



**HAL**  
open science

# Assimilation de mesures satellitaires dans des modèles numériques par méthodes de contrôle optimal

François van den Berghe

► **To cite this version:**

François van den Berghe. Assimilation de mesures satellitaires dans des modèles numériques par méthodes de contrôle optimal. Traitement du signal et de l'image [eess.SP]. École Nationale Supérieure des Mines de Paris, 1992. Français. NNT : . pastel-00954478

**HAL Id: pastel-00954478**

**<https://pastel.hal.science/pastel-00954478v1>**

Submitted on 3 Mar 2014

**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.

**T H E S E**

présentée à

L'ECOLE NATIONALE SUPERIEURE DES MINES DE PARIS

par

**François VAN DEN BERGHE**

Ingénieur ENSIEG

pour l'obtention du titre de

**DOCTEUR**

en

**SCIENCES ET TECHNIQUES D'IMAGE**

---

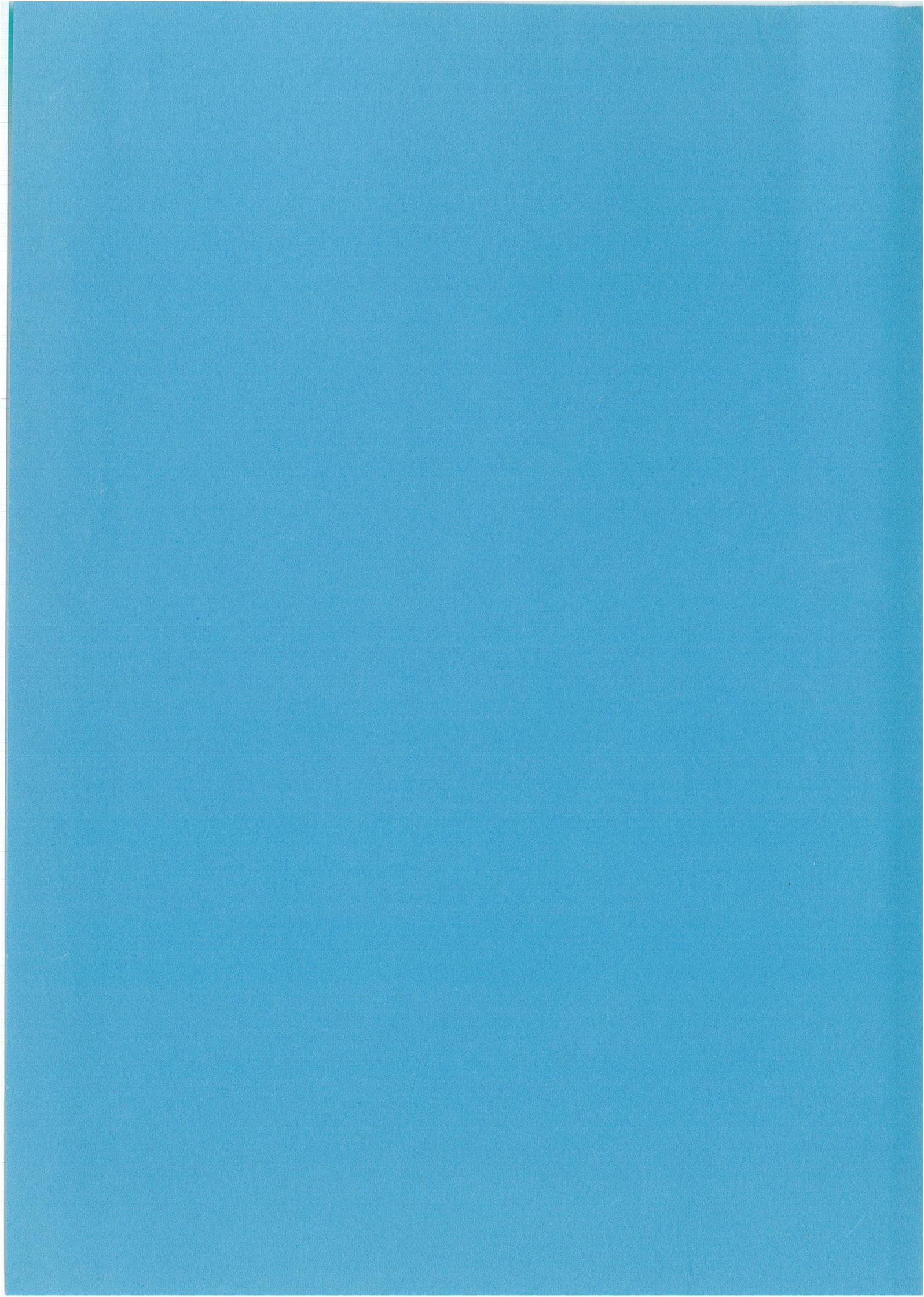
**ASSIMILATION DE MESURES SATELLITAIRES  
DANS DES MODELES NUMERIQUES  
PAR METHODES DE CONTROLE OPTIMAL**

---

Soutenu le 11 décembre 1992, devant le jury composé de :

Président : M. Pierre GUEVEL  
Membres : MM. Michel CREPON  
Hervé GUILLARD  
François-Xavier LE DIMET  
Lucien WALD  
Jean-Paul ZOLELIO

I.M.T  
Univ. Paris VI  
I.N.R.I.A  
I.M.A.G  
E.N.S.M.P  
Univ. Nice





## AVANT-PROPOS

Je voudrais tout d'abord remercier mes professeurs de DEA, et en particulier les Professeurs Chantal JOURDAIN et Jean-Louis LACOUME ainsi que Daniel BAUDOIS. Ce sont leurs enseignements qui ont suscité en l'élève-ingénieur, que j'étais alors, l'envie de découvrir la Recherche.

Toutefois, ma bonne volonté aurait été vaine, si Lucien WALD ne m'avait proposé ce travail et, surtout, s'il ne m'avait dirigé tout au long de son exécution. Je l'en remercie profondément, comme je n'oublierai pas de remercier le Professeur Jean-Marie MONGET qui m'a accueilli dans son laboratoire.

Mes remerciements vont aussi aux membres du jury qui ont accepté de juger ces travaux. Je remercie le Professeur Pierre GUEVEL dont les remarques et les critiques ont été très constructives dans la rédaction finale de ce manuscrit, Michel CREPON qui m'a inculqué les rudiments des sciences atmosphériques, Hervé GUILLARD qui m'a très souvent aidé lorsque mes recherches se trouvaient dans une impasse, le Professeur François-Xavier LE DIMET, beaucoup de résultats présentés ici sont obtenus grâce à ses travaux et Jean-Paul ZOLELIO qui était toujours disponible lorsque je butais sur des difficultés mathématiques.

Je ne voudrais pas non plus oublier tous ceux qui de près ou de (plus) loin m'ont aidé à la réalisation des ces travaux. Aussi, je remercie celle-même Roselyne ADDE, Pascale DELECLUSE, Françoise MARTEL, Grace WAHBA, Michel ALBUISSON, François CAUNEAU, Philippe COURTIER, Ross HOFFMAN, Claude MILLOT, Thierry RANCHIN, Olivier TALAGRAND, Jean-Noël THEPAUT... sans oublier ceux dont la place manque pour être cités.

*" Alles hat ein Ende, nur die Wurst hat zwei "*

Chanson populaire allemande.

## RESUME

Nous nous intéressons aux méthodes permettant d'assimiler les données de télédétection. Il y a principalement deux types de technique : les méthodes séquentielles, issues du filtrage de Kalman et les méthodes variationnelles, basées sur les équations adjointes du contrôle optimal. Les méthodes variationnelles sont récentes et nous essayons de présenter leur intérêt pour la télédétection à l'aide de deux applications.

Une première étude est menée sur un modèle numérique d'advection-diffusion. Elle montre la faisabilité numérique de la méthode : il est possible de minimiser une fonction de coût mesurant la distance entre la trajectoire du modèle et des observations réparties à la fois dans le temps et dans l'espace. La deuxième application est réalisée sur des observations effectuées par le diffusiomètre-vent embarqué à bord du satellite Seasat. Son objet est la cartographie des champs de vent à la surface des océans. Sur une période de douze heures, nous assimilons les observations dans un modèle numérique non linéaire, discrétisant l'équation de la vorticité sur un domaine rectangulaire couvrant à peu près tout l'océan Atlantique nord.

Une comparaison entre les champs de vent, obtenus par l'assimilation variationnelle, et ceux produits par une assimilation séquentielle, met en évidence la propriété de rétro-propagation de l'information dans le temps, que possèdent les équations adjointes. Il en résulte des analyses sensiblement meilleures, surtout en début de période d'assimilation, et une grande cohérence dynamique entre les champs analysés et le modèle numérique.

## SOMMAIRE

1.	INTRODUCTION .....	8
2.	THEORIE DE L'ESTIMATION ET ASSIMILATION DE DONNEES .....	18
	2.1 L'estimateur des moindres carrés.....	19
	2.2 Forme séquentielle de l'assimilation de données : le filtre de Kalman.....	23
	2.3 Commentaires.....	26
	2.4 Le lissage de Kalman.....	29
	2.5 Dualité entre l'estimation statistique et le contrôle optimal.....	31
	2.6 Choix d'une procédure d'assimilation .....	33
3.	ASSIMILATION VARIATIONNELLE ET EQUATIONS ADJOINTES .....	35
	3.1 Formulation variationnelle du problème de l'assimilation de donnée.....	36
	3.2 Utilisation des équations adjointes.....	39
	3.3 Modèles non linéaires et problèmes de convergence .....	44
	3.4 Détermination de la solution par pénalisation-régularisation.....	47
	3.5 Mise en oeuvre des techniques de contrôle optimal.....	51
4.	ASSIMILATION DANS UN MODELE D'ADVECTION-DIFFUSION : MODELE ET OBSERVATIONS .....	56
	4.1 Inversion, assimilation de données et traceurs .....	57
	4.2 Position du problème.....	58
	4.3 Le modèle numérique .....	59
	4.4. Distribution spatio-temporelle des observations .....	63

4.5 Synthèse d'observations de télédétection .....	64
4.6 Fonction de coût .....	68
4.7 Utilisation du modèle adjoint .....	70
<b>5. ASSIMILATION DANS UN MODELE D'ADVECTION-DIFFUSION : RESULTATS NUMERIQUES .....</b>	<b>73</b>
5.1 Test de la procédure d'assimilation.....	74
5.2 Présentation générale des résultats.....	75
5.3 Commentaires.....	81
5.4 Applications possibles .....	85
5.5 Conclusion du chapitre 5 .....	86
<b>6. ASSIMILATION D'OBSERVATIONS D'UN DIFFUSIOMETRE VENT SPATIO- PORTE.....</b>	<b>88</b>
6.1 Les données .....	89
6.2 Le modèle numérique .....	93
6.3 La fonction de coût .....	96
6.4 Assimilation variationnelle.....	98
6.5 Assimilation séquentielle.....	99
6.6 Résultats numériques.....	100
6.7 Commentaires.....	104
6.8 Conclusion du chapitre 6 .....	107
<b>7. CONCLUSION .....</b>	<b>110</b>
<b>8. REFERENCES BIBLIOGRAPHIQUES .....</b>	<b>115</b>

9. ANNEXES .....	124
9.1 Méthode adjointe.....	125
9.2 Utilisation du Lagrangien.....	129
9.3 Variational assimilation of remote sensing observations for the mapping of pollutant sources in lakes (communication) .....	130
9.4 Assimilation of satellite borne wind scatterometer observations (communication) ..	137

**Chapitre 1 :**

**INTRODUCTION**

L'environnement terrestre est le siège d'une multitude de phénomènes physiques complexes qui interagissent entre eux et évoluent dans le temps. Ces systèmes dynamiques sont difficilement observables, notamment lorsqu'il s'agit de l'océan ou de l'atmosphère. A chaque instant l'état de l'un de ces systèmes peut être décrit par un ensemble des valeurs, appelé champ, prises en chaque point du système par certaines variables physiques qui le caractérisent. Les difficultés d'obtention de la mesure font que ces champs ne sont connus de façon routinière qu'en très peu de points (comme par exemple les stations météorologiques disséminées à la surface de la terre ou embarquées sur des navires). Ils doivent donc être estimés en tous les autres points du système.

La procédure qui produit à un instant donné des champs physiques homogènes à partir des observations s'appelle *analyse*. Elle devient *assimilation de données* lorsqu'elle traite des observations distribuées dans le temps en utilisant explicitement un modèle numérique pour simuler l'évolution temporelle du système. Le problème que doit résoudre l'analyse ou l'assimilation de données est un problème d'estimation. Pour le résoudre, il est nécessaire d'utiliser toute l'information disponible : d'une part, des mesures réparties inégalement sur le domaine étudié, et fournies par différents systèmes d'observation ayant leurs propres caractéristiques, et, d'autre part, des relations mathématiques reliant certaines variables du système entre elles.

Dans le cas de l'analyse, les connaissances mathématiques se résument à des lois statistiques, déduites de l'observation du système sur de longues périodes (climatologie). De telles méthodes ont été utilisées pendant longtemps dans les centres météorologiques. Toutefois elles ne prennent pas en compte la dynamique du système, ce qui limite leur usage, notamment pour les prévisions météorologiques (Cressman 1959). Par contre, s'il existe des relations mathématiques décrivant la dynamique du système, il est possible à partir d'un état initial

connu, de calculer pour chaque instant ultérieur les états qu'occupera le système ainsi modélisé, fournissant ainsi une description complète du phénomène physique correspondant.

Ainsi en confrontant la réalité des observations aux connaissances théoriques existantes sur la physique du phénomène, l'assimilation de données se révèle être un puissant moyen d'étude de l'environnement terrestre. C'est pourquoi, longtemps restées l'apanage des météorologues, les techniques d'assimilation de données sont depuis quelques années appliquées à la climatologie ainsi qu'à l'océanographie. Aujourd'hui leur développement est inscrit comme une priorité dans presque tous les grands programmes d'études géophysiques à venir (Johnson *et al.* 1991).

En fait le récent intérêt des géophysiciens pour l'assimilation de données est stimulé par deux facteurs. Le premier est lié au progrès incessant de l'informatique : les super-calculateurs permettent aujourd'hui de mettre au point des modèles numériques très complexes simulant la majeure partie des phénomènes géophysiques. Le second facteur, qui a motivé par ailleurs ce travail, est dû à l'apparition des satellites d'observation de la terre. Les mesures rapportées au sol par les satellites sont de natures très diverses comme par exemple l'imagerie thermique ou visible, les mesures des vents à la surface des océans, ou la concentration des constituants chimiques dans l'atmosphère et la troposphère. Ces observations sont obtenues de façon répétitive sur tout le globe avec une résolution spatiale allant de la dizaine de mètres à la centaine de kilomètres.

Les données de télédétection offrent donc la possibilité d'études fines des phénomènes terrestres rapidement évolutifs, pourvu que l'on soit capable de les assimiler. En effet ces observations présentent certaines particularités qui rendent leur traitement difficile. Premièrement, bien que possédant une grande couverture spatiale, elles sont asynoptiques. Si les mesures sont obtenues pratiquement instantanément lors d'un passage du satellite, plusieurs

passages sont nécessaires pour couvrir géographiquement une région de la taille d'un océan et *a fortiori* le globe. Ainsi les différentes régions du système sont tour à tour observées avec un décalage dans le temps durant lequel l'évolution du système ne peut être négligée. Deuxièmement les satellites ne mesurent généralement pas directement les variables du système mais des grandeurs qui doivent être transformées. Les procédures d'inversion mises en jeu sont souvent complexes, et peuvent amplifier les erreurs de mesures inhérentes aux capteurs. Celles-ci doivent être prises en compte dans l'assimilation, d'autant plus que les équations mathématiques sous-tendant le modèle numérique sont généralement non linéaires et le rendent très sensible aux erreurs. Idéalement une bonne analyse doit être fournie avec l'incertitude qui l'accompagne. Enfin il est nécessaire de traiter en temps réel le flot de données qui est transmis chaque jour aux stations de réception terrestre.

