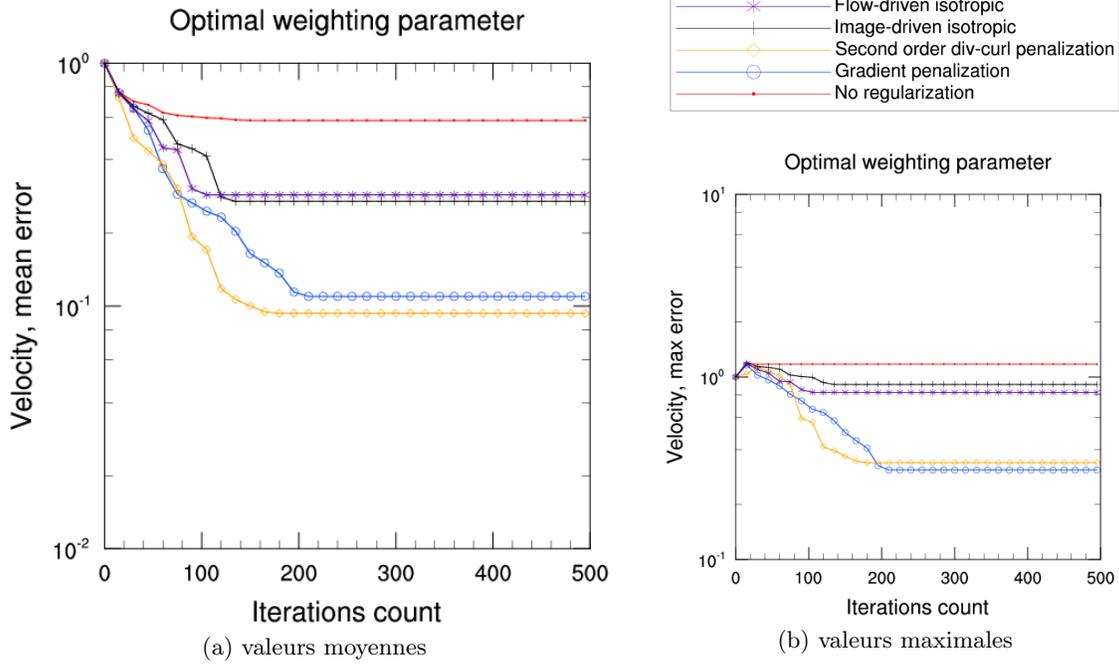


FIGURE E.19 – Estimation du mouvement avec différents termes de régularisation, comparaison de l'évolution de l'erreur moyenne normalisée sur le champ estimé au cours de la minimisation ; valeurs moyennes (a) et valeurs maximales (b)

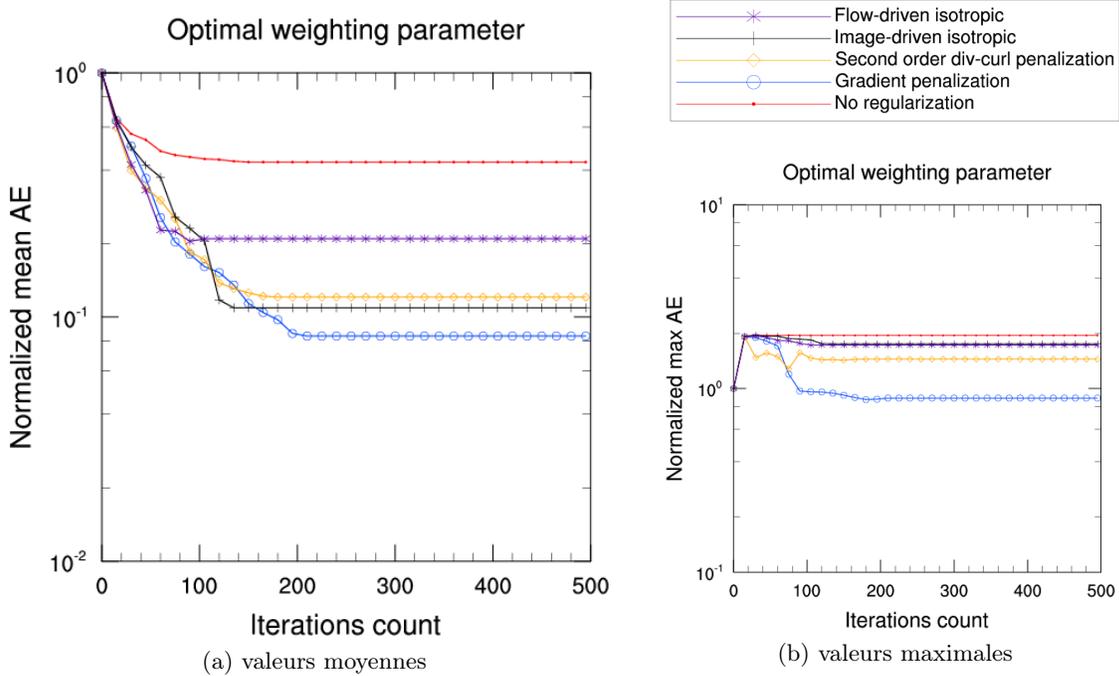


FIGURE E.20 – Estimation du mouvement avec différents termes de régularisation, comparaison de l'évolution de l'erreur angulaire moyenne normalisée au cours de la minimisation ; valeurs moyennes (a) et valeurs maximales (b)

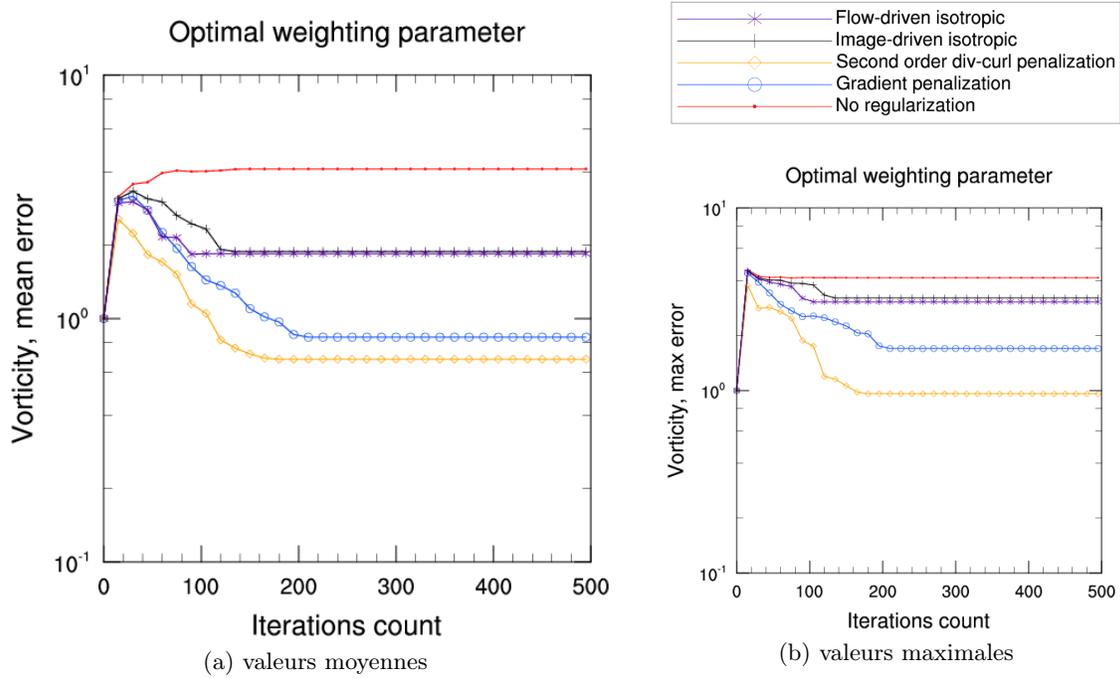


FIGURE E.21 – Estimation du mouvement avec différents termes de régularisation, comparaison de l'évolution de l'erreur moyenne normalisée sur la vortacité au cours de la minimisation ; valeurs moyennes (a) et valeurs maximales (b)