Ces particularités forment un obstacle à l'utilisation des données de télédétection. Ainsi en ce qui concerne la météorologie de l'hémisphère sud, pour lequel très peu d'observations conventionnelles sont disponibles, l'apport des données de télédétection est indéniable, alors qu'il est faible pour l'hémisphère nord, beaucoup mieux couvert par les stations météorologiques (Courtier et Talagrand 1990b). Néanmoins il existe d'autres domaines de la géophysique où elles constituent presque les seules données disponibles. Le cas de l'océanographie a déjà été cité : associées aux progrès des modèles numériques, les données de télédétection permettent d'envisager des prévisions réalistes de l'océan (Ghil et Malanotte-Rizzoli 1991).

Les plus récentes techniques d'assimilation revêtent deux formes : l'assimilation séquentielle et l'assimilation variationnelle (Talagrand 1992). Dans l'assimilation séquentielle, la seule forme utilisée jusqu'à présent pour les prévisions numériques de la météorologie, le modèle numérique est intégré durant l'intervalle de temps sur lequel les observations sont distribuées. Chaque fois que le modèle atteint un instant où une observation est disponible,

l'état calculé par le modèle (appelé d'après l'anglais *first guess*) est corrigé par l'observation. L'intégration du modèle est alors reprise à partir de cet état remis à jour et le processus est répété jusqu'à ce que toutes les observations aient été utilisées. Dans les centres météorologiques opérationnels, l'état ainsi obtenu à la fin de la période d'assimilation est utilisé comme condition initiale au modèle qui par intégration jusqu'à des instants futurs produit la prévision (*forecast* en anglais). L'opération qui consiste à corriger à un instant donné le "first guess" par les observations du moment revient à faire une analyse. L'assimilation séquentielle est donc une suite d'analyses réalisées aux instants d'observations suivies par des intégrations du modèle. Il faut noter, car c'est important pour la suite, que chaque observation est utilisée une seule fois, exactement au moment où elle est disponible.

L'intérêt de l'assimilation séquentielle consiste en la constante mise à jour de l'état prédit par le modèle : chaque observation est utilisée pour corriger la dernière estimation de l'état du système. L'information sur l'état réel du système est ainsi progressivement incorporée dans le modèle et la quantité d'information qu'il contient à la fin de la période d'assimilation lui permet de continuer tout seul pour produire la prévision. Cela fait de l'assimilation séquentielle une méthode très adaptée à la prévision, et c'est pourquoi elle est utilisée de façon opérationnelle depuis plus de dix ans dans certains centres météorologiques. Toutefois cette méthode possède un inconvénient majeur dû précisément à son caractère séquentiel : chaque observation influence les estimations faites sur l'état du système à des instants ultérieurs mais en aucun cas n'influence les estimations des états précédents. Autrement dit, durant la période d'assimilation, l'information apportée par les observations se propage seulement du passé vers le futur. Ceci est dommageable dans le cas où l'assimilation est réalisée sans contrainte de temps réel, comme pour les études climatologiques. Dans de telles situations, il est préférable de procéder à des analyses qui utilisent toutes les observations disponibles durant la période d'assimilation, c'est à dire les observations antérieures aussi bien que postérieures à la date de l'analyse.

L'assimilation variationnelle, quant à elle, vise à globalement ajuster le modèle à toutes les observations disponibles sur la période d'assimilation. Cet ajustement est simultané, et à chaque instant les états du modèle ainsi ajusté sont influencés par toutes les observations antérieures et postérieures à l'instant considéré, évitant ainsi le problème mentionné pour l'assimilation séquentielle. Les algorithmes d'assimilation variationnelle existants sont issues du contrôle optimal (Lions 1969). Une fonction scalaire mesure, pour chaque solution du modèle, prise sur tout l'intervalle d'assimilation, la distance entre cette solution et les observations. Cette fonction appelée fonction de coût est souvent une somme des écarts quadratiques entre les observations et les valeurs du modèle correspondantes. Comme une solution du modèle est uniquement définie sur toute la période d'assimilation par les conditions initiales au début de cette période, celles-ci peuvent être prises comme variables de contrôle, c'est à dire les variables par rapport auxquelles la minimisation va être effectuée. L'état initial qui minimise la fonction de coût est obtenue par une méthode itérative qui nécessite à chaque étape la connaissance de la valeur locale des dérivées partielles, ou gradient, de la fonction de coût par rapport à l'état initial. Ce gradient peut être déterminée au moyens des *équations adjointes* du modèle numérique utilisé par l'assimilation, et qui est appelé par opposition *modèle direct*.

L'assimilation séquentielle est utilisée opérationnellement en météorologie sous le nom d'interpolation optimale qui est une forme simplifiée du filtrage de Kalman, technique introduite en traitement du signal (Kalman 1960) et appliquée ensuite en météorologie (Ghil *et al.* 1981) puis en océanographie (Miller 1986). Les méthodes variationnelles basées sur les équations adjointes sont apparues plus tardivement (Lions 1969). Elles trouvent des applications dans des domaines très divers comme l'automatique ou la mécanique des fluides (Souli et Zolésio 1991) ainsi que dans les sciences atmosphériques (Marchuk 1974, Le Dimet et Talagrand 1986). Les premières études météorologiques ont été faites au début des années 80 (Cacuci et Hall 1984, Lewis et Derber 1985, Talagrand et Courtier 1987) et après 1985 pour les océanographes (Hoffman 1986, Tacker et Long 1988). Contrairement au filtrage de Kalman qui, sous des

formes simplifiées, est utilisé couramment pour les besoins opérationnels de la météorologie, les méthodes adjointes sont toujours en phase de développement et il faudra vraisemblablement attendre la future génération de super-calculateur pour leur utilisation routinière.

Il ne faut pas pour autant tirer des conclusions hâtives quant aux valeurs respectives de ces deux formes d'assimilation. Si l'utilisation opérationnelle des méthodes variationnelles en météorologie n'est pas encore abordable par les moyens de calcul actuels, il en est encore moins de l'algorithme complet du filtre de Kalman dont la mise en oeuvre ne peut même pas être envisagée. De fait les algorithmes d'assimilation séquentielle utilisés actuellement en géophysique sont des approximations à des degrés divers du véritable filtre de Kalman. Néanmoins il est d'usage de regrouper toutes ces variantes sous le nom de filtre de Kalman. Il semble par ailleurs que la solution qui sera retenue en météorologie ne sera ni séquentielle ni variationnelle mais combinera les deux formes (Ghil et Malanotte-Rizzoli 1991, Thépaut 1992). De plus, dans le cadre de la théorie de l'estimation linéaire, on montre que les deux formes d'assimilation conduisent toutes deux à l'estimateur des moindres carrés, et que, seule diffère la façon de calculer cet estimateur. Afin d'avoir une approche globale, dans le chapitre 2 nous poserons le problème de l'assimilation de données dans le contexte générale de la théorie de l'estimation statistique linéaire, et nous présenterons la solution séquentielle du filtre de Kalman. Dans le chapitre 3, nous nous étendrons davantage sur les méthodes variationnelles, telles qu'elles sont considérées en vue de leur utilisation opérationnelle, simultanément au centre européen de prévision météorologique à moyenne échéance et au centre météorologique national aux Etats Unis.

L'un des objets de l'activité du Groupe Télédétection & Modélisation de l'Ecole des Mines de Paris est de promouvoir l'utilisation de données de télédétection pour les études de la géosphère et de son environnement. Quelques travaux y ont été menés combinant les données de la télédétection et la modélisation numérique du comportement des systèmes étudiés. Des

études diverses, portant sur l'urbanisation (Méaille et Wald 1991) avec des méthodes simples d'interpolation, ou sur la cartographie du champ de vent à la surface des océans avec des procédures d'assimilation séquentielles (Cauneau et Chossat 1991), ont montré les potentialités de ce type d'approche. Elles ont aussi permis de cerner les problèmes inhérents à ce type d'approche.

Le but du travail que nous présentons est d'explorer les potentialités des méthodes d'assimilation variationnelles de données de télédétection dans les modèles numériques d'évolution. Pour le réaliser, nous avons choisi une approche pragmatique qui permet d'étudier précisément les difficultés, qu'elles relèvent de l'analyse numérique ou de la particularité de la mesure des satellites. Nous avons considéré deux cas. Le premier est artificiel, et a pour avantage la maîtrise de tous les paramètres, l'autre est un cas réel, relevant de la prévision météorologique marine à très court terme et possédant des applications industrielles importantes.

Notre première étude utilise un modèle numérique d'Advection-Diffusion fourni par l'INRIA de Sophia-Antipolis. L'intérêt de ce modèle pour notre travail est double. Il est robuste et connu, ce qui nous a permis de nous concentrer sur le développement du modèle adjoint. Ce développement a par ailleurs été facilité, et c'est son deuxième intérêt, par la linéarité des équations mathématiques exprimées dans le modèle numérique. Ce modèle, ainsi que le calcul de son adjoint sont présentés dans le chapitre 4. Le modèle direct est ensuite utilisé pour simuler des observations de télédétection telles qu'un radiomètre visible ou infra-rouge d'un satellite pourrait les fournir. Ces observations couvrent un bassin carré représentant schématiquement un lac dans lequel un polluant a subitement été déversé. Dans l'application que nous présentons, nous ne donnons pas de prévision, car elle correspond au cas où toutes les observations sont regroupées en un seul instant qui est justement l'instant final de la période d'assimilation. Le problème que nous nous posons est de retrouver l'état du modèle à l'instant initial, qui a pu produire la situation observée à l'instant final. Ce problème est connu sous le nom d'*inversion*

du modèle, et dans notre cas, les techniques d'assimilation variationnelle sont utilisées pour déterminer douze heures plus tard le point exact où a eu lieu le rejet initial de polluant. Cet exemple est purement illustratif et est peu proche de la réalité. Toutefois, dans cette étude préliminaire, l'intérêt de synthétiser les observations de télédétection est la maîtrise de la quantité et de la qualité des observations afin de déterminer leurs effets sur la convergence, puis sur la vitesse de convergence. Les résultats numériques de ces expériences sont donnés et commentés dans le chapitre 5.

A propos de cet exemple, il faut noter que, quoique simpliste, il dépasse largement la simple illustration des méthodes variationnelles et qu'il pourrait ouvrir un nouveau champ d'application de la télédétection à la lutte contre les pollutions. Ne serait-ce qu'en Europe, il y a de nombreux lacs ou de mers quasi-fermées de faible surface (comme la Baltique) pour lesquels les sources des agents polluants sont largement controversées, notamment lorsque plusieurs pays en sont riverains. L'utilisation de la méthode d'assimilation variationnelle pourrait certainement établir certains faits utiles à la dépollution.

Ce modèle numérique d'Advection-Diffusion et la procédure d'assimilation ont également été appliqués, avec un terme d'advection nul, au cas de la modélisation de l'extension de l'urbanisation dans le département du Var. Ces travaux reprenaient reprenant une étude réalisée avec une méthode simple d'interpolation par Méaille et Wald (1991). Ici, l'imagerie satellitaire fournissait en particulier les limites de l'urbanisation qui, ajoutées aux données démographiques, constituaient les observations. Ce cas n'est pas présenté dans cette thèse. D'abord parce qu'il est redondant d'un point de vue méthodologique avec le cas précédent, ensuite parce que l'apport de la procédure d'assimilation par rapport à la démarche empirique de Méaille et Wald, est très difficile à mettre en évidence, principalement à cause de la faible dynamique du phénomène, aggravée par la mauvaise qualité des observations.

Dans un deuxième temps nous avons repris, par des techniques d'assimilation variationnelle, une étude sur l'utilisation des champs de vents à la surface des océans en météorologie marine. Le projet initial répondait à une demande de l'industrie off-shore qui doit planifier ses travaux en fonction de l'état de la mer. Les champs de vents déduits d'observations satellitaires de l'océan Atlantique servent à initialiser un modèle régional de houle. Cette cartographie des champs de vents est rendue possible par le lancement en 1991 du satellite européen ERS-1 qui a emporté à son bord un diffusiomètre permettant de mesurer les vents. Grâce au réseau de distribution de l'agence spatiale européenne, les observations sont disponibles au sol une heure seulement après qu'elles aient été enregistrées par le satellite, permettant une assimilation de ces observations en temps réel. L'étude initiale utilisait un filtrage de Kalman simplifié et il a été montré que la précision des champs obtenus est satisfaisante pour l'utilisation finale qui en est faite (Cauneau et Chossat 1991).

L'approche variationnelle que nous développons sur ce sujet nécessite l'écriture d'un modèle numérique reflétant suffisamment la physique des phénomènes intervenants. Mais il doit rester d'une complexité assez faible, compatible avec les exigences de l'application industrielle et du temps réel. Cela nous a conduit à développer un modèle barotrope non linéaire basé sur l'Equation de la Vorticité. Le calcul du modèle adjoint et surtout son intégration introduit des problèmes numériques spécifiques à la non linéarité. Ces problèmes, ensemble avec les observations et les résultats de l'assimilation, sont développés et commentés dans le chapitre 6.

Dans la conclusion (chapitre 7) nous nous efforçons d'évaluer l'apport des méthodes d'assimilation variationnelle en télédétection, notamment par rapport au coût informatique (développement numérique et temps de calcul) qu'elles imposent.

**Chapitre 2 :**

**THEORIE DE L'ESTIMATION ET ASSIMILATION DE DONNEES**

L'assimilation de données est la procédure par laquelle toute l'information disponible (observations et relations physiques ou statistiques) est traitée afin de produire une description aussi précise et aussi complète que possible de l'état d'un système géophysique. Cette description consiste en la donnée, en des points régulièrement répartis sur le système (comme les sommets d'une grille), de l'évolution temporelle des champs physiques qui le caractérisent. Pour cela la procédure doit prendre en compte de façon quantitative la diversité de l'information et, comme ni les observations ni les relations ne sont absolument exactes, elle doit de plus évaluer la précision avec laquelle les champs physiques sont estimés. L'approche statistique du problème conduit au filtre de Kalman alors que l'approche déterministe débouche sur les techniques variationnelles. Pourtant dans les deux cas, tout au moins pour les problèmes linéaires, l'outil final utilisé est identique, il s'agit de l'estimateur linéaire des moindres carrés (Talagrand 1992). La théorie de l'estimation statistique linéaire nous permet dans ce chapitre de définir cet estimateur sous sa forme la plus générale (Lorenç 1986). Nous verrons ensuite comment grâce aux équations du filtre de Kalman cet estimateur peut être calculé de façon récursive permettant aux prévisionnistes de réaliser séquentiellement assimilation et prédiction en temps réel. Toutefois nous constaterons que les méthodes d'assimilation séquentielle sont peu adaptées au problème particulier que nous nous posons et c'est pourquoi nous ferons appel aux méthodes variationnelles. En effet, ces dernières, par des considérations déterministes, permettent de retrouver l'estimateur des moindres carrés (Wahba 1982).

## 2.1 L'estimateur linéaire des moindres carrés

La plupart des algorithmes d'assimilation qui ont été utilisés jusqu'à présent sont des formes plus ou moins évoluées de l'estimateur linéaire des moindres carrés (Talagrand 1992). L'estimation des moindres carrés consiste à chercher la combinaison linéaire des observations qui, en moyenne, approchera le mieux les champs physiques à estimer ; le degré d'approximation étant mesuré par la variance de l'estimation. L'usage courant de cet estimateur s'explique par la simplicité

de sa conception et ses bonnes propriétés. En effet, par son principe il est naturellement de variance minimale, ce qui en fait un estimateur linéaire optimal. On montre de plus que sous certaines hypothèses qu'il est efficace, c'est à dire qu'il n'existe pas d'estimateur (linéaire ou non linéaire) de variance inférieure.

La principale hypothèse à vérifier est que la densité de probabilité des erreurs sur l'information soit gaussienne. Hypothèse qui semble être vérifiée pour les modèles numériques, du moins en météorologie (Ghil and Manalotte-Rizzoli 1991, Balgovind *et al.* 1983), mais qui est beaucoup moins évidente pour les observations. En particulier en télédétection, le mouvement du satellite, ainsi que les procédures d'inversion, ont pour effet de corréler les observations entre elles.

Néanmoins, l'estimateur des moindres carrés conservent de bonnes propriétés statistiques, même lorsque les bruits ne sont pas gaussiens. Pour mettre en évidence sa simplicité nous allons l'introduire dans le cas particulier de l'analyse, c'est à dire lorsque toutes les observations sont disponibles au même instant. Puis, en introduisant un modèle numérique, nous généraliserons son application à l'assimilation de données.

Le problème de l'analyse consiste à vouloir estimer un vecteur de dimension  $n$  et de composantes  $x_i$ ,  $i = 1, \dots, n$ . Ce vecteur, dit *vecteur d'état* car il décrit l'état du système, n'est pas directement observable, mais au travers d'un opérateur linéaire  $C$ . Nous supposons donc que nous disposons d'un vecteur, de dimension  $m$  et de composantes  $y_i$ ,  $i = 1, \dots, m$ , regroupant toutes les observations effectuées sur le système à un même instant. Nous obtenons alors la relation :

$$y = C x + \epsilon$$

$C$  est une matrice  $m \times n$ , les  $m$  composantes du vecteur  $\epsilon$  sont les erreurs de mesures qui affectent chaque observation. Ces erreurs sont, bien sûr, inconnues, mais nous supposons qu'elles sont non

biaisées (sinon il suffit de retrancher le biais supposé connu), et que nous connaissons leur matrice de variance-covariance  $R$ , soit :

$$E(\varepsilon) = 0$$

$$E(\varepsilon \varepsilon^t) = R$$

où  $E$  est l'espérance mathématique et l'exposant  $t$  dénote l'opérateur de transposition.

Pour résumer, les quantités connues sont le vecteur d'observation  $y$  et la matrice d'observation  $C$ , alors que les erreurs d'observation ne sont connues qu'à travers leurs propriétés statistiques d'ordre 1 et 2 (biais nul et matrice de variance-covariance  $R$ ), le vecteur  $x$  doit être estimé. Pour cela nous recherchons un vecteur  $x'$ , qui soit une combinaison linéaire du vecteur  $x$ , et donc sous la la forme :

$$x' = Ay$$

où  $A$  est une matrice  $n \times m$  à déterminer. L'estimateur  $x'$  sera non biaisé si  $E(x-x') = 0$ , ce qui conduit à la relation matricielle :

$$A C = I_n$$

où  $I_n$  est la matrice identité de dimension  $n$ . L'estimateur sera de variance minimale, si, de plus, nous retenons la matrice  $A$  qui permettra de minimiser la quantité  $E[(x-x')(x-x')^t]$ . La solution de ce problème est :

$$A = [C R^{-1} C]^{-1} C R^{-1}$$

Les cas particuliers où  $R$  n'est pas inversible peuvent être contournés en réduisant la taille de l'estimateur. L'estimateur linéaire des moindres carrés est donc :

$$x' = [C R^{-1} C]^1 C R^{-1} y$$

et la matrice  $P$  de variance-covariance correspondante à l'erreur d'estimation est :

$$P = [C R^{-1} C]^1$$

Les éléments de  $P$  représente les covariances et, sur sa diagonale, les variances des erreurs qui sont commises sur chaque composante, lorsqu'on remplace  $x$  par  $x'$ .

L'estimateur linéaire des moindres carrés remplit complètement les charges qui lui sont assignées par l'analyse, puisque, en addition de l'estimée, il fournit la matrice  $P$  de variance-covariance de l'erreur associée à cette estimation. Remarquons que le calcul de la matrice  $P$  ne dépend pas des observations  $y$ , mais seulement de l'opérateur d'observation  $C$  et de l'erreur de mesure  $R$ . Cela signifie que pour analyser les observations au moyen de l'estimation linéaire des moindres carrés, il faut savoir exactement ce qui a été réellement observé d'utile pour l'estimation (matrice  $C$ ) et avec quelle précision les observations ont été mesurées (matrice  $R$ ).

Toutefois, l'assimilation de données est un problème d'estimation particulier, car les observations sont distribuées sur une période temporelle pendant laquelle l'état du système peut varier. La prise en compte de la dimension temporelle dans l'estimateur conduit au filtre de Kalman, nous allons voir qu'il est capable de traiter les observations séquentiellement au fur et à mesure de leur disponibilité.

## 2.2 Forme séquentielle de l'assimilation de données : le filtre de Kalman

Bien que certaines des équations précédentes peuvent être reprises, par souci de clarté, nous redonnons toutes les équations nécessaires à l'établissement du filtre de Kalman. Toutefois, nous avons repris des notations similaires, lorsque des analogies pouvaient être faites avec le paragraphe précédent. Les formules données ici sont celles du filtre de Kalman dans le cas précis de l'assimilation de données. Des compléments et des applications peuvent être trouvés dans Ghil and Manalotte-Rizzoli (1991) ou Cohn and Parrish (1991). Aux relations de l'estimation statistique linéaire, le filtrage de Kalman ajoute l'équation d'évolution du système qui est exprimée sous forme d'un modèle numérique linéaire  $S$  :

$$x_{k+1} = S x_k + b_k \quad 2.1$$

$x_k$  dénote l'état du système à l'instant  $k$ ,  $S$  est une matrice de transition exprimant la variation du vecteur d'état  $x$ , de dimension  $n$ , entre les instants  $k$  et  $k+1$ , et  $b_k$  représente les phénomènes physiques non modélisés par  $S$ . Le terme  $b_k$  est appelé erreur du modèle et est considéré généralement comme un vecteur aléatoire non biaisé et de matrice de variance-covariance  $Q$ . Comme nous supposons que ces erreurs sur le modèle sont non corrélées et indépendantes du passé du modèle, nous obtenons leurs propriétés statistiques à l'ordre 1 et 2 :

$$E(b_k) = 0 \quad 2.2 \text{ a}$$

$$E(b_k b_l^t) = Q \quad 2.2 \text{ b}$$

comme précédemment  $E$  est l'espérance mathématique et  $^t$  dénote l'opération de transposition.  $Q$  est la matrice de variance-covariance de l'erreur de modélisation. En particulier sa diagonale de  $Q$  est constituée des variances de cette erreur. Ici nous l'avons prise constante, mais cette matrice peut aussi varier dans le temps.

Nous supposons maintenant qu'il existe un opérateur linéaire  $C$  reliant les variables  $x$  du systèmes aux observations  $y$  de dimension  $m$ , et que ces observations soient disponibles aux instants  $k = 1, \dots, N$ , soit :

$$y_k = C x_k + \varepsilon_k \quad 2.3$$

Ici  $\varepsilon_k$  est une séquence de bruit blanc similaire à la séquence  $b_k$  :

$$E(\varepsilon_k) = 0 \quad 2.4 \text{ a}$$

$$E(\varepsilon_k \varepsilon_l^t) = R \quad 2.4 \text{ b}$$

où  $R$  est la matrice de variance-covariance de l'erreur d'observation, sa structure est similaire à celle de  $Q$ .

Le principe du filtrage de Kalman est d'utiliser de façon récurrente les équations de l'estimateur linéaire des moindres carrés. Supposons à l'instant (étape)  $k$  nous disposons d'une estimation, noté  $x_{k/k}$ , de l'état du système, accompagnée de la matrice de variance-covariance  $P_{k/k} = E[(x_{k+1/k} - x_{k+1})(x_{k+1/k} - x_{k+1})^t]$ . Cette matrice caractérise l'écart entre le vecteur d'état exact  $x_k$  et l'estimation  $x_{k/k}$  qui en a été faite à partir des observations  $y_k$  disponibles jusqu'à l'instant  $k$ . Alors, grâce à cet estimateur  $x_{k/k}$  disponible à l'instant  $k$ , nous pouvons prédire l'état qu'occupera le système à l'instant suivant  $k+1$  par la formule suivante :

$$x_{k+1/k} = S x_{k/k} \quad 2.5$$

Cette prédiction de l'état  $k+1$  a été notée  $x_{k+1/k}$ , car elle correspond à la prévision de l'état du système à l'instant  $k+1$  avant que les observations de l'instant  $k+1$  aient été prises en compte. L'erreur correspondante à la prévision est alors caractérisée par la matrice  $P_{k+1/k} = E[(x_{k+1/k} - x_{k+1})(x_{k+1/k} - x_{k+1})^t]$ . Après utilisation des hypothèses statistiques cette matrice vaut :

$$P_{k+1/k} = S P_{k/k} S^t + Q \quad 2.6$$

Le premier terme du membre de droite représente la propagation à l'instant  $k+1$  de l'erreur existant à l'instant  $k$ , tandis que le second terme est la contribution à l'erreur apportée par le modèle entre les instants  $k$  et  $k+1$ .

Lorsque les observations  $y_{k+1}$  deviennent disponibles, le prédicteur  $x_{k/k+1}$  de l'état  $k+1$  peut être corrigé, en recherchant comme estimateur une combinaison linéaire entre le prédicteur de l'instant  $k+1$  et les nouvelles observations. On recherche donc l'estimateur  $x_{k+1/k+1}$  de la forme :

$$x_{k+1/k+1} = x_{k+1/k} + K_{k+1} (y_{k+1} - C x_{k+1/k}) \quad 2.7$$

Cette dernière ligne met en évidence l'idée du filtre de Kalman. La prédiction  $x_{k+1/k}$  à l'instant  $k+1$  est corrigée par l'*innovation*  $y_{k+1} - C x_{k+1/k}$ , c'est-à-dire la différence entre l'état prévu et l'état observé. Cette correction se fait au moyen d'un coefficient de pondération  $K$  appelé *gain du filtre*. Dans notre cas,  $K$  est une matrice de dimension  $n \times m$ . Elle est évaluée à chaque étape de façon à réduire l'erreur d'estimation. Or la matrice de variance-covariance  $P_{k+1/k+1}$  associée à cette erreur vaut :

$$P_{k+1/k+1} = (I_n - K_{k+1} C) P_{k+1/k} (I_n - K_{k+1} C)^t + K_{k+1} R K_{k+1}^t \quad 2.8$$

ou  $I_n$  est la matrice identité  $n \times n$ . Le filtre sera dit optimal si le principe d'orthogonalité est respecté, c'est à dire si l'erreur d'estimation  $x_{k+1/k+1}$  est orthogonale à tous les observations  $y_j$   $0 \leq j \leq k+1$  (Larminat & Thomas 1977). Cette condition n'est remplie que pour une valeur unique du gain  $K_{k+1}$ , qui est :

$$K_{k+1} = P_{k+1/k} C^t (C P_{k+1/k} C^t + R)^{-1} \quad 2.9$$

Si, à l'instant  $k+1$ , il n'y a pas d'observation, alors  $C_{k+1} = 0$ , donc  $K_{k+1} = 0$  et l'état prédit  $x_{k+1/k}$  sert d'estimation  $x_{k+1/k+1}$ . Notons au passage que la valeur optimale du gain est aussi celle qui minimise la trace de la matrice de variance covariance  $P_{k+1/k+1}$ .

### *Algorithme du filtre de Kalman*

En reportant la valeur du gain 2.9 dans les expressions 2.7 et 2.8, nous obtenons un algorithme en deux étapes :

Etape de prédiction :

$$x_{k+1/k} = S x_{k/k} \quad 2.10$$

$$P_{k+1/k} = S P_{k/k} S^t + Q \quad 2.11$$

Etape d'estimation :

$$K_{k+1} = P_{k+1/k} H_k^t (C P_{k+1/k} C^t + R)^{-1} \quad 2.12$$

$$P_{k+1/k+1} = (I - K_{k+1}) P_{k+1/k} \quad 2.13$$

$$x_{k+1/k+1} = x_{k+1/k} + K_{k+1} (y_{k+1} - C x_{k+1/k}) \quad 2.14$$

Pour initialiser le filtre, nous supposons donné l'état initial  $x_0$ , ainsi que la covariance  $P_0$  de l'incertitude qui l'accompagne. Si la première observation a lieu à  $k = 0$ , alors nous poserons  $x_{0/-1} = x_0$  et  $P_{0/-1} = P_0$ . Si la première observation a lieu à  $k = 1$ , alors nous poserons  $x_{0/0} = x_0$  et  $P_{0/0} = P_0$ .

## 2.3 Commentaires

Cet algorithme met en évidence la forme séquentielle du filtrage de Kalman. A partir de l'état initial, les états du système sont successivement corrigés par les observations introduisant

ainsi graduellement, et de façon contrôlée, l'information dans le modèle. Pour comprendre comment cette assimilation s'effectue, il peut être utile de réécrire le gain de Kalman, ainsi que la matrice de variance-covariance, sous des formes plus compactes :

$$K_{k+1} = P_{k+1/k+1} C^t R^{-1} \quad 2.15$$

$$(P_{k+1/k+1})^{-1} = (P_{k+1/k})^{-1} + C^t R^{-1} C \quad 2.16$$

Il devient alors clair que le poids donné aux observations est inversement proportionnel à leur variance, alors que la précision de l'analyse est la somme de la précision sur l'état prévu calculé à partir des observations passées et de celle des observations courantes.

Remarquons d'abord que l'algorithme est basé sur une récurrence matricielle qui permet de calculer l'erreur associée. Normalement le nombre d'opérations élémentaires requis pour obtenir les matrices de variance-covariance est de l'ordre de  $n^2$ , si  $n$  est la dimension du vecteur d'état. Or, dans les applications météorologiques ou océanographiques, la taille des vecteurs d'état est de l'ordre de  $10^6$ - $10^7$  et va certainement continuer à croître, ce qui dépasse de loin les capacités de calcul actuelles ou prévues pour les dix prochaines années (Ghil and Malanotte-Rizzoli 1991). Dans les utilisations météorologiques opérationnelles, des artifices de calcul (exploitation des matrices symétriques ou creuses) et des simplifications (modélisation au préalable des corrélations) permettent d'aborder des formes dégradées du filtre de Kalman qui ne sont plus optimales, mais quand même réalistes. L'mise en oeuvre du filtre de Kalman relève donc d'un compromis entre les moyens de calcul disponibles et le degré d'optimalité désiré.

Nous remarquerons aussi que pour établir les équations du filtre de Kalman, nous avons dû faire l'hypothèse que le modèle était linéaire. Dans la réalité, les modèles ne sont pas toujours linéaires. Il existe plusieurs extensions du filtre aux modèles non-linéaires, la plus répandue étant le *filtre de Kalman étendu*, qui opèrent par linéarisations successives de la dynamique non-linéaire. Des expérimentations (Budgell 1986) menées en météorologie, ont montré que, malgré la perte

d'optimalité, le filtre de Kalman étendu avait un bon comportement. Néanmoins le problème de la non-linéarité demeure important dans la mesure où il n'est pas résolu mathématiquement. L'hypothèse de la linéarité n'est, par contre, pas requise dans la formulation de l'assimilation variationnelle, ce qui, d'un point de vue théorique, semble plus satisfaisant, même si, dans les applications pratiques non linéaires, nous retrouvons les mêmes difficultés.

La forme récurrente du filtre de Kalman offre la possibilité de fournir des prévisions en temps réel. L'information est introduite au fur et à mesure qu'elle est disponible et est utilisée pour calculer l'état du système à l'instant immédiatement suivant. Cette information apportée par les observations s'accumule dans le modèle en améliorant au fil du temps les estimations. L'inconvénient de la méthode est, qu'en début de la période d'assimilation, les estimations sont faites avec très peu d'observations, en tous cas beaucoup moins que celles réalisées en fin de la période. En particulier le filtre de Kalman est incapable d'estimer l'état initial du système au démarrage de l'assimilation. Il faut donc faire appel à une autre source d'information qui pourra lui fournir un état initial. Evidemment, si cet état initial n'est pas très réaliste, l'apport d'observations le corrigera au cours du temps ; néanmoins les premières estimations en seront affectées. En ce sens nous pouvons nous demander dans quelle mesure la qualité des estimations est sacrifiée à la récursivité dans le filtre de Kalman.