Bibliographie

- [Adrian(1991)] Ronald J. Adrian. Particle imaging techniques for experimental fluid mechanics. *Annual Review of Fluid Mechanics*, 23 :261–04 (1991). 57
- [Alvarez et al.(1999)] Luis Alvarez, Julio Esclarin, Martin Lefbure, and Javier Snchez. A pde model for computing the optical flow. In *XVI congreso de ecuaciones diferenciales y aplicaciones (C.E.D.Y.A. XVI)*, pages 1349–1356. Las Palmas de Gran Canaria (1999). 114, 116, 117, 139
- [Ames(1992)] William F. Ames. *Numerical methods for partial differential equations*. Academic Press (1992). 3rd edition. 142
- [Aminou et al.(1997)] Donny M. A. Aminou, Bernard Jacquet, and Frederick Pasternak. Characteristics of the meteosat second generation (msg) radiometer/imager : Severi. In *SPIE*, volume 3221, pages 19–31 (1997). doi :10.1117/12.298084. 52
- [Amodei and Benbourhim(1991)] L. Amodei and M. N. Benbourhim. A vector spline approximation. *Journal of Approximation Theory*, 67(1) :51–79 (1991). ISSN 0021-9045. doi :10.1016/0021-9045(91)90025-6. 116
- [Anandan(1989)] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Journal of Computer Vision*, 2(3) :283–310 (1989). doi :10.1007/BF00158167. 57
- [Andrews et al.(1987)] David G. Andrews, Conway B. Leovy, and James R. Holton. *Middle atmosphere dynamics*. Academic Press (1987). 47
- [Anh et al.(1996)] Vo Anh, Ji Yu Shi, and Hung Tat Tsui. Scaling theorems for zero crossings of bandlimited signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(3) :309–320 (1996). ISSN 0162-8828. doi :http://doi.ieeecomputersociety.org/10.1109/34.485558. 56
- [Antoniadis(2007)] Anestis Antoniadis. Wavelet methods in statistics : some recent developments and their applications. *Statistics Surveys*, 1 :16–55 (2007). 82
- [Auroux(2003)] Didier Auroux. *Étude de différentes méthodes d’assimilation de données pour l’environnement*. Ph.D. thesis, Université de Nice Sophia Antipolis (2003). 173
- [Ballabrera-Poy et al.(2001)] Joaquim Ballabrera-Poy, Pierre Brasseur, and Jacques Verron. Dynamical evolution of the error statistics with the seek filter to assimilate altimetric data in eddy-resolving ocean models. *Quarterly Journal of the Royal Meteorological Society*, 127(571) :233–253 (2001). DOI : 10.1002/qj.49712757113. 168
- [Barron et al.(1994)] John L. Barron, David J. Fleet, and Steven Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1) :43–77 (1994). 89, 118, 119, 133, 176

- [Bazi et al.(2007)] Yakoub Bazi, Lorenzo Bruzzone, and Farid Melgani. Image thresholding based on the em algorithm and the generalized gaussian distribution. *Pattern Recogn.*, 40(2) :619–634 (2007). ISSN 0031-3203. doi :<http://dx.doi.org/10.1016/j.patcog.2006.05.006>. 55
- [Bedka and Mecikalski(2005)] Kristopher M. Bedka and John R. Mecikalski. Application of satellite-derived atmospheric motion vectors for estimating mesoscale flows. *Journal of Applied Meteorology*, 44(11) :1761–1772 (2005). doi :10.1175/JAM2264.1. 59
- [Bedka et al.(2009)] Kristopher M. Bedka, Christopher S. Velden, Ralph A. Petersen, Wayne F. Feltz, and John R. Mecikalski. Comparisons of satellite-derived atmospheric motion vectors, rawinsondes, and noaa wind profiler observations. *Journal of Applied Meteorology and Climatology*, 48(8) :1542–1561 (2009). doi :10.1175/2009JAMC1867.1. 58
- [Béréziat et al.(2000)] Dominique Béréziat, Isabelle Herlin, and Laurent Younes. A generalized optical flow constraint and its physical interpretation. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2 :2487 (2000). ISSN 1063-6919. doi :<http://doi.ieeecomputersociety.org/10.1109/CVPR.2000.854890>. 183
- [Bertero et al.(1988)] Mario Bertero, Tomaso A. Poggio, and Vincent Torre. Ill-posed problems in early vision. In *Proceedings of the IEEE*, volume 76, pages 869–889 (1988). 184
- [Bischof et al.(1992)] Christian Bischof, Alan Carle, George Corliss, Andreas Griewank, and Paul Hovland. Adifor-generating derivative codes from fortran programs. *Sci. Program.*, 1(1) :11–29 (1992). ISSN 1058-9244. 40
- [Bischof et al.(1996)] Christian Bischof, Peyvand Khademi, Andrew Mauer, and Alan Carle. Adifor 2.0 : Automatic differentiation of fortran 77 programs. *IEEE Comput. Sci. Eng.*, 3(3) :18–32 (1996). ISSN 1070-9924. doi :<http://dx.doi.org/10.1109/99.537089>. 40
- [Bischof et al.(2003)] Christian Bischof, Bruno Lang, and Andre Vehreschild. Automatic differentiation for matlab programs. *Softw. Pract. Exper.*, 2(1) :1617–7061 (2003). ISSN 0038-0644. doi :10.1002/pamm.200310013. 40
- [Bischof et al.(1997)] Christian Bischof, Lucas Roh, and A. J. Mauer-Oats. Adic : an extensible automatic differentiation tool for ansi-c. *Softw. Pract. Exper.*, 27(12) :1427–1456 (1997). ISSN 0038-0644. doi :[http://dx.doi.org/10.1002/\(SICI\)1097-024X\(199712\)27:12<1427::AID-SPE138>3.3.CO;2-H](http://dx.doi.org/10.1002/(SICI)1097-024X(199712)27:12<1427::AID-SPE138>3.3.CO;2-H). 40
- [Black and Anandan(1991)] Michael J. Black and P. Anandan. Robust dynamic motion estimation over time. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 296–302. IEEE Computer Society Press (1991). 114
- [Bormann et al.(2003a)] Niels Bormann, Sami Saarinen, Graeme Kelly, and Jean-Noël Thépaut. The spatial structure of observation errors in atmospheric motion vectors from geostationary satellite data. *Monthly Weather Review*, 131(4) :706–718 (2003a). doi :10.1175/1520-0493(2003)131<0706:TSSOOE>2.0.CO;2. 59, 71
- [Bormann et al.(2003b)] Niels Bormann, Jean-Noël Thépaut, and Graeme Kelly. Eclipse anomalies in atmospheric motion vectors derived from meteosat imagery and their implications on data assimilation. *Journal of Applied Meteorology*, 42(7) :1035–1042 (2003b). doi :10.1175/1520-0450(2003)042<1035:EAIAMV>2.0.CO;2. 59, 71