Ce trait du filtre de Kalman est particulièrement mis en exergue dans certaines études dont le but n'est pas la prévision immédiate. C'est le cas notamment de la climatologie et c'est aussi le cas d'une des études que nous présentons. Alors le problème de l'assimilation de données peut être abordé d'une façon différente : nous disposons d'un jeu d'observations distribuées sur une période de temps  $[0, T]$  déjà écoulée, et nous voulons reconstruire l'évolution du système sur toute la période. Dans le filtrage de Kalman, une observation faite à l'instant  $t$  de la période va être utilisée pour calculer tous les futurs états du système de la période  $[t, T]$ . Pourtant, compte tenu du type d'équations (différentielles) impliquées dans les modèles numériques géophysiques, il est clair que l'état du système à l'instant  $T$  dépend de l'état à l'instant initial. En d'autres termes, les

observations qui sont effectuées à l'instant final, ou à tout autre instant  $t$ , contiennent aussi de l'information sur les états occupés par le système aux instants précédents. Cette information n'est pourtant pas exploitée par le filtre de Kalman. Il existe certaines techniques d'assimilation capables de reprendre ce genre d'information et de la redistribuer dans le temps. En particulier ces techniques peuvent estimer l'état initial du système du début de la période d'observation. Il va de soi que, contrairement au filtre de Kalman, les estimations de ce type ne peuvent pas être effectuées l'une après l'autre au cours du temps, mais sont réalisées en bloc à la fin de la période d'assimilation. Toutefois nous allons voir que l'une de ces techniques peut être déduite des équations du filtrage.

#### 2.4 Le lissage de Kalman

Il existe une approche statistique capable de redistribuer l'information sur toute la période d'observation, c'est le *lissage de Kalman*. En fait, comme l'indique son nom, le lisseur de Kalman relève aussi de la théorie générale du filtrage de Kalman. La différence avec l'estimateur que nous avons présenté dans le paragraphe 2.2 (qui, dans la théorie de l'estimation, s'appelle en fait *prédicteur* de Kalman), réside dans le fait qu'il ne recherche plus à prédire l'état du système mais à estimer les états passés et présents.

Pour saisir cette différence, il est nécessaire de reprendre le problème générale d'estimation auquel s'adresse le filtrage de Kalman. Ce problème est le suivant (en reprenant les notations du paragraphe 2.2). A partir de la matrice constituée de tous les vecteurs d'observation disponibles à l'instant  $k$  :

$$Y_k^t = [y_1, y_2, \dots, y_k] \quad 2.17$$

il s'agit d'estimer la matrice inconnue :

$$X_k^t = [x_1, x_2, \dots, x_k, x_{k+1}] \quad 2.18$$

Dans le paragraphe 2.2, nous n'avons pas cherché à estimer toutes les composantes (tous les états) du vecteur  $X_k^t$ , mais uniquement les deux dernière composantes correspondant à l'état présent  $x_k$  et l'état futur  $x_{k+1}$ , ce qui allège considérablement les calculs. Pourtant les données du problème (modèles, matrices de covariance, etc.) permettent en théorie de calculer les espérances, variances et covariances de  $X_k$  et  $Y_k$ , d'où la possibilité d'appliquer les formules générales de l'estimateur statistique linéaire des moindres carrés, et de calculer toutes les composantes de  $X_k$ . Ces calculs sont effectués par le lisseur de Kalman, qui à chaque instant  $t_k$ , estime toutes les composantes de  $X_k^t$ , d'où son intérêt pour l'assimilation de données lorsque l'instant  $t_k$  correspond à l'instant final  $T$ . Cependant la construction de l'algorithme du lissage de Kalman est très complexe. Il recalcule les états du système à partir des états estimés par le passage au préalable du filtre de Kalman. Cela nécessite donc d'abord la mise en place du filtre puis ensuite celle du lisseur et c'est pourquoi il ne donne lieu qu'à très peu d'applications pratiques. Nous donnons cet algorithme sans démonstration (Wunsch 1988).

#### *Algorithme du lisseur de Kalman*

Nous supposons que les états  $x_{k/k}$  ont été estimés préalablement par le filtre de Kalman du paragraphe 2.2. En notant  $x'_{k+1}$  l'état estimé à l'instant  $k+1$  par le lisseur de Kalman et  $P'_{k+1}$  la matrice de variance-covariance de l'erreur accompagnant le lissage, nous déduisons à l'instant  $k$  l'estimation  $x'_k$  et la matrice de covariance  $P'_k$  :

$$x'_k = x_{k/k} + J_k (x'_{k+1} - S x_{k/k}) \quad 2.19$$

$$P'_k = P_{k/k} + J_k (P'_{k+1} - P_{k+1/k}) J_k^t \quad 2.20$$

$$J_k = [P_{k/k} S^t P'_{k+1/k}]^{-1} \quad 2.21$$

La base du calcul du lisseur de Kalman repose sur la comparaison entre l'état  $x_{k/k}$ , estimé lors du passage du filtre de Kalman, et une valeur  $x'_{k+1}$  calculée de façon rétrograde à partir des observations ultérieures à l'instant  $k+1$ . Cette comparaison est ensuite pondérée pour corriger l'état

$x_{k/k}$  qui avait été estimé à l'instant  $k$  par le filtre de Kalman. Les observations n'apparaissent pas dans les équations précédentes, car l'information qu'elles apportaient a déjà été extraite lors du passage du filtre de Kalman. Notons aussi l'apparition de la matrice  $S^t$ , transposée du modèle  $S$ . Nous retrouverons cette matrice qui est l'opérateur fondamental de transport de l'information dans le sens rétrograde.

Pour des problèmes non opérationnels, le lisseur de Kalman est certainement la méthode d'assimilation de données la plus satisfaisante. A chaque instant, elle estime les champs physiques en utilisant toutes les observations disponibles (passées, présentes et futures) de la période d'observation et fournit l'erreur d'estimation qui accompagne le calcul de ces champs. Néanmoins le lisseur de Kalman n'a fait l'objet jusqu'à présent que de développements théoriques (Wunsch 1988). Il semble que de bonnes performances pourraient effectivement être attendues dans le domaine de l'assimilation de données, mais les équations mathématiques sont très lourdes à manipuler, et surtout, les perspectives de son utilisation sont pour le moment très limitées par les coût de calcul qu'il requiert.

## 2.5 Dualité entre l'estimation statistique et le contrôle optimal

Il existe cependant un espoir de trouver des algorithmes qui puissent apporter des estimations proches des solutions du lisseur de Kalman, mais à un coût de calcul plus réduit. Ces algorithmes sont basés sur la *dualité* entre les estimations de Kalman et certains problèmes de contrôle optimal. Nous n'allons pas nous étendre sur la théorie de la dualité (Kalman 1960), mais nous retiendrons qu'à chaque problème d'estimation statistique linéaire correspond un problème de contrôle déterministe (Ghil and Malanotte-Rizzoli 1991, Wahba 1982). Ainsi les problèmes d'estimation statistique peuvent être résolus par la minimisation d'une certaine fonctionnelle déterministe.

La dualité nous permet alors de ramener le problème de l'assimilation de données à un problème de minimisation. Bien qu'en théorie le problème dual de l'assimilation variationnelle ne soit pas exactement le filtre de Kalman mais plutôt le lisseur de Kalman, il est possible de saisir cette notion sur un exemple simple.

Si nous reprenons le cas de l'analyse décrite dans le paragraphe 2.1, il est possible de montrer assez simplement, que l'estimateur linéaire des moindres carrés, qui y a été obtenu, est en tout point similaire au vecteur  $x'$  solution de la minimisation de la fonctionnelle suivante :

$$J = 1/2 (y - C x)^t R^{-1} (y - C x) \quad 2.22$$

Nous pouvons mieux apprécier cette similarité entre l'estimation statistique et la minimisation d'une fonctionnelle quadratique, si nous nous rappelons que le filtre de Kalman est l'estimateur qui minimise la trace de la matrice de variance-covariance  $P_{k/k}$ . Or la diagonale de cette matrice est justement constituée des carrés des erreurs d'estimation.

Dans le cas de l'assimilation de données, la fonctionnelle  $J$  précédemment citée prend la forme suivante :

$$J = 1/2 \int_{[0,T]} (y_k - C x_k)^t R^{-1} (y_{k+1} - C x_k) \quad 2.23$$

Si nous considérons  $J$  comme une fonction implicite de l'état initial  $x_0$  à travers le modèle  $S$ , et notons  $\nabla J$  le gradient de  $J$  par rapport à l'état initial, alors la matrice Hessienne  $J''$  est égale au produit scalaire  $J'' = \langle \nabla J, \nabla J \rangle$ . Il est alors possible de montrer (Press *et al.* 1990, chapitre 14.3, et Gauthier 1992 pour les applications météorologiques) que l'inverse  $J''^{-1}$  de cette matrice est égale à la matrice de covariance caractérisant l'erreur qui accompagne la donnée de l'état initial  $x_0$ , solution de la minimisation par rapport à la valeur exacte  $x_0^v$  :

$$J''^{-1} = E[(x_0 - x_0^v)(x_0 - x_0^v)^t] \quad 2.24$$

Ce résultat permet de déduire une méthode d'estimation de la matrice Hessienne (Rabier et Courtier 1992), matrice dont la connaissance peut être très utile pour le préconditionnement du problème lors de la minimisation.

Il reste à savoir maintenant s'il est effectivement plus facile de minimiser la fonctionnelle définie par 2.23 que de calculer les équations du lissage de Kalman. La réponse est apportée par les méthodes variationnelles. En effet ces méthodes permettent de résoudre certains problèmes de contrôle optimal qui, bien que n'étant pas exactement des problèmes duaux du lissage de Kalman, ont des applications importantes dans l'assimilation de données. Le développement de ces techniques de contrôle optimal a donné naissance à l'assimilation de données variationnelle.

Les méthodes variationnelles font depuis quelque temps l'objet d'actives recherches en météorologie ainsi qu'en océanographie. En effet l'intérêt qu'elles présentaient initialement en assimilation de données pour des études postérieures (climatologie) s'amplifie par la perspective que pourraient offrir les prochains super-calculateurs de les utiliser en prévision numérique opérationnelle.

## 2.6. Choix d'une procédure d'assimilation

Dans le cas de l'étude sur le modèle d'Advection-Diffusion, le choix de la procédure d'assimilation ne se pose pas, puisqu'il n'y a que les méthodes variationnelles qui permettent de retrouver l'état initial, et c'est même une de leurs caractéristiques.

En ce qui concerne la cartographie des champs de vent à la surface des océans, nous voulons à la fois pouvoir faire des prévisions à 12 heures de l'état de la surface de la mer et

produire des cartes suffisamment fiables pour pouvoir être reprises, principalement comme "forcing", dans des études numériques postérieures d'océanographie et de génie côtier. D'autre part les modèles que nous utilisons sont de taille raisonnable, tout au moins si nous les comparons aux modèles de circulation atmosphérique, si bien que les coûts de calcul ne sont pas *a priori* une contrainte majeure, dans la mesure bien sûr où ils restent réalisables. Notre préoccupation est donc surtout la qualité des champs analysés. La procédure doit être choisie de façon à satisfaire un certain cahier des charges, qui peut être résumé de la façon suivante : champs de vent homogènes couvrant tout l'océan Atlantique nord sur une période de 12 heures, assimilation en temps réel sur station de travail, analyse à des temps intermédiaires quelconques les plus fiables possibles et calcul des prévisions à 12 heures. Des premiers travaux (Cauneau and Chossat 1991), utilisant l'assimilation séquentielle, ont permis de montrer que l'ébauche calculée à la fin de la période d'assimilation avec les ressources informatiques d'une station de travail était acceptable comme prévision, mais que la fiabilité des analyses obtenues à des temps intermédiaires de la période d'assimilation, laissait à désirer, surtout en début de période. Ces travaux confirment le fait que, d'une part, les ressources informatiques sont suffisantes pour supporter une procédure d'assimilation variationnelle qui est d'un coût semblable, d'autre part, que le filtre de Kalman qui est très adapté pour la prévision numérique, n'est néanmoins pas optimal dans le sens où il ne propage l'information que du passé vers le futur, affectant ainsi la qualité des premières estimations.

Une sensible amélioration peut donc être espérée par l'emploi d'une procédure possédant la faculté de propager l'information dans le sens rétrograde sans pour autant altérer les prévisions. Le lisseur de Kalman possède cette qualité, mais ses mécanismes ne sont pas encore assez bien dominés pour être utilisables. C'est pourquoi nous nous sommes tournés vers les méthodes variationnelles qui, tout en assurant une bonne répartition temporelle de l'information, disposent d'un formalisme mathématique bien établi, facilitant ainsi leur mise en oeuvre.

**Chapitre 3 :**

**ASSIMILATION DE DONNEES VARIATIONNELLE  
ET EQUATIONS ADJOINTES**

L'intérêt de l'approche statistique dans l'assimilation de données réside dans sa capacité à prendre en compte les incertitudes qui accompagnent les différentes formes d'information qui lui sont soumises. En effet, sauf cas particuliers, les observations ne sont jamais parfaites et ne vérifient pas en général les équations du modèle. Aussi les méthodes statistiques fournissent une solution qui est un compromis entre ces deux sources d'information, l'influence de chacune des sources étant pondérée par la fiabilité de l'information qu'elle apporte. Malheureusement le calcul de la solution n'est pas en pratique réalisable et il est nécessaire de procéder à des simplifications.

Une simplification possible du problème de l'assimilation peut être alors de réduire le nombre de champs candidats à la solution du problème. Cela peut être effectivement réalisé si nous remarquons que les champs résultants de l'assimilation statistique sont différents des observations, tout en ne vérifiant pas les équations du modèle. Une réduction pourrait donc consister à rechercher des solutions qui vérifient les équations du modèle, et donc de faire l'hypothèse que le modèle est parfait. Dans ce cas nous rechercherons parmi tous les champs physiques solutions du modèle, ceux qui approchent au mieux les observations. Cet ajustement du modèle aux observations conduit à une formulation variationnelle du problème de l'assimilation de données.

Ce chapitre présente ce formalisme, puis nous montre comment des techniques de contrôle optimal sont à même de résoudre le problème variationnel qui en découle, et enfin, permettent de déduire un algorithme d'assimilation de données.

### 3.1 Formulation variationnelle du problème de l'assimilation de données

Pour définir mathématiquement la phrase "approcher au mieux les observations", nous pouvons construire une fonctionnelle qui évalue de façon déterministe l'écart entre les observations et les solutions du modèle. Cette fonctionnelle peut être tout simplement le carré de la norme de l'écart entre les observations (vecteur  $Y$ ) et la solution du modèle (vecteur  $X$ ) projetée dans l'espace

des observables (opérateur C). En reprenant les notations du chapitre 2, nous obtiendrons, par exemple, la fonctionnelle suivante :

$$J(X, Y) = \| Y - CX \|^2 \quad 3.1$$

Parce que J pénalise la différence entre le modèle et les observations, elle est souvent appelée fonction de coût. Cette définition 3.1 porte la forte empreinte de l'approche statistique. En effet il suffit de définir le produit scalaire dans l'espace des observations par  $\langle CX, Y \rangle = (CX)^t R^{-1} Y$ , (comme au chapitre 2, R est la matrice de variance-covariance des observations et  $^t$  l'opérateur de transposition) et nous aurons réintroduit les erreurs statistiques dans ce problème déterministe. Ceci est directement lié aux aspects de la *dualité* mentionnés précédemment (§ 2.5). En particulier l'usage de fonctions de coût quadratique permet de retrouver des formes de l'estimateur des moindres carrés.