- [Brink(1994)] Anton Brink. Maximum entropy segmentation based on the autocorrelation function of the image histogram. *J. Comput. Inf. Technol.*, 2(2) :77–85 (1994). ISSN 1330-1136. 55
- [Bühler and Holmlund(1993)] Y. Bühler and K. Holmlund. The cmw extraction algorithm for mtp/mpef. In *2nd International Winds Workshop*. Tokyo, Japan (1993). 58
- [Byrd et al.(1987)] Richard H. Byrd, Robert B. Schnabel, and Gerald A. Shultz. A trust region algorithm for nonlinearly constrained optimization. *SIAM Journal on Numerical Analysis*, 24(5) :1152–1170 (1987). doi :10.1137/0724076. 30, 170
- [Candes et al.(2006)] Emmanuel Candes, Laurent Demanet, David Donoho, and Lexing Ying. Fast discrete curvelet transforms. *Multiscale Modeling and Simulation*, 5(3) :861–899 (2006). doi :10.1137/05064182X. 68, 81
- [Candes and Donoho(2004)] Emmanuel Candes and David Donoho. New tight frames of curvelets and optimal representations of objects with piecewise-c2 singularities. *Comm. on Pure and Appl. Math*, 57 :219–266 (2004). 81
- [Candes and Donoho(2005)] Emmanuel Candes and David Donoho. Continuous curvelet transform i. resolution of the wavefront set. *Applied and Computational Harmonic Analysis*, 19(3) :162–197 (2005). 67, 81
- [Canny(1986)] J. Canny. A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 8 :679–714 (1986). 56
- [Caselles(1995)] Vincent Caselles. Geometric models for active contours. In *ICIP95*, pages III : 9–12 (1995). 56
- [Caselles et al.(1993)] Vincent Caselles, Francine Catte, Tomeu Coil, and Françoise Dibos. A geometric model for active contours in image processing. *NumMath*, 66 :1–31 (1993). 56
- [Celis et al.(1985)] Maria Rosa Celis, John E. Dennis, and R.A. Tapia. *Numerical Optimization*, chapter A trust-region strategy for nonlinear equality constrained optimization, pages 71–82. SIAM : Philadelphia (1985). 30, 170
- [Chan et al.(1998)] F.H.Y. Chan, F.K. Lam, and Hui Zhu. Adaptive thresholding by variational method. *IEEE Transactions on Image Processing*, 7(3) (1998). 55
- [Chan and Shen(2005)] Tony F. Chan and Jianhong Shen. *Image processing and analysis*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2005). ISBN 0-89871-589-X. 68
- [Chen and Suter(1998)] Fang Chen and David Suter. Image coordinate transformation based on div-curl vector splines. In *ICPR '98 : Proceedings of the 14th International Conference on Pattern Recognition-Volume 1*, page 518. IEEE Computer Society, Washington, DC, USA (1998). ISBN 0-8186-8512-3. 115
- [Cohen(1992)] Albert Cohen. *Ondelettes et traitement numérique du signal*. MASSON (1992). 66
- [Cohen(1991)] Laurent D. Cohen. On active contour models and balloons. *CVGIP*, 53(2) :211–218 (1991). 56
- [Coifman(1990)] R. R. Coifman. Wavelet analysis and signal processing. In Louis Auslander, Tom Kailath, and Sanjoy K. Mitter, editors, *Signal Processing, Part I : Signal Processing Theory*, pages 59–68. Springer-Verlag, New York, NY (1990). 66

- [Concus et al.(1976)] Paul Concus, Gene H. Golub, and Dianne P. O’Leary. A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations. *Sparse Matrix Computations*, pages 309–332 (1976). 171
- [Corpetti(2002)] Thomas Corpetti. *Estimation et analyse de champs denses de vitesses d’écoulements fluides*. Ph.D. thesis, Université de Rennes 1, Ecole Doctorale MATISSE (Mathématiques et Applications des Mathématiques, Informatique, Automatique, Robotique, Traitement du Signal et des Images, électronique et Micro-électronique, Opto-électronique et Micro-ondes) (2002). 57
- [Corpetti(2004)] Thomas Corpetti. *Analyse d’écoulements fluides à partir de séquences d’images*. Hermes Science (2004). 113, 116
- [Corpetti et al.(2009)] Thomas Corpetti, P. Héas, Etienne Mémin, and Nicolas Papadakis. Pressure image assimilation for atmospheric motion estimation. *Tellus A*, 61(1) :160–178 (2009). doi :10.1111/j.1600-0870.2008.00370.x. 9, 73
- [Corpetti et al.(2002)] Thomas Corpetti, Etienne Mémin, and P. Pérez. Régularisation div-curl et équation de continuité pour l’estimation du mouvement fluide. In *13ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, RFIA 2002*, volume 3, pages 887–898. Angers, France (2002). 116
- [Courtier(1997)] Philippe Courtier. Dual formulation of four-dimensional variational assimilation. *The Quarterly Journal of the Royal Meteorological Society*, 133(544) :2449 – 2461 (1997). 33
- [Courtier et al.(1998)] Philippe Courtier, E. Andersson, W. Heckley, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, and J. Pailleux. The ecmwf implementation of three-dimensional variational assimilation (3d-var). i : Formulation. *Quarterly Journal of the Royal Meteorological Society*, 124(550) :1783–1807 (1998). 33
- [Daley(1993)] Roger Daley. *Atmospheric Data Analysis*. Cambridge University Press (1993). 69
- [Deb et al.(2008)] S. K. Deb, C. M. Kishtawal, P. K. Pal, and P. C. Joshi. A modified tracer selection and tracking procedure to derive winds using water vapor imagers. *Journal of Applied Meteorology and Climatology*, 47(12) :3252–3263 (2008). doi : 10.1175/2008JAMC1925.1. 58
- [Dennis and Schnabel(1983)] John E. Dennis and Robert B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice Hall (1983). 30, 170
- [Derber and Bouttier(1999)] J. Derber and F. Bouttier. A reformulation of the background error covariance in the ecmwf global data assimilation system. *Tellus*, 51A :195–221 (1999). 33, 69
- [Deriche(1987)] R. Deriche. Using canny’s criteria to derive an optimal edge detector recursively implemented. *Int. J. Computer Vision*, 2 :15–20 (1987). 56
- [Dew and Holmlund(2000)] G. Dew and K. Holmlund. Investigations of cross-correlation and euclidian distance target matching techniques in the mpef environment. In *5th International Winds Workshop*. Lorne, Australia (2000). 58
- [Durand and Froment(2003)] Sylvain Durand and Jacques Froment. Reconstruction of wavelet coefficients using total variation minimization. *SIAM Journal of Scientific Computing*, 24(5) :1754–1767 (2003). 82

- [Elizondo et al.(2002)] David Elizondo, Bernard Cappelaere, and Christèle Faure. Automatic versus manual model differentiation to compute sensitivities and solve non-linear inverse problems. *Comput. Geosci.*, 28(3) :309–326 (2002). ISSN 0098-3004. doi :[http://dx.doi.org/10.1016/S0098-3004\(01\)00048-6](http://dx.doi.org/10.1016/S0098-3004(01)00048-6). 39
- [EUMETSAT(2004)] EUMETSAT. Msg end-user requirements document. Technical Report 2.6, EUMETSAT, Am Kavalleriesand 31, Postfach D-64295 Darmstadt, Germany (2004). 53
- [Evans(1967)] David J. Evans. The use of pre-conditioning in iterative methods for solving linear equations with symmetric positive definite matrices. *J. Inst. Math. Applic.*, pages 295–314 (1967). 171
- [Evensen(1997)] Geir Evensen. Advanced data assimilation for strongly nonlinear dynamics. *Monthly Weather Review*, 125(6) :1342–1354 (1997). 168
- [Fermüller and Aloimonos(2000)] Cornelia Fermüller and Yiannis Aloimonos. Observability of 3d motion. *International Journal of Computer Vision*, 37(1) :43–63 (2000). 57
- [Fitzpatrick(1985)] J. Michael Fitzpatrick. A method for calculating velocity in time dependent images based on the continuity equation. In *Computer Vision and Pattern Recognition*, pages 78–81 (1985). 183
- [Fitzpatrick(1988)] J. Michael Fitzpatrick. The existence of geometrical density-image transformations corresponding to object motion. *Computer Vision, Graphics, and Image Processing*, 44(2) :155–174 (1988). 183
- [Fletcher and Reeves(1964)] R. Fletcher and C. M. Reeves. Function minimization by conjugate gradients. *The Computer Journal*, 7(2) :149–154 (1964). 171
- [Flór and Eames(2002)] Jan-Bert Flór and Ian Eames. Dynamics of monopolar vortices on a topographic beta-plane. *J. Fluid Mech.*, 456 :353–376 (2002). doi :[10.1017/S0022112001007728](https://doi.org/10.1017/S0022112001007728). 76, 78, 84, 92, 185
- [Gelb(1974)] Arthur Gelb. *Applied Optimal Estimation*. MA : MIT Press (1974). 167
- [Giering and Kaminski(1998)] Ralf Giering and Thomas Kaminski. Recipes for adjoint code construction. *ACM Trans. Math. Softw.*, 24(4) :437–474 (1998). ISSN 0098-3500. doi :<http://doi.acm.org/10.1145/293686.293695>. 40
- [Gilbert and Jonsson(2007)] Jean Charles Gilbert and Xavier Jonsson. LIBOPT - an environment for testing solvers on heterogeneous collections of problems - version 1.0. Technical Report RT-0331, INRIA (2007). 83, 184
- [Glowinski and Le Tallec(1987)] Roland Glowinski and Patrick Le Tallec. *Augmented Lagrangian and Operator Splitting Methods in Nonlinear Mechanics*. Society for Industrial and Applied Mathematics ; illustrated edition edition (1987). 110
- [Golub(1983)] Gene H. Golub. *Matrix Computations*. The Johns Hopkins University Press ; 3rd edition (October 15, 1996) (1983). 171
- [Golub and Loan(1996)] Gene H. Golub and Charles F. Van Loan. *Matrix Computations ; 3rd edition*. The Johns Hopkins University Press (1996). 31
- [Gonzalez and Woods(2001)] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*, chapter Thresholding, pages 595–611. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (2001). 54