### 3.1.1 Formulation variationnelle : minimisation avec contrainte

Lorsque la fonction de coût est définie, l'assimilation de donnée consiste à chercher des champs physiques (sous forme d'un vecteur X), qui minimisent la fonction de coût sous la contrainte de la vérifications des équations du modèle. Ce problème d'optimisation peut être résumé dans la formulation variationnelle suivante :

$$\text{Trouver } X', \text{ tel que } J(X') = \min J(X) \quad X \in \mathcal{X} \quad 3.2$$

$$\text{et } F(X') = 0 \quad 3.3$$

Ici, J est la fonction de coût mesurant l'écart entre les observations et les champs analysés, F représente les équations du modèle numérique, mais peut aussi exprimer d'autres contraintes (relations physiques par exemple) imposées à la solution X', et  $\mathcal{X}$  est le domaine sur lequel la solution est recherchée. Par commodité, nous avons exprimé ces contraintes sous forme d'égalité, mais elles peuvent être aussi du type inégalité (telle une température qui doit être positive). Les

résultats de l'assimilation de données variationnelle sont donc directement liés aux choix des fonctions J et F. Il convient de donner un sens physique facilement interprétable.

### 3.1.2 Réduction des variables de contrôle

Généralement, les problèmes d'optimisation avec contrainte sont difficiles à résoudre. D'ailleurs, dans le cas particulier de l'assimilation de données, la complexité du problème mathématique 3.2-3 qui est posé, est à la mesure de son interprétation physique. Il s'agit de chercher parmi toutes les solutions du modèle possibles, celle qui, sur la période d'observation, est la plus proche des observations. La minimisation est donc à effectuer sur tout les états possibles que peut occuper le système au cours de cette période. Cela semble *a priori* difficilement réalisable, sauf si nous remarquons que, dans la mesure où tous ces états sont en fait déterminés par la seule donnée de l'état initial, il est possible d'opérer *une réduction des variables de contrôle* (Le Dimet and Talagrand 1986) et de minimiser la fonction de coût par rapport à l'état initial du système.

Nous avons mentionné ici l'état initial car c'est généralement une inconnue pour les systèmes géophysiques. Il est rarement possible d'obtenir instantanément assez d'observations pour en faire une description complète. Toutefois le contrôle peut aussi bien porter sur des paramètres inconnus du modèle, comme, par exemple, des conditions aux limites ou des forçages externes.

### 3.1.3 Calcul du gradient de la fonction de coût

Pour déterminer le minimum de la fonction de coût, nous avons besoin de calculer les gradients de la fonction de coût par rapport aux variables de contrôle. Dans le cas général, ces gradients ne peuvent pas être déterminés analytiquement. Ils doivent être évalués numériquement en déterminant les effets sur le système d'une perturbation des variables de contrôle. Hoffman (1986) a utilisé avec succès cette méthode pour analyser des champs de vent, mais elle se révèle très coûteuse en calcul car elle demande autant d'intégrations du modèle que les variables de contrôle ont de degrés de liberté.

## 3.2 Utilisation des équations adjointes

Afin de présenter la méthode des équations adjointes dans sa forme la plus générale mais aussi la plus simple, nous développerons le cas d'un système dépendant de façon continue du temps. Des résultats similaires peuvent être obtenus pour les applications numériques qui nécessitent une discrétisation en temps (Thacker and Long 1988).

### 3.2.1 *Présentation de la méthode*

Supposons un système dont les équations d'évolution aient été discrétisées par rapport aux variables d'espace, permettant ainsi de décrire sur un intervalle de temps  $[0, T]$ , l'état de ce système comme solution d'une équation différentielle ordinaire :

$$dX/dt = F(X, U) \quad 3.4$$

$$X(0) = V \quad 3.5$$

où  $X$  représente l'état du système et appartient à un espace vectoriel  $X$  de dimension  $n$ ,  $F$  est une fonction (linéaire ou non linéaire) de  $X$  dans  $X$ ,  $V$  est l'état initial, et  $U$  est un ensemble de paramètres ajustables du modèle. Nous supposons qu'à chaque couple  $(U, V)$ ,  $F$  fait correspondre une unique solution  $X(t)$  sur l'intervalle  $[0, T]$ .

Le système est observé sur la période  $[0, T]$ . Soit  $X_{\text{obs}}(t)$ , le vecteur constitué des observations. Il appartient à l'espace  $O$  de dimension  $m$ . Les espaces  $X$  et  $O$  peuvent être différents, car il existe des points d'observations qui ne coïncident pas avec les points de la grille de calcul du modèle, mais aussi parce que les variables du système peuvent ne pas être directement observables. Néanmoins, nous supposons qu'il existe un opérateur linéaire  $C$  de  $X$  dans  $O$  qui permet de comparer les observations à la solution du modèle au travers d'une fonction de coût  $J$  définie par :

$$J(U, V) = 1/2 \int_{[0, T]} H(X, t) dt \quad 3.6$$

H est une fonction de X et de t, et donc une fonction implicite de U et V. Dans le cas particulier de l'équation 3.1 où H est une fonction quadratique de l'erreur entre les observations et le modèle, J devient :

$$J(U, V) = 1/2 \int_{[0, T]} H(X, t) = 1/2 \int_{[0, T]} \| W(C X(U, V, t) - X(t)_{\text{obs}}) \|^2 dt \quad 3.7$$

Ici, nous avons choisi d'écrire explicitement dans l'expression de la fonction de coût la matrice W qui pondère les différences entre elles. Idéalement W devrait être l'inverse de la matrice R de variance-covariance des observations, et peut aussi varier dans le temps. Comme W est une matrice positive, elle peut être incorporée dans la norme  $\| \cdot \|$  qui est associée au produit scalaire, défini sur l'espace des observations  $\mathcal{O}$  par  $\| X \|^2 = \langle X, X \rangle = X^t W X$ . Dans l'équation 3.7, nous avons préféré expliciter la matrice de pondération W en utilisant la forme canonique du produit scalaire, c'est-à-dire  $\langle X, X \rangle = X^t X$ .

Le problème de contrôle induit par l'assimilation de données consiste à déterminer les valeurs  $U^*$  et  $V^*$  des variables de contrôle, telles que la solution des équations 3.4-5 minimise la fonction de coût J.  $U^*$  et  $V^*$  doivent vérifier :

$$J(U^*, V^*) \leq J(U, V) \quad \forall (U, V) \in \mathcal{V} \quad 3.8$$

$\mathcal{V}$  est l'espace des contrôles admissibles, c'est à dire l'ensemble des valeurs permises pour les variables de contrôles.

Afin de pouvoir appliquer une méthode de minimisation performante, il convient de déterminer le gradient  $J_u$  et  $J_v$  de la fonction de coût par rapport aux variables de contrôle. La difficulté du problème de minimisation 3.8 réside dans le fait que la fonction de coût n'est qu'une

fonction implicite de  $U$  et  $V$ . Pour dériver la fonction de coût définie par la formule 3.8, il convient d'appliquer la règle de dérivation composée. Ce calcul différentiel peut être réalisé d'une façon assez simple et élégante en introduisant l'état adjoint solution du modèle adjoint au système 3.4-5. Les calculs nécessaires sont donnés dans l'annexe 9.1 et nous retiendrons seulement ici les résultats utiles aux applications pratiques.

Si nous définissons l'état adjoint  $P(t)$  solution du système différentiel :

$$- dP/dt = [\partial F/\partial X]^t P + \partial H/\partial X \quad 3.9$$

$$P(T) = 0 \quad 3.10$$

ces équations différentielles sont déduites des équations différentielles 3.4-5 et sont appelées *équations adjointes* par opposition aux équations 3.4-5 qui sont appelées *équations directes*. La théorie du contrôle optimal permet d'établir des relations directes entre les gradients de la fonction de coût et l'état adjoint :

$$J_v(U, V) = P(0) \quad 3.11$$

$$J_u(U, V) = \int_{[0, T]} [\partial F/\partial U]^t P dt \quad 3.12$$

Le gradient de la fonction est ainsi obtenu par intégration rétrograde du système 3.9-10.

Pour résumer, pour un couple  $(U, V)$  les gradients de la fonction de coûts  $J_u(U, V)$  et  $J_v(U, V)$  peuvent être obtenus par les opérations suivantes :

- i) Intégrer l'état initial  $V$  par le modèle direct 3.4-5 de 0 à  $T$  ( $U$  étant fixé) et sauvegarder tous les états calculés  $X(t)$ ,  $0 \leq t \leq T$ .

- ii) A partir de  $P(T) = 0$ , intégrer le modèle adjoint de façon rétrograde dans le temps de  $T$  à  $0$ . Le modèle adjoint 3.9-10 est déterminé à chaque instant  $t$  par les valeurs de  $X(t)$  calculées et sauvegardées au cours de l'intégration directe.
- iii) La valeur finale  $P(0)$  est le gradient de  $J$  par rapport à l'état initial (variable  $V$ ).
- iv) Le calcul du terme 3.12 fournit le gradient de la fonction de coût par rapport aux paramètres de contrôle (variable  $U$ ).

### 3.2.2 Remarques:

Ici nous introduisons les équations adjointes de façon arbitraire et nous vérifions qu'elles conduisent au calcul du gradient de la fonction de coût. En fait, les états adjoints du système sont les valeurs prises par les *Multiplieurs de Lagrange* du *Lagrangien* associé au problème de minimisation 3.8 (cf. annexe 9.2). Le système adjoint est constitué des équations permettant de calculer ces multiplieurs. Ainsi, la justification de l'usage des méthodes adjointes relève d'une théorie générale bien établie (Lions 1969, Courtier 1987).

Dans les problèmes d'assimilation de données, la variable de contrôle est l'état initial  $V$  qu'il est difficile d'obtenir complètement par observation. Le contrôle par la variable  $U$  est moins courant, et seulement utilisé pour étudier la sensibilité des modèles numériques à certains de leurs paramètres d'entrée (Cacuci et Hall 1984, Courtier 1987). C'est pourquoi dans la suite, sauf mention spéciale, nous nous restreindrons au contrôle sur l'état initial, sachant que des résultats similaires peuvent être obtenus pour les autres variables de contrôle.

### 3.2.3 algorithmes de descente

Un fois le gradient de la fonction de coût obtenu il est possible d'appliquer un algorithme de descente dans l'espace des contrôles afin de déterminer la valeur  $V^*$  qui minimise la fonction de coût  $J$ . Des approximations successives  $V_n$  de  $V^*$  sont obtenues par étape de la forme :

$$V_{n+1} = V_n - \rho_n D_n$$

3.13

où pour chaque  $n$ ,  $D_n$  est la direction de descente directement calculée à partir du gradient de  $J$ , et  $\rho_n$  est un scalaire approprié. Si  $\rho_n$  et  $D_n$  sont convenablement choisis, alors la séquence  $(V)_n$  va converger vers  $V^*$ .

L'aspect itératif des méthodes variationnelles est un facteur important car il conditionne leur mise en oeuvre. En effet pour déterminer une approximation  $V_1$  de la valeur optimale du contrôle  $V$ , il faut disposer d'une valeur  $V_0$  servant de *point de départ* à la descente, puis intégrer le modèle direct et enfin intégrer le modèle adjoint. Chaque itération nécessite donc au moins autant, et même plus, de calcul que la prévision numérique toute seule. C'est pourquoi il est nécessaire de choisir un algorithme de descente capable de converger vers le minimum de la fonction de coût en peu d'itérations.

Divers algorithmes de descente ont été mis au point (Gill *et al.* 1982). Le plus simple est *l'algorithme de la plus grande pente du gradient* qui correspond au cas où la direction de descente est exactement celle du gradient ( $D_n = J_v$ ). Cet algorithme assure une convergence régulière, mais très lente voire nulle (Faurre 1986). C'est pourquoi on lui préfère souvent *l'algorithme du gradient conjugué* dont les directions de descente sont conjuguées entre elles. En effet, dans l'espace des phases, la fonction de coût a initialement une structure elliptique. Or, conjuguer les directions de descente revient à changer la métrique du problème (pré-conditionnement) et de ramener ainsi l'ellipsoïde en une sphère. Dès lors, le gradient pointe vers le minimum, alors qu'initialement, il pouvait pointer dans une direction quelconque, voire orthogonale si la fonction de coût était fortement elliptique (Thépaut and Courtier 1992). Ainsi, pour des fonctions de coût quadratiques, le gradient-conjugué assure une convergence en un nombre d'itérations au plus égal à la dimension du système.

L'intérêt de cet algorithme est qu'il peut aussi présenter de bonnes propriétés de convergence, lorsqu'il est appliqué aux fonctions de coût non quadratiques générées par les

modèles non linéaires. En effet, au voisinage du minimum, on peut remplacer la fonction de coût par son développement de Taylor au deuxième ordre, et retrouver ainsi une fonction quadratique. Ces algorithmes, dit du deuxième ordre, sont très performant car ils utilisent la *matrice Hessienne* pour re-conditionner le système à chaque pas de descente. Avec *l'algorithme de Newton*, la matrice Hessienne est explicitement calculée, alors qu'elle est seulement estimée par *l'algorithme de quasi Newton* (Navon and Legler 1987). La réduction du nombre d'itérations, et donc du temps de calcul, obtenu par le conditionnement du système est importante et même indispensable pour les gros systèmes météorologiques. La matrice Hessienne est un outil décisif dans le processus de minimisation. Malheureusement, sa détermination, sinon son estimation, n'est pas aisée.

### 3.3 Modèles non linéaires et problèmes de convergence

#### 3.3.1 *Convergence de l'algorithme de descente*

Les techniques de contrôle optimal permettent de calculer le gradient de la fonction de coût pour tous les modèle numériques dont les équations sont différentiables. Néanmoins la nullité du gradient en un point n'implique pas forcément que ce point soit le minimum, mais indique seulement un extremum local (minimum, maximum ou point selle). Il est toutefois des situations où ces points sont toujours des minimas, c'est par exemple, les cas où la fonction de coût est strictement convexe (Le Dimet *et al.* 1991). Alors la convergence des algorithmes de descente vers le minimum est assurée. Il n'est pas surprenant de trouver des restrictions à la méthode sur la forme de la fonction de coût. Elles expriment simplement, que pour trouver le minimum de cette fonction il faut d'abord, que ce minimum existe puis qu'il soit unique.

Sachant que la fonction strictement convexe la plus simple est certainement la fonction quadratique, il n'est pas surprenant de retrouver une nouvelle fois ce type de fonction dans les applications de l'assimilation de données. Néanmoins, le choix d'une fonction de coût quadratique n'assure pas l'unicité du minimum. En effet, la convexité est requise par rapport aux variables de

contrôle. Si la fonction  $F$  définie par 3.4 est linéaire en  $X$ , alors la fonction  $J$  définie par 3.7 est quadratique par rapport à  $X$  et donc par linéarité par rapport à l'état initial  $V$ . Mais si  $F$  n'est pas linéaire, alors la dépendance de  $J$  par rapport à  $V$  peut être tout à fait non quadratique.

Des expérimentations menées en météorologie (Lacarra et Talagrand 1988, Courtier et Talagrand 1990), montrent que, pour les modèles non linéaires couramment utilisés en météorologie, et sur des durées d'intégration de 24 à 48 heures, il est licite de considérer les fonctions de coût comme quadratiques. Ces résultats signifient que l'on peut localement remplacer le modèle numérique défini par 3.4-5 par un modèle linéaire qui lui est tangent, la durée de la validité de l'approximation variant de 24 à 48 heures. Nous noterons aussi que le modèle adjoint défini par les formules 3.9-10 est toujours linéaire, car ce modèle n'est pas en réalité l'adjoint du modèle direct mais *l'opérateur adjoint* de l'opérateur défini par le *modèle linéaire tangent* que nous venons de citer.

Dans tous les cas, puisque nous travaillons sur un domaine spatio-temporel limité, nous savons qu'il existe au moins un minimum à la fonction de coût. Alors le problème de convergence se posera non pas en terme d'existence du minimum, mais en terme d'unicité.

### 3.3.2 *Unicité de la solution*

Que ce soit pour les modèles linéaires ou pour les modèles non linéaires approchés par leur modèle linéaire tangent, le problème de l'unicité du minimum de la fonction de coût se pose essentiellement lorsque l'information apportée par les observations n'est pas suffisante pour reconstruire l'état initial du problème. En clair, il faut que les observations permettent de déterminer tous les degrés de liberté du modèle. Remarquons qu'il ne suffit pas seulement que le nombre d'observations soit exactement égal au nombre de degrés de liberté, il faut de plus qu'elles soient indépendantes. Car il se peut que les observations soient suffisantes en quantité mais redondantes, réduisant ainsi l'information réellement apportée. Dans ce cas, une fonction de coût, même quadratique, n'est plus strictement convexe et peut présenter plusieurs minima locaux. En fait, ces

minima forment des continuums, c'est-à-dire que le graphe de la fonction de coût ne sera pas parabolique dans toutes les directions, mais constant dans les directions où la solution ne peut être déterminée (Thacker 1988).

Avec une fonction de coût strictement convexe, les algorithmes de descente convergent vers ce minimum quelque soit son point de départ (état initial  $V_0$ ). Si la fonction de coût n'est plus convexe, alors, suivant le point de départ, la descente pourra se terminer sur l'un ou l'autre minimum local, ou même sur un point selle.

### 3.3.3 Analyse du deuxième ordre et conditionnement

Il est cependant possible de déterminer si le point stationnaire obtenu par l'algorithme de descente est un minimum. Il faut pousser l'analyse au second ordre et calculer la matrice Hessienne  $H$  de la fonction de coût. En effet, un minimum est caractérisé par un gradient nul et une matrice Hessienne définie positive. La mise en application d'une analyse du second ordre est assez complexe et peu utilisée en pratique. Nous ne mentionnons ici que certains résultats (Le Dimet *et al.* 1991).

Dans le cas linéaire, la matrice Hessienne est constante et l'étude du signe de ses valeurs propres, qui sont toutes réelles puisque la matrice est symétrique, permet de savoir si le point stationnaire qui a été trouvé est un minimum et s'il est unique. Cependant, le calcul des valeurs propres d'une matrice n'est pas une chose aisée et il existe un critère simple qui correspond à ce qui est appelé *observabilité* du système dans la théorie du contrôle optimal. Dans le cas linéaire seulement, et avec la fonction de coût définie par l'équation 3.7 ( $W = I_m$  matrice identité de dimension  $m$ ) à minimiser sous la contrainte du système d'équation 3.4-5 (de dimension  $n$ ), on dira que le système est observable si :

$$\text{rang } [C^t, A^t C^t, \dots, (A^t)^{n-1} C^t] = n$$

Alors la matrice Hessienne  $H$  sera définie positive et il y aura unicité du minimum.

Si le modèle est non linéaire, les points stationnaires ne peuvent être étudiés que localement. Le Hessien n'est plus constant et doit être calculé en chaque point (état initial possible) ; il n'est donc plus possible d'exhiber un critère algébrique d'observabilité. Ce calcul est réalisable en utilisant une nouvelle fois les équations adjointes. Il faut introduire l'opérateur adjoint de l'opérateur qui, à un état initial associe, son état adjoint. De nouveau un critère sur "l'adjoint de l'état adjoint" permet de déduire le signe des valeurs propres de la matrice Hessienne.

Nous retiendrons que l'analyse du second ordre est en générale lourde à mettre en oeuvre, mais qu'elle est, d'une part, l'unique moyen de déterminer rigoureusement la validité des solutions de l'assimilation de données variationnelle et, d'autre part, elle peut être utilisée pour la conception de système d'observation optimal (Le Dimet *et al.* 1991).

Ces résultats sur l'impact de la matrice Hessienne du système sur l'unicité de la solution sont à rapprocher de ceux obtenus dans le paragraphe 3.2.3 sur la convergence par pré-conditionnement. Le récent intérêt pour l'estimation de cette matrice n'est donc pas surprenant et a permis la mise au point d'une technique d'estimation très astucieuse (Gautier 1992, Rabier and Courtier 1992, Courtier 1992).

Il existe aussi des méthodes pour pallier systématiquement le manque d'observations et lever l'indétermination qui en découle ; nous en présentons une dans le paragraphe suivant.

#### 3.4 Détermination de la solution par pénalisation-régularisation

Les résultats de l'assimilation de données sont très sensibles à l'information qui est utilisée. Cette dépendance est à la fois quantitative, puisque l'insuffisance d'observations multiplie le

nombre de solutions possibles, mais aussi qualitative, car des observations trop bruitées apportent des solutions mathématiques non réalistes.

Dans le cas d'observations en quantité apparemment suffisante mais très bruitées, l'effet principal consiste en l'apparition dans la solution d'ondes sans signification physique, mais qui mathématiquement apparaissent dans la solution pour suivre la variabilité des observations (Courtier et Talagrand 1987, Courtier et Talagrand 1990). Pour assimiler des observations bruitées, on pourra limiter à une certaine valeur l'amplitude des variations temporelles de la solution. On pourra ainsi demander la condition  $\| dX/dt \| \leq H$ , où  $H$  est la limitation souhaitée. Dans ce cas il sera nécessaire d'utiliser une fonction de coût appropriée, de la forme :

$$J_r = J + r/2 \int_{[0, T]} \psi( \| dX/dt \| ) dt \quad 3.14$$

où  $J$  est la fonction de coût définie par 3.7,  $r$  est un coefficient positif, et  $\psi$  la fonction définie par :  $\psi(z) = 1/2 (z - H)^2$  si  $z \geq H$  et  $\psi(z) = 0$  si  $z < H$ . Il est démontré (Le Dimet *et al.* 1991) que, si  $U_r^*$  (respectivement  $U^*$ ) est la valeur optimale minimisant  $J_r$  (respectivement  $J$ ), alors  $U_r^* \rightarrow U^*$  quand  $r \rightarrow +\infty$ . Cependant pour des valeurs de  $r$  trop grandes, le problème devient mal conditionné et la convergence est très lente (Courtier et Talagrand 1990). Il existe des algorithmes (Bertsekas 1982, Le Dimet et Talagrand 1986, Ghil et Malanotte-Rizzoli 1991, Le Dimet *et al.* 1991), comme par exemple l'algorithme du *Lagrangien augmenté*, qui, en apportant d'autres modifications à la fonction de coût, permettent d'éviter ce problème de conditionnement. Néanmoins les problèmes dus au bruit d'observation peuvent être abordés de la même façon que les problèmes de manque d'information. En effet mathématiquement ils relèvent de la même classe des problèmes dit *mal posés* qui font l'objet d'études particulières en géophysique (Wahba 1987).

Concrètement l'effet de la mauvaise qualité de l'information s'observe sur les solutions par l'apparition dans les champs physiques, de petites structures qui n'ont pas de sens physique. Or, en étudiant l'évolution de la fonction de coût, certains auteurs (Wahba 1987, Courtier et Talagrand

1987) ont remarqué qu'après avoir reconstruit les principales structures physiques, accompagnées d'une forte diminution de la fonction de coût, la procédure d'assimilation développait ces petites structures, tandis que la fonction de coût ne diminuait que très peu. Une interprétation de ce phénomène peut être la suivante (Wahba 1987) : les premières itérations retrouvent la projection sur les vecteurs de bases des grandes valeurs de la solution, alors que les itérations suivantes retrouvent les faibles valeurs de la solution qui sont beaucoup plus sensibles au bruit de mesure. Un moyen d'éviter ces bruits est d'arrêter l'itération avant qu'ils n'apparaissent. Le problème est de faire la part des choses dans la solution entre ce qui a une signification physique et ce qui est du bruit. Il existe des méthodes (Wahba 1987) pour déterminer objectivement ce nombre critique d'itérations. Mais il est plus satisfaisant de s'attaquer à la cause du problème et de chercher dans la méthode d'assimilation un moyen de lever l'indétermination due aux observations.

En effet, Bennett et Miller (1990) ont montré que, pour des assimilations variationnelles utilisant des fonctions de coût de la formule 3.7, les solutions de l'assimilation n'étaient définies qu'à une fonction *non-observable* près. Si des observations viennent à manquer ou sont de mauvaise qualité, alors il pourra exister à l'échelle du modèle des structures qui ne seront pas résolues par les observations. Afin de pallier au manque d'observations, il est nécessaire d'avoir recours à d'autres sources d'information, notamment de disposer d'information *a priori* sur la nature des champs physiques recherchés.

Le type d'information portera essentiellement sur des contraintes que la solution attendue devra vérifier. Par exemple on peut supposer que les champs physiques, qui constitueront l'état initial solution de l'assimilation, auront une certaine régularité et que leurs amplitudes, ou leurs dérivées, ne varieront pas trop fortement dans l'espace. Une façon d'obtenir des solutions régulières consiste à pénaliser leur variabilité spatiale dans la fonction de coût. Cette méthode, dite de *pénalisation-régularisation*, ajoute à la fonction de coût initiale  $J$ , définie par la formule 3.7, un terme quadratique supplémentaire proportionnel à cette variabilité :

$$J_r(U) = J + r/2 \| U \|^2 \quad 3.15$$

où  $r$  est un coefficient positif. Il est clair la que valeur de  $U$  qui minimisera cette fonction de coût sera faible en amplitude. Mathématiquement, l'ajout de ce terme de pénalité suffit à lever l'indétermination. En effet, si nous regardons le Hessien de la fonction de coût  $J_r$  ainsi pénalisée, nous obtenons :

$$H_r = H + r I \quad 3.16$$

où  $H$  est le Hessien de  $J$  (forme non pénalisée) et  $I$  est la matrice identité. Puisque  $H$  et  $H_r$  ont les mêmes vecteurs propres, si  $\lambda$  est une valeur propre de  $H$ , alors  $\lambda + r$  est la valeur propre de  $H_r$  associée au même vecteur propre. En choisissant  $r$  assez grand, il devient donc possible de rendre toute les valeurs propres de  $H_r$  strictement positives et le problème a une unique solution globale dans le cas linéaire et une solution locale dans le cas non linéaire. Malheureusement, pour de faibles valeurs de  $r$ , le problème est *mal conditionné*, c'est-à-dire que la convergence vers la solution est lente, alors que pour de grandes valeurs de  $r$ , la solution obtenue tendra vers la solution nulle.

Une possibilité d'obtenir une solution unique, sans toutefois altérer la vitesse de convergence, est de pénaliser les dérivées de la solution. Ainsi nous utiliserons un terme de pénalisation que nous pouvons écrire sous la forme générale :

$$J_{\lambda,m} = J + \lambda \int_{[0, T]} \| \partial^{(m)} X \|^2 dt \quad 3.17$$

où  $\partial^{(m)}$  est la dérivé  $m$ -ième de  $X$ . Ajouter le terme de pénalisation défini ci-dessus revient à passer la solution retenue dans un filtre passe-bas, dont la fonction de transfert est (Shiew *et al.* 1986) :

$$F(\omega) = 1/[1 + \lambda(2\pi\omega)^{2m}] \quad 3.18$$

ce qui illustre l'effet régulateur de ce type de pénalisation.

Lorsque  $m$  est supérieur ou égal à 2, Wahba et Wandelberger (1980) ont proposé une méthode (*général cross validation*) qui permet de déterminer de façon optimale la valeur de  $\lambda$  et  $m$  à partir des observations. Pour notre part nous retiendront le cas où  $m = 2$  et déterminerons la valeur de  $\lambda$  par corrections successives.

### 3.5 mise en oeuvre des techniques de contrôle optimal dans une procédure variationnelle d'assimilation de données.

Dans le paragraphe 3.2, nous avons calculé le gradient de la fonction de coût d'un problème continu dans le temps. Dans la pratique, nous disposons d'un modèle numérique discret qui n'est qu'une approximation des équations continues. L'opérateur dont il faut prendre l'adjoint est l'opérateur qui sert effectivement à calculer les états du système, c'est-à-dire le modèle numérique (quand nous parlons de modèle numérique, nous sous-entendons qu'il s'agit d'un opérateur discret). En effet, suivant le schéma de discrétisation adopté, l'opérateur adjoint (discret) qui résulterait de la discrétisation des équations adjointes continues (obtenues elles-mêmes à partir des équations continues du système) peut, dans certains cas, différer du modèle adjoint qui est l'opérateur adjoint du modèle numérique direct. Il existe une caractérisation des schémas pour lesquels cette égalité est vérifiée (Talagrand and Courtier 1987). Nous retiendrons que les schémas classiques du type Euler ont cette propriété.

Néanmoins, comme nous utilisons des modèles numériques discrets, il est plus satisfaisant de raisonner directement sur les équations discrètes. En utilisant le formalisme des multiplieurs de Lagrange, il est possible de démontrer tous les résultats exposés au paragraphe 3.2 dans le cas discret (Thacker et Long 1988, Thacker 1988). Nous nous contenterons de montrer comment, avec

les équations adjointes discrètes, il est possible de mettre en oeuvre la procédure variationnelle d'assimilation de données.

### 3.5.1 Equations discrétisées du modèle direct

A chaque instant  $t_k$   $k = 0, 1, \dots, N$ , nous supposons que l'état du système étudié peut être décrit par un vecteur d'état  $X_k$  de dimension  $n$ . A l'instant  $t_{k+1}$ , l'état  $X_{k+1}$  est relié à l'état  $X_k$  de l'instant  $t_k$  par le modèle numérique :

$$X_{k+1} = X_k + \Delta t_k F_k(X_k, U) \quad k = 0, \dots, N-1 \quad 3.19$$

$$X_0 = V \quad 3.20$$

où  $\Delta t_k = t_{k+1} - t_k$  est le pas temporel d'intégration,  $F_k$  est le modèle numérique, éventuellement non linéaire, donc dépendant du temps. Dans le cas linéaire, la fonction  $F_k$  sera remplacée par un produit matriciel  $F_k X_k$ .

### 3.5.2 Fonction de coût discrète

Au cours du temps, le système est observé et nous supposons que le nombre maximal d'observations disponibles instantanément est  $m$ . A chaque instant  $t_k$  un vecteur  $X_k^{\text{obs}}$  de dimension  $m$  est défini. Ses composantes seront constituées des observations disponibles à cet instant et zéro partout ailleurs. Nous supposons de plus qu'il existe un opérateur linéaire  $C$  qui fait correspondre les observations avec l'état du système. Dans ce cas, la fonction de coût, équivalente de 3.7, est :

$$J_N = \sum_{0 \leq k < N} [(CX_k - X_k^{\text{obs}})^t W_k (CX_k - X_k^{\text{obs}}) + \lambda (D^m X_k)^t (D^m X_k)] \Delta t_k \quad 3.21$$

$W_k$  est une matrice de pondération, nulle aux temps  $t_k$  où aucune observation n'est disponible,  $D$  est l'opérateur de différentiation itéré à l'ordre  $m$ , et  $\lambda$  est un réel positif dont la valeur doit être déterminée (cf. paragraphe 3.2).

### 3.5.3 Equations adjointes discrétisées du modèle adjoint

Le modèle adjoint conduisant au gradient de cette fonction de coût est :

$$P_k = P_{k+1} + [\partial F_k / \partial X]^t P_{k+1} + C^t W_k (C X_k - X_k^{obs}) + \lambda D^{mt} D^m X_k \quad 3.22$$

$$k = N-1, \dots, 0$$

L'état final adjoint  $P(0)$  obtenu par intégration rétrograde à partir de la condition initiale est :

$$P_N = C^t W_N (C X_N - X_N^{obs}) + D^{mt} D^m X_N \quad 3.23$$

Telle qu'elle est donnée ici, l'initialisation du modèle adjoint discret peut paraître différente de l'initialisation de l'état adjoint continu de la formule 3.10. C'est parce qu'ici nous avons inclus le cas particulier où pourraient exister des observations sur l'état final.

La valeur de l'état adjoint à  $t = 0$  fournit le gradient de la fonction de coût par rapport à l'état initial  $V$  :

$$J_V = P(0) \quad 3.24$$

alors que le gradient par rapport aux paramètres  $U$  est obtenu par le calcul :

$$J_U = \sum_{0 \leq k < N} [\partial F_k / \partial U]^t P_k \Delta t_k \quad 3.24$$

### 3.5.4 problèmes de mise en oeuvre

La mise en oeuvre des méthodes adjointes nécessite de grandes ressources à la fois en moyen de calcul et en temps de développement. Bien entendu, les deux sont liés dans la mesure où, lors de la phase de développement, il faut prendre des décisions qui dépendent essentiellement

des ressources de calcul disponibles. Mais nous voulons insister sur le fait que, par rapport au modèle numérique direct seul, la mise en oeuvre d'une procédure d'assimilation de données impose un surcoût informatique très important.