- [Goody and Yung(1995)] Richard M. Goody and Y. L. Yung. *Atmospheric Radiation : Theoretical Basis*. Oxford University Press 2 edition, USA, 2 edition (1995). 47
- [Gould and Leyffer(2003)] Nicholas I. M. Gould and Sven Leyffer. *Frontiers in Numerical Analysis*, chapter An introduction to algorithms for nonlinear optimization, pages 109–197. Springer Verlag (2003). 26, 27, 30, 170
- [Grossmann and Morlet(1984)] A. Grossmann and J. Morlet. Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis*, 15(4) :723–736 (1984). doi :10.1137/0515056. 66
- [Gupta and Prince(Feb 1996)] Sandeep Gupta and Jerry L. Prince. Stochastic models for div-curl optical flow methods. *Signal Processing Letters, IEEE*, 3(2) :32–34 (Feb 1996). ISSN 1070-9908. doi :10.1109/97.484208. 112, 139
- [Hannon et al.(1996)] Scott Hannon, Larrabee Strow, and Wallace Mcmillan. Atmospheric infrared fast transmittance models : A comparison of two approaches. In *In Proceedings of SPIE Conference 2830, Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research II*, pages 94–105 (1996). 47
- [Hascoët and Pascual(2004)] Laurent Hascoët and Valérie Pascual. Tapenade 2.1 user’s guide. Technical report, INRIA (2004). 41
- [Hellier et al.(2001)] Pierre Hellier, Christian Barillot, Étienne Mémin, and P. Perex. Hierarchical estimation of a dense deformation field for 3d robust registration. *IEEE Trans. Med. Imaging*, 20(5) :388–402 (2001). 57
- [Herlin et al.(2006)] Isabelle Herlin, Etienne Huot, Jean-Paul Berroir, François-Xavier Le Dimet, and Gennady Korotaev. Estimation of a motion field on satellite images from a simplified ocean circulation model. In *ICIP International Conference on Image Processing*, pages 1077–1080. Atlanta, USA (2006). ISSN 1522-4880. doi :10.1109/ICIP.2006.312742. 58
- [Herlin et al.(2004)] Isabelle Herlin, François-Xavier Le Dimet, E. Huot, and J.-P. Berroir. Coupling models and data : which possibilities for remotely-sensed images. In P. Prastacos, U. Cortés, J.-L. Diaz De León, and M. Murillo, editors, *e-Environment : Progress and Challenge of Research on Computing Science*, volume 11, pages 365–383. Instituto Politécnico Nacional, Mexico (2004). 69
- [Hinterberger et al.(2002)] Walter Hinterberger, Otmar Scherzer, Christoph Schnörr, and Joachim Weickert. Analysis of optical flow models in the framework of calculus of variations. *Numerical Functional Analysis and Optimization*, 23(1) :69–89 (2002). doi : 10.1081/NFA-120004011. 114
- [Holmlund(1998)] Kenneth Holmlund. The utilization of statistical properties of satellite-derived atmospheric motion vectors to derive quality indicators. *Weather and Forecasting*, 13(4) :1093–1104 (1998). doi :10.1175/1520-0434(1998)013<1093:TUOSPO>2.0.CO;2. 59
- [Honorat(2007)] Marc Honorat. *Lagrangian data assimilation for river hydraulic numerical simulation*. Ph.D. thesis, INP Grenoble (2007). (in French). 69
- [Horn and Schunck(1981)] Berthold K.P. Horn and Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, pages 185–203 (1981). 57, 113, 114, 116, 183
- [Hoteit(2001)] Ibrahim Hoteit. *Filtres de Kalman réduits et efficaces pour l’assimilation de données en océanographie*. Ph.D. thesis, Université Joseph-Fourier - Grenoble I (2001). 168

- [Huber(1981)] Peter J. Huber. *Robust Statistics*. Wiley-Interscience (1981). 114
- [Hummel and Gidas(1984)] R. A. Hummel and B. C. Gidas. Zero crossings and the heat equation. Technical report, Div. Comput. Sci., Courant Inst. Math. Sci., New York Univ (1984). 56
- [Huot et al.(2006)] Etienne Huot, Till Isambert, Isabelle Herlin, Jean-Paul Berroir, and Gennady Korotaev. Data assimilation of satellite images within an oceanographic circulation model. *IEEE International Conference on Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings, 2006*, 2 :265–268 (2006). ISSN 1520-6149. doi :10.1109/ICASSP.2006.1660330. 58
- [Ide et al.(1997)] K. Ide, P. Courtier, M. Ghil, and A. C. Lorenc. Unified notation for data assimilation : Operational, sequential and variational. *J. Meteorol. Soc.*, 75 :181–189 (1997). 11
- [Jazwinski(1970)] Andrew H. Jazwinski. *Stochastic processes and filtering theory*. New York : Academic Press (1970). 167, 168
- [Jones(1965)] Richard H. Jones. Optimal estimation of initial conditions for numerical prediction. *Journal of the Atmospheric Sciences*, 22(6) :658–663 (1965). 167
- [Kass et al.(1988)] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes : Active contour models. *International Journal of Computer Vision*, 1(4) :321–331 (1988). 56
- [Kay(1993)] Steven M. Kay. *Fundamentals of statistical signal processing : estimation theory*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA (1993). ISBN 0-13-345711-7. 17
- [Kishtawal et al.(2009)] C. M. Kishtawal, S. K. Deb, P. K. Pal, and P. C. Joshi. Estimation of atmospheric motion vectors from kalpana-1 imagers. *Journal of Applied Meteorology and Climatology*, 48(11) :2410–2421 (2009). doi :10.1175/2009JAMC2159.1. 58
- [Korotaev et al.(2007)] Gennady Korotaev, Etienne Huot, F.-X. Le Dimet, I. Herlin, S.V. Stanichny, and D.M. Solovyev. Analysis of the black sea surface currents retrieved from space imagery. In *Rapport du 38ème Congrès de la Commission Internationale pour l'Exploration Scientifique de la mer Méditerranée*, volume 38. Istanbul, Turkey (2007). 58
- [Le Dimet(1982)] François-Xavier Le Dimet. A general formalism of variational analysis. Technical Report 73091, CIMMS (1982). 34
- [Le Dimet et al.(2002)] François-Xavier Le Dimet, Michael I. Navon, and Dacian N. Daescu. Second order information in data assimilation. *Weather Rev*, 130 :629–648 (2002). 38
- [Le Dimet et al.(2009)] François-Xavier Le Dimet, Innocent Souopgui, Olivier Titaud, and Arthur Vidard. Models, data and dmages for dredicting the evolution of geophysical fluids. *Acta Mathematica Vietnamica*, 34 :37–47 (2009). 63
- [Le Dimet and Talagrand(1986)] François-Xavier Le Dimet and Olivier Talagrand. Variational algorithms for analysis and assimilation of meteorological observations : Theoretical aspects. *Tellus*, 38A :97–110 (1986). 34
- [Lehahn et al.(2007)] Y. Lehahn, F. d'Ovidio, M. Lévy, and E. Heifetz. Stirring of the northeast atlantic spring bloom : A lagrangian analysis based on multisatellite data. *J. Geophys. Res.*, 112(C08005) :1–15 (2007). doi :10.1029/2006JC003927. 74
- [Levy(2008)] Bernard C. Levy. *Principles of Signal Detection and Parameter Estimation*. Springer Publishing Company, Incorporated (2008). ISBN 0387765425, 9780387765426. 17