Tout d'abord, il est nécessaire d'écrire et de coder les équations adjointes. Cette tâche doit être menée avec grand soin car il est primordial de prendre exactement l'adjoint de l'opérateur qui est défini par le modèle numérique. Il existe des méthodes pour coder systématiquement les modèles adjoints à partir du modèle direct (Long *et al.* 1990, Talagrand 1991), mais il faut reconnaître qu'actuellement, il est souvent plus rapide de réécrire le modèle direct et de développer simultanément le modèle adjoint, plutôt que d'écrire l'adjoint d'un code direct déjà existant (Courtier 1987).

La phase de développement des codes adjoints pourrait être conduite rapidement par la mise au point de générateurs automatiques de code adjoint. Des recherches se poursuivent actuellement en France et aux Etats-Unis, et permettront, à terme, de lever cette difficulté.

Lorsque la procédure d'assimilation est prête, il faut s'assurer que les moyens de calcul, qui sont normalement utilisés pour exécuter le modèle direct seul, sont suffisants pour le déroulement de la procédure. En effet les modèles utilisées en géophysique sont en général très gros ( $10^6$  à  $10^7$  degrés de liberté pour certains modèles météorologiques). Les divers algorithmes du gradient itératifs utilisés durant la minimisation nécessitent entre 30 et 100 itérations. Or, une itération du gradient implique l'intégration du modèle direct et celle du modèle adjoint. Comme l'intégration du modèle adjoint nécessite environ le même nombre d'opérations que celle du modèle direct, l'assimilation de données peut multiplier, parfois jusqu'à 100, le temps de calcul par rapport à une intégration simple du modèle direct.

Il se pose de plus un problème de stockage. L'intégration du modèle adjoint a besoin de tous les états du système qui ont été calculés lors de l'intégration du modèle direct. cela nécessite au

préalable de les avoir conservés en mémoire. Or dès que le domaine spatial est important, ou que l'intégration se prolonge, nous obtenons des quantités importantes de valeurs qui doivent parfois être stockées sur support physique. Bien évidemment, la relecture de ces données par le modèle adjoint en est considérablement ralentie.

Il n'y a guère que les super-calculateurs qui puissent offrir les ressources informatiques que demande l'assimilation de données variationnelle. Il est vrai que, jusqu'à présent, ce sont aussi les seuls à supporter les calculs des modèles directs de la météorologie. Cependant ils n'offrent pas encore les moyens de réaliser des assimilation de données variationnelle en mode opérationnel.

Dans le cadre de notre étude, nos modèles sont de taille modeste (environ  $10^3$  degrés de liberté) car ils ont été originellement développés sur un calculateur de puissance réduite. Aussi nous avons commencé par développer la procédure d'assimilation variationnelle sur ce calculateur qui fonctionnait alors aux limites de ses capacités. Puis nous avons transporté les modèles sur le Cray Y-MP2E de l'institut méditerranéen de technologie à Marseille. Là nous disposions largement des moyens de calcul nécessaires à l'exécution de nos travaux.

**Chapitre 4 :**

**ASSIMILATION DANS UN MODELE D'ADVECTION-DIFFUSION :  
MODELE ET OBSERVATIONS**

Cette première étude pratique sur l'assimilation de données variationnelle cherche surtout à identifier les principales difficultés inhérentes à la mise en oeuvre des méthodes adjointes, sans application immédiate. Néanmoins le modèle numérique qui est utilisé présente un certain intérêt puisqu'il est déduit de l'Equation d'Advection-Diffusion. Ces travaux sont méthodologiques et c'est pourquoi, bien qu'ils aient fait l'objet d'une communication (voir annexe 9.3), nous préférons reprendre leur description complète.

#### 4.1 Inversion, assimilation de données et traceurs

L'Equation d'Advection-Diffusion peut modéliser le comportement de certains constituants physiques qui évoluent dans un fluide sans interagir chimiquement avec lui. Ces corps, appelés *traceurs* ou *traceur passifs* lorsque l'on veut insister sur leur inertie chimique, sont ainsi uniquement soumis à l'action physique du fluide (principalement écoulement et mélange) et c'est pourquoi le suivi des déplacements de certains de ces corps dans des systèmes géophysiques comme l'atmosphère ou l'océan permet d'élucider certains aspects de leur dynamique. Toutefois l'extraction de ce genre d'information à partir des mesures de traceur n'est pas simple, et fait appel à des outils mathématiques complexes regroupés sous l'appellation d'*inversion* (Wunsch 1987).

Il existe aussi des situations où le suivi du traceur peut être intéressant non plus pour le milieu ambiant, mais pour le traceur lui-même. Par exemple, il est des polluants qui se comportent chimiquement comme des traceurs. Dans ces conditions, on pourra chercher à exploiter les connaissances disponibles sur la dynamique du milieu pollué, pour suivre le plus précisément possible, le parcours des traceurs (Sarmiento 1983 pour l'océan et Petri *et al.*1990 pour l'atmosphère). Nous retrouvons ainsi le problème de l'assimilation de données.

Nous nous apercevons donc que, si l'inversion et l'assimilation de données ont des objectifs différents, voire contraires, elles ont pourtant beaucoup de points communs.

Effectivement la théorie du contrôle optimal permet d'unifier dans une même approche les problèmes relatifs au suivi des traceurs.

Dans ce chapitre, nous nous proposons de développer cette approche pour évaluer l'apport des images satellitaires de télédétection dans l'étude des traceurs. Nous utilisons un modèle numérique pour simuler la diffusion d'un traceur quelconque. Nous employons le terme "traceur quelconque" pour traduire le fait que ce traceur n'est pas réel. Dans la mesure où notre étude est méthodologique, nous n'utiliserons pas d'images satellitaires existantes. C'est pourquoi nous synthétisons des données de concentration dont les caractéristiques (erreurs, distribution, etc.) seront similaires à celles que présentent généralement les observations de télédétection.

#### 4.2 Position du problème

Nous envisageons le cas où un constituant chimique, appelé dans la suite traceur, est déversé accidentellement et de manière quasi-instantanée dans un bassin fermé (tel qu'un lac). Le traceur diffuse dans les eaux et, après un certain laps de temps, sa répartition est plus ou moins homogène. Il devient alors très difficile de déterminer géographiquement le point où a eu lieu le rejet initial. Une situation analogue se produit lorsque plusieurs sources émettent simultanément de façon continue.

Si le traceur est par exemple un polluant, la situation à laquelle peuvent être confrontée les autorités locales, s'apparente mathématiquement à un problème d'estimation. En effet l'urgence impose d'abord de suivre et d'essayer de prévoir, à partir de relevés ponctuels, le déplacement des nappes de polluant sur l'ensemble du bassin. Puis, à partir de l'observation des effets finaux de la pollution, il s'agit de déterminer le lieu où le polluant a été initialement déversé afin de trouver les responsables de la catastrophe.

Si la résolution de la première tâche peut être grandement facilitée par la multiplication des observations, la seconde tâche sous-entend qu'il faille d'une certaine façon remonter dans le temps, ce qui peut se révéler très délicat dans les cas où le pollueur a opéré quasiment instantanément et que la pollution qui en résulte n'est découverte que tardivement. Nous avons simulé une telle situation au moyen d'un modèle numérique, et nous avons résolu les problèmes qui en découlaient au moyen d'une procédure variationnelle d'assimilation des données de concentration.

#### 4.3 Le modèle numérique

Le modèle numérique qui nous sert à la fois pour simuler les effets du rejet du traceur dans un bassin ainsi que pour réaliser l'assimilation des données de concentration, est basé sur l'équation de diffusion à laquelle est rajouté un terme d'advection.

##### 4.3.1 *l'équation d'Advection-Diffusion*

Nous considérons un bassin D rectangulaire de longueur  $L_x$ , de largeur  $L_y$ , et dont la profondeur ( $L_z \approx 0$ ) est faible devant les dimensions horizontales. Dans ces conditions, nous pouvons étudier les variations horizontales (dans le plan  $(x,y)$ ) de la valeur moyenne de la concentration du traceur. Elle est donnée en tout point du bassin par les équations différentielles suivantes :

$$\partial C / \partial t = \nabla(k \nabla C) - \nabla(\mathbf{v} C) + S \quad 4.1$$

$$C(t=0) = C_0 \quad 4.2$$

où  $t \in [0,T]$  représente le temps,  $C(x,y)$  est la concentration du traceur dans le plan horizontal,  $\mathbf{v}(x,y)$  est le vecteur vitesse horizontal du courant existant dans le bassin,  $k(x,y)$  est le coefficient

de diffusion,  $S(x,y)$  est un terme source qui sera toujours nul dans la suite,  $\nabla$  est l'opérateur gradient dans le plan  $(x,y)$  et  $C_0 = C_0(x,y)$  est une fonction connue.

#### 4.3.2 Conditions aux limites $\partial D$ du bassin $D$

Par analogie avec la diffusion d'un corps chimique inerte dans un lac, nous supposons que le traceur ne peut s'échapper du bassin, ce qui conduit aux conditions limites du type *Neumann* :

$$\frac{\partial C}{\partial \mathbf{n}} = 0, \text{ où } \mathbf{n} \text{ est le vecteur normal à } \partial D \quad 4.3a$$

Auxquelles nous ajouterons la condition suivante :

$$\mathbf{v} \cdot \mathbf{n} = 0, \text{ le long de } \partial D \quad 4.3b$$

Alors une fonction initiale  $C_0 = C_0(x, y)$ , définit une unique fonction  $C(x, y, t)$  solution des équations 4.1-3 sur l'intervalle de temps  $[0,T]$ .

Pour pouvoir être intégrée sous forme de modèle numérique, l'équation différentielle doit être discrétisée en temps et en espace. Il existe plusieurs possibilités pour cette discrétisation, donnant lieu à autant de modèles numériques différents (Cirillo and Poli 1992). Pour notre part, nous avons utilisé un modèle développé à l'INRIA de Sophia-Antipolis.

#### 4.3.3 Discrétisation de l'équation d'Advection-diffusion

Le modèle numérique est obtenu par discrétisation dans l'espace du système d'équations 4.1-3 en éléments finis triangulaires, et en utilisant le schéma *Euler forward* pour l'intégration dans le temps. Les calculs induits par la discrétisation sont classiques pour ce genre de schéma. Nous retiendrons surtout que les valeurs prises par la concentration aux  $ns$  sommets du

maillage délimité par les  $nt$  triangles sont regroupées en un vecteur d'état de dimension finie égale à  $ns$ . Ce vecteur des concentrations est calculé en  $N$  dates  $0, t_1, \dots, t_N = T$ , régulièrement espacées de l'intervalle de temps  $\Delta t$  dans la période d'intégration  $[0, T]$ . Les concentrations  $C_{n+1}$  à l'instant  $t_{n+1}$  sont obtenues à partir des concentrations  $C_n$  à l'instant  $t_n$  par la récurrence matricielle suivante :

$$C_{n+1} = C_n + \Delta t [A_d C_n - A_v C_n] \quad 4.4$$

où  $A_d$  et  $A_v$  sont des matrices carrées de dimension  $ns^2$  et  $\Delta t = t_{n+1} - t_n$  est le pas d'intégration temporelle.

Les matrices  $A_d$  et  $A_v$  représentent les opérateurs discrets qui réalisent respectivement les opérations nécessaires aux calculs des termes de diffusion et de transport. En particulier, grâce aux conditions aux limites du type Neumann qui ont été adoptées, la matrice  $A_d$  est symétrique alors que la matrice  $A_v$  est anti-symétrique. Nous remarquerons aussi que l'équation 4.4 est linéaire par rapport à la variable  $C_n$ .

#### 4.3.4 Stabilité du schéma de discrétisation

Le schéma de discrétisation utilisé pour intégrer le modèle numérique de l'équation 4.4 sera stable si les conditions suivantes sont vérifiées (cf. Lions 1984 pour le détail des calculs) :

$$\Delta t < 1/2 \Delta x^2 / k \quad \text{et} \quad \Delta t < \Delta x / v \quad 4.5$$

où  $\Delta t$  est le pas d'intégration temporelle,  $\Delta x$  est la hauteur des triangles du maillage.

La vitesse  $v$  ainsi que le coefficient de diffusion  $k$  étant normalement imposés par la physique du problème, nous voyons que le pas de temps  $\Delta t$  maximal varie comme le carré de la

résolution spatiale. Ainsi, travailler à haute résolution ( $\Delta x$  petit) imposera de faibles valeurs du pas d'intégration  $\Delta t$ , ce qui nécessitera un nombre important d'itérations pour atteindre l'instant final  $T$ .

En particulier dans le cas extrême où les inégalités 4.5 deviennent des égalités, nous obtenons la valeur maximale de la hauteur des triangles du maillage :

$$\Delta x < k / 2v \quad 4.6$$

Ici nous obtenons une condition de stabilité du schéma numérique de discrétisation. Au cours de nos simulations nous veillerons donc à vérifier la relation :  $v \Delta x / k < 2$ .

#### 4.3.5 Maillage du domaine et choix des paramètres

Le domaine est un carré de 30 km de côté, partagé en 1800 triangles semblables, isocèles et rectangles, générant ainsi 961 sommets. Chaque triangle possède un côté mesurant 1 km, une hypoténuse de 1,4 km et une hauteur de 0,7 km, ce qui donne une aire de 0,5 km<sup>2</sup>. La période d'étude est de 12 heures.

Seule la composante  $v_x$  du champ de vitesse  $v(v_x, v_y)$  est non nulle. Ainsi le long de ce côté (direction  $Ox$ ) les conditions aux limites  $\mathbf{v} \cdot \mathbf{n} = 0$  sont naturellement vérifiées. Cette vitesse est uniforme (1 km/h) (voir figure 4.2) sur tout le domaine, sauf au voisinage des deux autres bords du bassin (direction  $Oy$ ) où elle varie très rapidement pour s'annuler le long des bords. Avec ce champ de vitesse ainsi défini, les principales structures seront transportées sur une distance d'environ un demi-bassin à la fin de l'intégration. Cette dernière condition implique un courant de retour en profondeur afin de conserver la masse du traceur. Le champ des vitesses est donc divergent localement sur les bords gauche et droit. Une autre approche possible, et proche de la réalité, aurait pu consister à laisser le bord droit ouvert (c'est à dire ne pas lui imposer la condition 4.3b). Ces dernières remarques n'altèrent en rien les conclusions que nous tirons de cet exemple

méthodologique. Afin d'équilibrer transport et diffusion, le coefficient de diffusion est constant sur tout le domaine et égal à  $1 \text{ km}^2/\text{h}$ .

La valeur maximale du pas d'intégration permise est d'un peu plus de 15 mn. Comme nous allons ensuite intégrer le modèle adjoint de ce modèle, il est préférable de disposer d'un schéma d'intégration aussi stable que possible (Ghil et Malanotte-Rizzoli 1991), et c'est pourquoi nous utiliserons un pas de 7,5 mn. Quarante vingt seize itérations seront nécessaires pour intégrer le modèle de l'état initial ( $t = 0$ ) jusqu'à l'instant final à ( $T = 12$  heures).

#### 4.4 Distribution spatio-temporelle des observations

La première nécessité d'une assimilation de données est de disposer d'observations et cela en quantité suffisante. Il semble même (Zou *et al.* 1991) que leur quantité prime sur leur qualité. Toutefois un nombre d'observations important ne garantit pas forcément le succès (mathématique) de l'assimilation. En effet, plus importante que la quantité est la répartition spatio-temporelle de ces observations (cf. chapitre 2). Cette remarque est particulièrement importante en télédétection car les observations sont à la fois distribuées en temps et en espace.