- [Lions(1971)] Jacques Louis Lions. *Optimal Control of Systems Governed by Partial Differential Equations*. Springer Verlag, Berlin (1971). (Translation from the French edition, Dunod, Paris, 1969). 34
- [Liu et al.(2002)] Fei Liu, Xiaodan Song, Yupin Luo, and Dongcheng Hu. Adaptive thresholding based on variational background. *IEEE Electronics Letters*, 38(4) (2002). 55
- [Longuet-Higgins(1981)] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293(5828) :133–135 (1981). doi :10.1038/293133a0. 57
- [Lorenç et al.(2000)] A. C. Lorenç, S. P. Ballard, and coauthors. The met. office global three-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 126(570) :2991–3012 (2000). doi :10.1002/qj.49712657002. 69
- [Luenberger(1984)] David G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley ; 2nd edition (1984). 30, 170, 171
- [Luong et al.(1998)] Bruno Luong, Jacques Blum, and Jacques Verron. A variational method for the resolution of a data assimilation problem in oceanography. *Inverse Problems*, 14(4) :979–997 (1998). 83
- [Ma et al.(2008)] Jianwei Ma, Olivier Titaud, Arthur Vidard, and François-Xavier Le Dimet. Spatio-temporal structure extraction and denoising of geophysical fluid image sequences using 3D curvelet transforms. Research Report RR-6683, INRIA, Grenoble, France (2008). 82
- [Maksyutov et al.(2003)] Shamil Maksyutov, Prabir K. Patra, and Gen Inoue. Pseudo-data inversion of column-co₂ observations by remote sensing using a high resolution inverse model. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume XXXIV, Part. 7/W14. Kyoto, Japan (2003). 48
- [Malardel(2005)] Sylvie Malardel. *Fondamentaux de météorologie : à l'école du temps*. Cépaduès (2005). 45
- [Malladi et al.(1994)] Ravikanth Malladi, James A. Sethian, and Baba C. Vemuri. Evolutionary fronts for topology-independent shape modeling and recovery. In *ECCV '94 : Proceedings of the third European conference on Computer vision (vol. 1)*, pages 3–13. Springer-Verlag New York, Inc., Secaucus, NJ, USA (1994). ISBN 3-540-57956-7. 56
- [Mallat(1989a)] S. Mallat. Multiresolution approximations and wavelet orthonormal bases of $l_2(\mathbb{R})$. *Transactions of the American Mathematical Society*, 315(1) :69–87 (1989a). ISSN 00029947. 66
- [Mallat(1989b)] S. G. Mallat. A theory for multiresolution signal decomposition : The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(7) :674–693 (1989b). ISSN 0162-8828. doi :http://dx.doi.org/10.1109/34.192463. 66
- [Marr and Hildreth(1980)] D. Marr and E.C. Hildreth. Theory of edge detection. In *Roy. Soc. London.*, pages 187–217 (1980). 54, 56
- [Marr and Ullman(1981)] D. Marr and S. Ullman. Directional selectivity and its use in early visual processing. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 211(1183) :151–180 (1981). doi :10.1098/rspb.1981.0001. 54
- [Maurizot(1997)] Mariette Maurizot. *Analyse du mouvement fluide 2D dans des séquences d'images numériques : localisation, caractérisation et suivi des points singuliers*. Ph.D. thesis, Université de Rennes 1, Informatique (1997). 57

- [Mémin and Pérez(1998)] Etienne Mémin and Patrick Pérez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *Image Processing, IEEE Transactions on*, 7(5) :703–719 (1998). ISSN 1057-7149. doi :10.1109/83.668027. 114, 116
- [Meyer(1992)] Yves Meyer. *Wavelets and operators*, volume 37 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge (1992). ISBN 0-521-42000-8; 0-521-45869-2. Translated from the 1990 French original by D. H. Salinger. 66
- [Michel(2008)] Yann Michel. *Assimilation de données d’images télédéteectées en météorologie*. Ph.D. thesis, Université de Paris Est (2008). 77, 161
- [Nagel(1983)] Hans-Hellmut Nagel. Constraints for the estimation of displacement vector elds from image sequences. In *Eighth International Joint Conference on Artificial Intelligence*, pages 945–951. Karlsruhe, West Germany (1983). 114, 117
- [Nodet(2005)] Maëlle Nodet. *Modélisation mathématique et assimilation de données lagrangiennes pour l’océanographie*. Ph.D. thesis, Université de Nice Sophia-Antipolis (2005). (in French). 69
- [Osher and Fedkiw(2003)] Stanley Osher and Ronald Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*, volume 153. Springer Verlag (2003). 56
- [Osher and Sethian(1988)] Stanley Osher and James A. Sethian. Fronts propagating with curvature dependent speed : Algorithms based on hamilton-jacobi formulations. *JOURNAL OF COMPUTATIONAL PHYSICS*, 79(1) :12–49 (1988). 56
- [Otsu(1979)] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trns. on Systems, Man, and Cybernetics*, 9 :62–66 (1979). 54
- [Papadakis and Mémin(2008a)] Nicolas Papadakis and Etienne Mémin. Variational assimilation of fluid motion from image sequences. *SIAM Journal on Imaging Science*, 1(14) :343–363 (2008a). doi :10.1137/080713896. 9
- [Papadakis and Mémin(2008b)] Nicolas Papadakis and Etienne Mémin. A variational technique for time consistent tracking of curves and motion. *Journal of Mathematical Imaging and Vision* (2008b). 9, 73, 117, 161
- [Pham et al.(1996)] Dinh Tuan Pham, Jacques Verron, and Marie Christine Roubaud. A singular evolutive extended kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, 16 :323–340 (1996). 168
- [Phillips(1982)] Norman A. Phillips. On the completeness of multi-variate optimum interpolation for large-scale meteorological analysis. *Monthly Weather Review*, 110(10) :1329–1334 (1982). 165
- [Powell(1984)] M. J. D. Powell. *Lecture Notes in Mathematics*, volume 1066, chapter Non-convex minimization calculations and the conjugate gradient method, pages 122–141. Springer Verlag (1984). 171
- [Purser et al.(2003)] James Purser, Wan-Shu Wu, David F. Parrish, and Nigel M. Roberts. Numerical aspects of the application of recursive filters to variational statistical analysis. part ii : Spatially inhomogeneous and anisotropic general covariances. *Monthly Weather Review*, 131(8) :1536–1548 (2003). 34
- [Rohn et al.(2001)] Michael Rohn, Graeme Kelly, and Roger W. Saunders. Impact of a new cloud motion wind product from meteosat on nwp analyses and forecasts. *Monthly*

- Weather Review*, 129(9) :2392–2403 (2001). doi :10.1175/1520-0493(2001)129<2392:IOANCM>2.0.CO;2. 59, 71
- [Saad(2003)] Yousef Saad. *Iterative Methods for Sparse Linear Systems, Second Edition*. Society for Industrial and Applied Mathematics 2 edition (2003). 31
- [Sasaki(1955)] Yoshi K. Sasaki. A fundamental study of the numerical prediction based on the variational principle. *J. Meteor. Soc. Japan*, 33(6) :262–275 (1955). 22
- [Sasaki(1958)] Yoshi K. Sasaki. An objective analysis based on the variational method. *J. Meteor. Soc. Japan*, 36(3) :77–88 (1958). 22
- [Scharf(1991)] Louis Scharf. *Statistical Signal Processing : Detection, Estimation, and Time Series Analysis*. Prentice-Hall, Inc. (1991). ISBN 0201190389. 17
- [Schmetz et al.(1993)] J. Schmetz, K. Holmlund, J. Hoffman, B. Strauss, B. Mason, V. Gaertner, A. Koch, and L. Van De Berg. Operational cloud-motion winds from meteosat infrared images. *J. Appl. Meteor.*, 32 :1206–1225 (1993). 9
- [Schnabel et al.(1985)] Robert B. Schnabel, John E. Koonatz, and Barry E. Weiss. A modular system of algorithms for unconstrained minimization. *ACM Trans. Math. Softw.*, 11(4) :419–440 (1985). ISSN 0098-3500. doi :http://doi.acm.org/10.1145/6187.6192. 171
- [Schnörr(1994)] Christoph Schnörr. Segmentation of visual motion by minimizing convex non-quadratic functionals. In *12th IAPR International Conference on Pattern Recognition, 1994.*, volume 1, pages 661–663. IEEE Computer Society Press, Jerusalem, Israel (1994). doi :10.1109/ICPR.1994.576391. 114, 117, 139
- [Sezgin and Sankur(2004)] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation (2004). 54
- [Snieder and Trampert(2000)] Roel Snieder and Jeannot Trampert. *Inverse Problems in Geophysics*, chapter Wavefield inversion, pages 119–190. Springer Verlag, New York (2000). 108
- [Sobel and Feldman(1968)] Sobel and Feldman. A 3x3 isotropic gradient operator for image processing (1968). Presented at a talk at the Stanford Artificial Project in 1968, unpublished but often cited, orig. in *Pattern Classification and Scene Analysis*, Duda,R. and Hart,P., John Wiley and Sons,'73, pp271-2. 56
- [Souopgui et al.(2010a)] Innocent Souopgui, Emmanuel Kamgnia, François-Xavier Le Dimet, and Arthur Vidard. A new approach for regularization of inverse problems in images processing. In *10th African Conference on Research in Computer Science and Applied Mathematics*. Yamoussoukro, Côte d'Ivoire (2010a). Accepted. 107
- [Souopgui et al.(2009)] Innocent Souopgui, François-Xavier Le Dimet, and Arthur Vidard. Vector field regularization by generalized diffusion. Research Report RR-6844, INRIA, Grenoble, France (2009). 82, 83, 154
- [Souopgui et al.(2010b)] Innocent Souopgui, François-Xavier Le Dimet, and Arthur Vidard. Generalised diffusion based regularization for inverse problem in images processing. In *2nd International Conference on Engineering Optimization(ENGOPT2010)*. Lisbon, Portugal (2010b). Accepted. 107
- [Souopgui et al.(2010c)] Innocent Souopgui, François-Xavier Le Dimet, and Arthur Vidard. Régularisation par diffusion généralisée pour les problèmes inverses mal posés : application au traitement d'images en géophysique. In *40th French National Meeting on Numerical Analysis (CANUM 2010)*. Carcans-Maubuisson, France (2010c). 107

- [Starck et al.(2002)] Jean-Luc Starck, Emmanuel J. Candes, and David L. Donoho. The curvelet transform for image denoising. *Image Processing, IEEE Transactions on*, 11(6) :670–684 (2002). doi :10.1109/TIP.2002.1014998. 102
- [Storto and Tveter(2009)] Andrea Storto and Frank Thomas Tveter. Assimilating humidity pseudo-observations derived from the cloud profiling radar aboard cloudsat in aladin 3d-var. *Meteorological Applications*, 16 :461–479 (2009). doi :10.1002/met.144. 48
- [Strow et al.(2006)] Larrabee Strow, Scott E. Hannon, Sergio De Souza Machado, Howard E. Motteler, and David Tobin. Validation of the atmospheric infrared sounder radiative transfer algorithm. *J. Geophys. Res*, 111 (2006). 47
- [Strow et al.(2003)] Larrabee Strow, Scott E. Hannon, Sergio De Souza Machado, Howard E. Motteler, and David Tobin. An overview of the airs radiative transfer model. *J. Geophys. Res*, 41(2) :303–313 (2003). 47
- [Suter(1994)] David Suter. Motion estimation and vector splines. In *CVPR94*, pages 939–942 (1994). 115
- [Tao et al.(2003)] Wen-Bing Tao, Jin-Wen Tian, and Jian Liu. Image segmentation by three-level thresholding based on maximum fuzzy entropy and genetic algorithm. *Pattern Recogn. Lett.*, 24(16) :3069–3078 (2003). ISSN 0167-8655. doi :http://dx.doi.org/10.1016/S0167-8655(03)00166-1. 55
- [Tikhonov(1963)] A. N. Tikhonov. Regularization of incorrectly posed problems. *Soviet Math*, 4 :1624–1627 (1963). 111
- [Titaud et al.(2008)] Olivier Titaud, Arthur Vidard, Innocent Souopgui, and François-Xavier Le Dimet. Assimilation of image sequences in numerical models. Technical Report RR6701, INRIA (2008). 63
- [Titaud et al.(2009)] Olivier Titaud, Arthur Vidard, Innocent Souopgui, and François-Xavier Le Dimet. Assimilation of image sequences in numerical models. *Tellus A*, 62(1) :30–47 (2009). doi :10.1111/j.1600-0870.2009.00416.x. 63, 185
- [Torre and Poggio(1986)] V. Torre and T. Poggio. On edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(2) :147–163 (1986). ISSN 0162-8828. 54
- [Toye(2009)] Habib Toye. Interfacing image processing libraries with geophysical fluid forecasting numerical systems (2009). Second year engineering school training (ENSIMAG). 161
- [Tsuyuki(2000)] T. Tsuyuki. Summary of comments from nwp centers represented in wgne on the large differences in satellite wind observation errors assigned at nwp centers. Technical report, CGMS, Geneva, Switzerland, 5 pp (2000). 59, 71
- [Vidard(2001)] Arthur Vidard. *Vers une prise en compte des erreurs modèle en assimilation de données 4D variationnelle : Application à un modèle réaliste d’océan*. Ph.D. thesis, Université Joseph-Fourier (2001). 33, 79
- [Vidard et al.(2008)] Arthur Vidard, François-Xavier Le Dimet, Innocent Souopgui, and Olivier Titaud. Assimilation of images in numerical models in geophysics. In José Herskovits, editor, *EngOpt 2008*. COPPE/UFRJ, Rio de Janeiro Brésil (2008). ISBN 978-85-7650-152-7. 63
- [Vidard et al.(2004)] Arthur Vidard, Andrea Piacentini, and François-Xavier Le Dimet. Variational data analysis with control of the forecast bias. *Tellus*, 56A :177–188 (2004). 82

- [Vigan et al.(2000)] Xavier Vigan, Christine Provost, Rainer Bleck, and Philippe Courtier. Sea surface velocity from sea surface temperature image sequences. 1. method and validation using primitive equation model output. *Journal of Geophysical Research*, 105(C8) :19,499–19,514 (2000). Estimation du mouvement dans les images. 116
- [Wang et al.(1992)] Zhi Wang, François-Xavier Le Dimet Michael I. Navon, and X. Zou. The second order adjoint analysis : Theory and applications. *Meteorology and Atmospheric Physics*, 50 :3–20 (1992). 38
- [Weaver and Courtier(2001)] Anthony Weaver and Philippe Courtier. Correlation modelling on the sphere using a generalized diffusion equation. *The Quarterly Journal of the Royal Meteorological Society*, 127(575) :1815–1846 (2001). ISSN 1477-870X. 33
- [Weickert et al.(2006)] Joachim Weickert, Andrés Bruhn, Thomas Brox, and Nils Papenberg. *Mathematical Models for Registration and Applications to Medical Imaging*, volume 10 of *Mathematics in Industry*, chapter A Survey on Variational Optic Flow Methods for Small Displacements, pages 103–136. Springer Berlin Heidelberg (2006). doi :10.1007/978-3-540-34767-5. 114
- [Weickert and Schnörr(2001)] Joachim Weickert and Christoph Schnörr. A theoretical framework for convex regularizers in pde-based computation of image motion. *Int. J. Comput. Vision*, 45(3) :245–264 (2001). ISSN 0920-5691. doi :http://dx.doi.org/10.1023/A:1013614317973. 114, 115, 132
- [Williams and Shah(1993)] Donna J. Williams and Mubarak Shah. Edge characterization using normalized edge detector. *CVGIP : Graph. Models Image Process.*, 55(4) :311–318 (1993). ISSN 1049-9652. doi :http://dx.doi.org/10.1006/cgip.1993.1021. 54
- [Wong and Sahoo(1989)] A.K.C. Wong and P. K. Sahoo. A gray-level threshold selection method based on maximum entropy principle. *IEEE Transactions on Systems, Man and Cybernetics*, 19(4) (1989). 55
- [Xu and Prince(1998a)] Chenyang Xu and Jerry L. Prince. Generalized gradient vector flow external forces for active contours. *Signal Processing — An International Journal*, 71(2) :131–139 (1998a). 56
- [Xu and Prince(1998b)] Chenyang Xu and Jerry L. Prince. Snakes, shapes, and gradient vector flow. *TIP*, 7(3) :359–369 (1998b). 144
- [Ziou and Tabbone(1998)] D. Ziou and S. Tabbone. Edge detection techniques - an overview. *International Journal of Pattern Recognition and Image Analysis*, 8 :537–559 (1998). 54

Liste des tableaux

1.1	Convention d'utilisation des symboles	11
1.2	Convention d'utilisation des symboles	12
3.1	Résolutions horizontales et dimensions des images METEOSAT	50
5.1	Quelques termes de régularisation pour le flot optique	114
5.2	Estimation du mouvement avec différents termes de régularisation, comparaison des fonctions de diagnostic à convergence	130
5.3	Expression du gradient de la fonction d'énergie pour les fonctions de régularisation usuelles	140
5.4	Comparaison des fonctions de diagnostic pour différentes approches de régularisation	151

Table des figures

2.1	Illustration des minima en minimisation sans contrainte	27
2.2	Illustration graphique du minimum sous contraintes	28
3.1	Répartition et couverture de observations du réseau synoptique	45
3.2	Répartition des pseudo-Observations de champs de vitesse	46
3.3	Couverture géographique par les satellites météorologiques géostationnaires . .	51
3.4	Images METEOSAT du canal vapeur d'eau	51
3.5	Principe de l'imagerie par le radiomètre SEVERI	52
3.6	Couverture nominale de la Terre par les images MSG.	52
3.7	Carte de la résolution au sol des images MSG 1.5.	53
4.1	Structures observée dans les images du Gulf Stream vu du satellite Aqua	65
4.2	Séquence d'images METEOSAT de la couverture nuageuse au dessus de l'Europe	70
4.3	Photographie de la plate-forme Coriolis.	79
4.4	Comparaison du champ de vitesse vrai et du champ de vitesse analysé	85
4.5	Comparaison de la séquence d'images vraies et de la séquence d'images analysées	86
4.6	Courbe d'évolution des erreurs normalisées	88
4.7	Assimilation directe d'images, terme d'écart aux observations par image	89
4.8	Comparaison de l'état analysé et de l'état vrai	90
4.9	Diagnostics sur le champ de vitesse analysé par assimilation d'images	91
4.10	Etat analysé par assimilation directe d'une séquence d'images représentées en pixels	95
4.11	Détails sur le champ reconstruit par assimilation d'images représentées en pixels	96
4.12	Etat analysé par assimilation d'images avec décomposition en curvelets et seuillage brut	97
4.13	Détails sur l'état analysé par assimilation d'images avec décomposition en cur- velets et seuillage brut	98
4.14	Etat analysé par assimilation d'images avec décomposition en curvelets et seuillage brut, échelle graphique normale	99
4.15	Etat analysé par assimilation d'images avec décomposition en curvelets et seuillage échelle par échelle	100
4.16	Détails sur l'état analysé par assimilation d'images avec décomposition en cur- velets et seuillage échelle par échelle	101
4.17	Détails sur l'état analysé par assimilation d'images avec différents opérateurs d'observation	103

4.18	Assimilation directe d'images, comparaison de l'écart aux observations avec différents opérateurs d'observation	104
4.19	Assimilation directe d'images, comparaison de l'écart aux observations pour différents seuillages	105
5.1	Estimation du mouvement sans régularisation : évolution des fonctions de diagnostic	121
5.2	Comportement de la régularisation de Tikhonov en fonction du paramètre de pondération	122
5.3	Comportement de la pénalisation du gradient en fonction du paramètre de pondération	124
5.4	Comportement de la régularisation isotropique basée sur les données en fonction du paramètre de pondération	125
5.5	Comportement de la pénalisation du gradient au cours des itérations	126
5.6	Evolution comparée de l'écart aux observations pour différents termes de régularisation	127
5.7	Evolution comparée de l'erreur moyenne pour différents termes de régularisation	128
5.8	Evolution comparée de l'erreur angulaire pour différents termes de régularisation	129
5.9	Evolution comparée de l'erreur sur la vorticité pour différents termes de régularisation	131
5.10	Comportement de la diffusion généralisée en fonction du paramètre de pondération	145
5.11	Comportement de la diffusion généralisée au cours des itérations	147
5.12	Evolution de l'écart aux observations pour différentes approches de régularisation	148
5.13	Evolution de l'erreur moyenne pour différentes approches de régularisation . . .	149
5.14	Evolution de l'erreur angulaire pour différentes approches de régularisation . . .	150
5.15	Evolution de l'erreur sur la vorticité pour différentes approches de régularisation	152
5.16	Champs reconstruit par différentes approches de régularisation	153
5.17	Sensibilité du terme d'écart aux observations sur la fonction de confiance	155
5.18	Evolution de l'erreur moyenne avec deux fonctions de confiance	156
5.19	Evolution de l'erreur angulaire avec deux fonctions de confiance	157
5.20	Evolution de l'erreur sur la vorticité avec deux fonctions de confiance	158
6.1	Illustration d'une fonction constante à support limité	163
6.2	Illustration de la norme \mathcal{L}^2 sur des translations de faible amplitude	163
6.3	Illustration de la norme \mathcal{L}^2 sur des translations de grandes amplitudes	164
B.1	Comportement de la descente de plus grande pente	170
C.1	Répartition et couverture de observations du réseau synoptique	178
C.2	Répartition des observations par bouées	178
C.3	Répartition des observations par avions de ligne	179
C.4	Trace d'un satellite à orbite quasi polaire lors d'une révolution autour de la Terre . . .	180
C.5	Répartition des observations par le sondeur AMSU-A	181
C.6	Répartition des observations de concentration d'ozone	181
D.1	champ de vitesse vrai pour les expériences	186

E.1 Estimation du mouvement sans régularisation : évolution des fonctions de diagnostic 187

E.2 Comportement de la régularisation de Tikhonov en fonction du paramètre de pondération 189

E.3 Comportement de la pénalisation du gradient en fonction du paramètre de pondération 190

E.4 Comportement de la pénalisation du rotationnel en fonction du paramètre de pondération 191

E.5 Comportement de la pénalisation de la divergence en fonction du paramètre de pondération 192

E.6 Comportement de la pénalisation du gradient du rotationnel en fonction du paramètre de pondération 193

E.7 Comportement de la pénalisation du gradient de la divergence en fonction du paramètre de pondération 194

E.8 Comportement de la régularisation isotropique basée sur les données en fonction du paramètre de pondération 195

E.9 Comportement de la régularisation isotropique basée sur le champ en fonction du paramètre de pondération 195

E.10 Comportement de la pénalisation du gradient au cours des itérations 196

E.11 Comportement de la régularisation div-curl du second ordre au cours des itérations 196

E.12 Comportement de la pénalisation du rotationnel au cours des itérations 197

E.13 Comportement de la pénalisation de la divergence au cours des itérations 197

E.14 Comportement de la pénalisation du gradient du rotationnel au cours des itérations 198

E.15 Comportement de la pénalisation du gradient de la divergence au cours des itérations 198

E.16 Comportement de la régularisation isotropique basée sur les données au cours des itérations 199

E.17 Comportement de la régularisation isotropique basée sur le champ au cours des itérations 199

E.18 Evolution comparée de l'écart aux observations avec différents terme de régularisation 200

E.19 Evolution comparée de l'erreur moyenne avec différents terme de régularisation 201

E.20 Evolution comparée de l'erreur angulaire avec différents terme de régularisation 201

E.21 Evolution comparée de l'erreur sur la vorticité avec différents terme de régularisation 202

Index

- adjoint, 34, 38
 - discret, 36
 - modèle, 34, 35, 74
- AMV, 46, 58, 59, 69
- assimilation de données, 8, 29, 43, 107, 159, 165
 - approche stochastique ou séquentielle, 8, 165
 - approche variationnelle, 8, 17, 22, 32
- assimilation directe d'images, 9, 56, 71, 79, 107, 160
- binarisation, 54
- conditionnement, 25, 30
 - mauvais, 25, 108, 109
 - pré, 25, 30
- contour actif, 56
- crête, 56
 - algorithme de Canny, 56
 - opérateur, 56
 - opérateur de Robert, 56
 - opérateur de Sobel, 56
- curvelet, 67
 - décomposition, 67, 80, 81, 161
- différentiation automatique, 39, 40
- diffusion généralisée, 33, 133, 140, 142, 143
- ébauche, 110, 165, 167
 - covariance d'erreur, 22, 32, 110, 165, 167
 - erreur, 165
- erreur
 - angulaire, 119, 148, 154, 176
 - représentativité, 44, 47
- estimateur du minimum de variance, 18, 20
- estimateur du mouvement, 143
- estimation bayésienne, 17
- estimation du mouvement, 57
 - flot optique, 57
- estimation par moindre carré, 20
- filtre de Kalman, 167
 - ensemble, 168
 - étendu, 168
 - SEEK, 168
- flot optique, voir estimation du mouvement, 73
- fonction coût, 17, 21–26, 34, 72, 80, 154
 - gradient, 36
- fonction de confiance, 152, 154
- fonction objectif, voir fonction coût, voir fonction coût
- fourier, 66
 - décomposition, 66
- géophysique
 - fluide, 7
 - système, 106
- géophysique, 7
- gradient, 34
- image, 8, 49, 50, 56, 63, 65, 72, 80, 159
 - modèle, 58
 - niveau d'abstraction, voir niveau d'interprétation
 - niveau d'interprétation, 63
 - opérateur, 73, 75, 77, 80, 161
 - segmentation, 54
 - séquence d'images, 8, 69, 159
- interpolation optimale, 165
- mal posé, 25, 109, 133
- maximum de vraisemblance, 19–22
 - estimateur, 20
- mesure in situ, voir observation, voir observation
- minimisation, 25, 29, 154, 169
- minimum
 - local, 26, 29

- modèle, 7, 11, 43, 78, 107, 108, 159, 160, 175
 - continu, 11
 - direct, 23, 35
 - discret, 13
 - erreur, 84
 - linéaire tangent, 35
- modèles, 24
 - covariances d'erreur, 167
 - discret, 24, 167
- multiplicateur de Lagrange, 28, 29
- observation, 7, 21, 24, 25, 43, 79, 108, 159, 165, 167
 - conventionnelle, 43, 59, 177
 - covariance d'erreur, 22, 47, 75
 - écart, 72, 74, 154
 - erreur, 165
 - in situ, 44, 177
 - opérateur, 13, 22, 75, 77, 109, 159, 165, 167
 - pseudo, 47, 48, 56, 69, 159
 - pseudo-, 9
 - réseau, 44, 177
- ondelette, 66
 - décomposition, 66, 161
- optimalité, 165
 - condition, 26, 28
 - système, 38, 72
 - système, 69
- optimisation, 26, 83
 - sous contraintes, 28, 29
- paramètre, 7, 8, 21, 25, 107
- préconditionnement, 82, 143
- problème d'ouverture, 77
- problème inverse, 25, 32, 107, 108, 133
- régularisation, 25, 38, 77, 82, 107, 111, 112, 133, 159
- seuillage, 54, 66, 80, 105
 - adaptatif, 55
 - brut, 55, 81
 - doux, 55
 - échelle par échelle, 81
- statistiques, 8, 13, 159
- structure, 64
- télémétrie, voir observation
- variable d'état, 11, 78
- variable de contrôle, 24, 79
- vecteur d'état, voir variable d'état
- vision par ordinateur, 53

Résumé

La compréhension et la prévision de l'évolution des fluides géophysiques sont d'une importance capitale et constituent un domaine de recherche scientifique aux enjeux conséquents. Une bonne prévision est basée sur la prise en compte de toutes les informations disponibles sur le système considéré. Ces informations incluent les modèles, les observations et les connaissances *a priori*. L'assimilation de données permet de les combiner de façon optimale pour déterminer les entrées du modèle. Les dernières décennies ont vu croître en densité et en qualité la couverture satellitaire produisant, entre autres, des séquences d'images montrant l'évolution dynamique de certains phénomènes géophysiques tels que les dépressions et les fronts. Ces séquences d'images sont jusqu'à présent sous-utilisées en assimilation de données. Cette thèse propose une extension de l'assimilation variationnelle de données aux observations de type séquence d'images. Après avoir présenté les images, leur utilisation actuelle et ses limites, nous introduisons les notions de niveau d'interprétation, d'espaces et d'opérateur image. Ces notions sont utilisées pour formuler l'assimilation directe de séquences d'images. Nous proposons également une nouvelle approche de régularisation par diffusion généralisée pour les problèmes inverses. Les résultats préliminaires en traitement d'images et en assimilation directe de séquences d'images montrent une méthode prometteuse qui résout la plupart des problèmes rencontrés avec les approches classiques de régularisation.

Mots clés : assimilation de données, assimilation d'images, fluide géophysique, régularisation.

Abstract

Understanding and forecasting the evolution of geophysical fluids is a major scientific and societal challenge. A good forecast must take into account all available information on the studied system. These informations include models, observations and *a priori* knowledge. Data assimilation techniques combine all these informations in a consistent way to produce model inputs. During the last decades, many satellites were launched to increase the knowledge of earth. They produce, among others, image sequences showing the dynamical evolution of geophysical processes such as depressions and fronts. These images sequences are currently under-utilized in data assimilation.

This thesis presents a consistent approach for taking into account image sequences in variational data assimilation. After a presentation of images, their current use and its limitation, we introduce the concepts of interpretation level, image space and image operator used for direct image sequences assimilation. We also propose a new approach of regularization based on generalized diffusion for ill-posed inverse problems. Preliminary results on image processing and image sequences assimilation show a promising approach that solve most of the problems encountered with classical approaches of regularization.

Keywords : data assimilation, images assimilation, geophysical fluid, regularization.