Dans la théorie du contrôle optimal, il est *a priori* tout à fait possible de prendre en compte des observations distribuées irrégulièrement à la fois dans l'espace et dans le temps. Il suffit pour cela que l'information qu'elles apportent soit suffisante pour faire converger la procédure d'assimilation. *La condition d'observabilité* (Le Dimet *et al.* 1991) est un critère permettant de déterminer à partir du modèle numérique et du système d'observation, si le problème mathématique posé par l'assimilation de données est solvable ou non.

Parce que le domaine étudié est relativement peu important ( $30 \times 30 \text{ km}^2$ ) et la période d'étude plutôt courte (12 heures), nous ne simulerons pas plusieurs passages du satellite mais nous supposerons qu'un satellite a couvert tout le domaine instantanément. C'est à dire qu'à un instant donné, et unique, nous disposons d'observations sur la concentration en chaque sommet du maillage. L'instant de la période d'assimilation auquel surviennent les observations est primordial. Puisque le jeu d'observations recouvre complètement le domaine, il est clair que si les données sont disponibles à l'instant initial, alors nous disposons d'un état initial et il n'est pas nécessaire de réaliser l'assimilation de données. Si maintenant les observations sont disponibles à un instant ultérieur, alors l'information contenue dans l'état initial que nous recherchons, aura diffusé dans le domaine et il sera plus difficile de retrouver cet état. Nous nous sommes placé dans le cas extrême où les observations sont effectuées à l'instant final, c'est-à-dire à 12 heures.

#### 4.5 Synthèse des observations de télédétection

##### 4.5.1 *représentation graphique*

Les valeurs de concentration étant données en chaque sommet du maillage, il est nécessaire de procéder à une interpolation pour créer un champ de concentration continu se présentant comme une image de télédétection. Pour cela, nous avons défini une image couvrant le bassin en  $512 \times 512$  pixels, offrant ainsi une résolution  $d = 58,5 \text{ m}$ . Chaque pixel  $p$  est ensuite affecté selon ses coordonnées cartésiennes géographiques au triangle  $t$  du maillage qui l'entoure. Dans ce triangle, les coordonnées barycentriques  $\alpha, \beta, \gamma$  du pixel  $p$  sont calculées à partir de ses coordonnées cartésiennes géographiques ainsi que ceux des sommets  $s_1, s_2, s_3$  du maillage définissant le triangle  $t$ . La valeur de la concentration  $C_p$  affectée au pixel  $p$  est la valeur obtenue par l'interpolation barycentrique, c'est à dire  $C_p = G (\alpha C_{s1} + \beta C_{s2} + \gamma C_{s3})$ ,  $G$  étant un coefficient d'étalonnage qui permet de faire correspondre des valeurs de concentration (en réels dans le modèle) en des niveaux de gris de l'image (en octets, de 0 à 255).

L'intérêt de l'interpolation barycentrique est double. Les éléments nécessaires au calcul des coordonnées  $\alpha$ ,  $\beta$ ,  $\gamma$  sont en grande partie déduits de fonctions déjà codées dans le modèle numérique. De plus, avec ce type d'interpolation, les vraies valeurs des concentrations aux sommets du maillage sont conservées à l'étalonnage près. Donc, lors des simulations numériques, les valeurs des pixels correspondants géographiquement aux sommets du maillage peuvent être directement introduites dans le modèle.

#### 4.5.2 *Etat initial vrai*

A l'instant initial, un rejet quasi ponctuel du traceur a lieu sur le milieu du bord gauche du bassin ( $x=0$ ,  $y=15$  km), la concentration du traceur étant nulle partout ailleurs. Nous disons quasi-ponctuel car, pour ne pas introduire de fonction discontinue, le rejet a la forme d'une cloche de Gauss, très concentrée autour du point supposé être la source. Ainsi en chaque sommet  $s(x_v, y_v)$  du maillage, l'état initial est donné par la fonction suivante :

$$C_0(x_s, y_s) = 1000 \exp[-x_v^2 + (y_v - 15)^2 / 1,5^2] \quad 4.7$$

L'image résultant de l'interpolation est donnée par la figure 4.2. Cet état initial sera par la suite appelé *état initial vrai*.

#### 4.5.3 *observations pures*

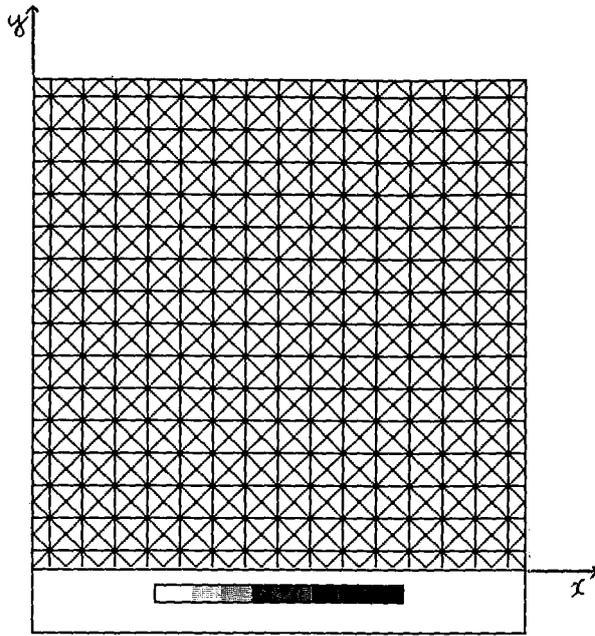
A partir de l'état initial vrai, le modèle est intégré. L'état obtenu à  $t = 12$  heures par cette simulation, appelée simulation de contrôle, est ensuite interpolé pour fournir une image de l'état du bassin (figure 4.3). Cette image constitue les observations de télédétection, qualifiées de *pures*, car elles sont solutions exactes du modèle. Cette propriété est importante car, avec ce jeu de données, nous connaissons déjà une solution au problème posé par leur assimilation : il s'agit de l'état initial vrai.

#### 4.5.4 observations bruitées

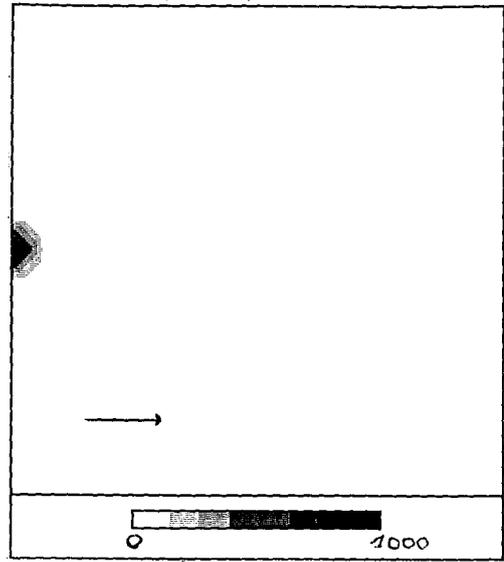
Nous avons bruité les observations pures obtenant ainsi une seconde image qui représente des observations plus proches de ce que fournirait un capteur spatioporté. Le bruitage des observations pures a été réalisé par la transformation suivante :

$$C_{\text{bruit}}(x,y) = (1 + R(x,y)) C_{\text{pure}}(x,y) \quad 4.8$$

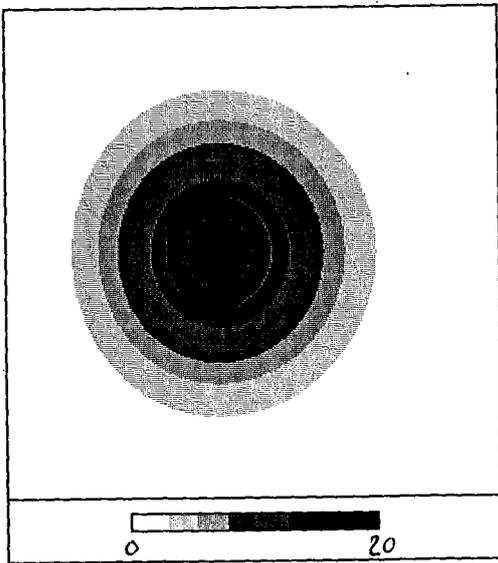
où  $R(x,y)$  est une variable aléatoire uniforme qui ne prend que deux valeurs : + 0,2 et - 0,2. Le bruitage est assez extrême puisque, sans modifier la moyenne et l'écart-type des observations, il impose une erreur de 20 % sur toutes les données. Comme cela apparaît sur l'image des observations bruitées (figure 4.4), le bruit obtenu par le générateur aléatoire utilisé n'est pas parfait, mais présente certaines structures. Dans la réalité les erreurs de mesure des appareils embarqués à bord des satellites présentent souvent des structures similaires à nos observations artificiellement bruitées.



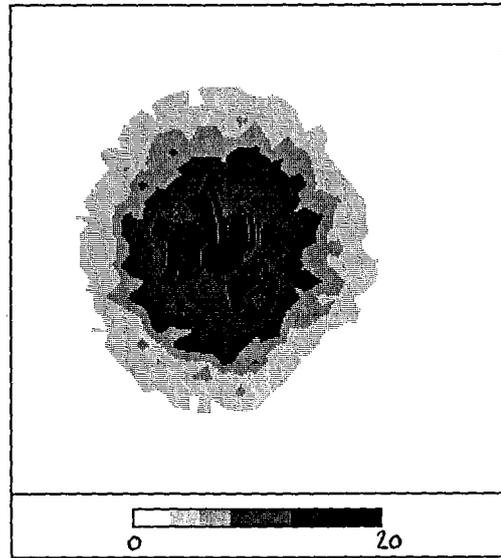
**Figure 4.1 :**  
 Maillage du bassin carré 30 x 30 km<sup>2</sup>  
 961 sommets - 1800 triangles.



**Figure 4.2 :**  
 Etat initial vrai, la flèche indique  
 la direction de la vitesse d'advection,  
 $v = 1\text{km/h}$ ,  $t = 0$  heure.



**Figure 4.3 :**  
 Observations pures obtenues par intégration de  
 l'état initial vrai à  $t = 12$  heures.



**Figure 4.4 :**  
 Observations bruitées à  $t = 12$  heures :  
 observations pures  $\pm 20\%$  de bruit aléatoire.

## 4.6 La fonction de coût

### 4.6.1 *Ecart avec les observations*

La fonction de coût évalue l'écart entre la sortie du modèle et les observations. Nous utiliserons une fonction quadratique éventuellement pondérée, ce qui conduit à la fonction de coût suivante :

$$J_{\text{obs}} = 1/2 \sum_{1 \leq i \leq n_s} w_i | C_i(t=12h) - C_i^{\text{obs}} |^2 \quad 4.9$$

Les coefficients  $w_i$  permettent de pondérer les observations entre elles, généralement suivant le degré de confiance accordée à chaque mesure. Ici comme toutes les observations ont la même erreur et les coefficients  $w_i$  sont tous égaux à 1.

### 4.6.2 *Unicité du minimum de la fonction de coût*

L'assimilation de données variationnelle va consister à faire varier l'état initial  $C_0$  afin que 12 heures plus tard cet état  $C(12h)$  intégré par le modèle ajuste au mieux les observations  $C^{\text{obs}}$ . La qualité de l'ajustement étant indiquée par la valeur de la fonction de coût  $J_{\text{obs}}$ , le meilleur ajustement sera celui qui minimisera cette fonction. L'assimilation ne sera donc possible que si, tout d'abord, il existe un minimum à cette fonction de coût, et, si de plus, ce minimum est unique.

Comme nous travaillons sur l'intervalle  $[ 0, +\infty [$ , il existe au moins un minimum à la fonction de coût. Le problème de la multiplicité des minimas est résolu immédiatement par la condition d'observabilité, puisqu'ici la matrice  $C$  est égale à l'identité. Toutefois, notre problème est très mal conditionné et il faudrait donc calculer la matrice Hessienne pour pré-conditionner le système lors de la minimisation. Malheureusement, l'absence de librairie mathématique appropriée lors du développement du code nous a privé des moyens d'obtenir cette matrice. Aussi nous avons

essayé de résoudre le problème du conditionnement par l'usage de termes de pénalisation dans la fonction de coût.

#### 4.6.3 Pénalisation-régularisation

Plusieurs formes de pénalisation sont possibles. En automatique, il est souvent requis que le contrôle soit réalisé avec une dépense d'énergie minimale, et pour cela, on cherche à limiter l'amplitude de l'état initial solution. Dans ce cas la fonction de coût comportera un terme supplémentaire du type :

$$J_{amp} = 1/2 \sum_{1 \leq i \leq ns} w_i | C_i(t=0) |^2 \quad 4.10$$

Toutefois si ce type de pénalisation a une signification évidente pour les systèmes industriels, son utilisation est difficilement justifiable pour les systèmes géophysiques. Il n'y a en effet *a priori* aucune raison de limiter seulement l'amplitude initiale d'un phénomène physique indépendamment de son évolution ultérieure. Par ailleurs les essais que nous avons effectués avec ce type de pénalisation se montrèrent complètement infructueux du fait du caractère particulier du point de départ retenu pour la minimisation.

C'est pourquoi nous utiliserons de préférence des contraintes sur les dérivés de la solution. C'est à dire que nous choisirons parmi les solutions possibles, celle qui est la plus régulière. Ici nous utiliserons un terme de pénalisation construit sur la dérivée seconde de la solution qui pénalisera les grandes amplitudes de la dérivée :

$$J_{smo} = 1/2 \sum_{1 \leq i \leq ns} [ (\partial^2 C_i / \partial^2 x)^2 + (\partial^2 C_i / \partial^2 y)^2 + 2(\partial^2 C_i / \partial y \partial x)^2 ]^2 (t=0) \quad 4.11$$

Bien qu'assez particulière, cette forme de pénalisation a été étudiée en détail (Wahba 1987). En particulier la méthode GCV (*General Cross Validation*) issues de ces travaux, permet d'estimer la valeur optimale du paramètre de lissage  $\lambda$  pondérant l'importance de la pénalisation dans la fonction de coût finale :

$$J_{\text{cost}} = 1/ns J_{\text{obs}} + \lambda J_{\text{smo}} \quad 4.12$$

Dans laquelle le facteur  $1/ns$  est introduit afin de faciliter l'application de la méthode GCV.

#### 4.6.4 Détermination du paramètre de lissage

La méthode GCV permet d'estimer la valeur du paramètre de lissage  $\lambda$  à partir des observations. Nous ne détaillerons pas cette méthode qui a fait l'objet de plusieurs publications (Wahba et Wandelberger 1980, Villalobos et Wahba 1987, Bates *et al.* 1986) et qui peut être appliquée en utilisant les fonctions de la librairie GCVPACK, disponible sur le serveur *statlib* ([statlib@orn.edu.gov](mailto:statlib@orn.edu.gov)). Néanmoins l'exécution de ces fonctions nécessite des ressources de calcul dont nous ne disposons pas initialement. Aussi nous avons déterminé  $\lambda$  en traçant la décroissance de la fonction  $J_{\lambda}(X)$  pour plusieurs valeurs de  $\lambda$ , et avons retenu la valeur pour laquelle  $J(x)$  diminue le plus en 30 itérations de descente du gradient. Que les observations soient bruitées ou non, nous avons trouvé une valeur identique  $\lambda = 50$ . Ce résultat un peu surprenant s'expliquera au vu des résultats exposés dans le chapitre suivant.

#### 4.7 Utilisation du modèle adjoint.

La fonction de coût  $J_{\text{cost}}$  étant définie, l'assimilation de donnée consiste en sa minimisation par rapport aux variables de contrôle. Il est donc nécessaire de calculer le gradient de la fonction de coût par rapport à chaque variable de contrôle. Ici la variable de contrôle est l'état initial  $C_0$ , ce qui

introduit une difficulté dans la mesure où la fonction de coût n'est qu'une fonction implicite de la variable  $C_0$ .

D'après la théorie du contrôle optimal, une possibilité de calculer, relativement simplement, le gradient de la fonction de coût par rapport à l'état initial est offerte par l'introduction de l'état adjoint  $P(t)$ , solution du système adjoint du système d'équations 4.1-3. L'ajout d'un terme de pénalisation modifie légèrement la forme de l'équation 3.11 du chapitre 3 qui devient :

$$\text{Grad}[J_{\text{cost}}]/C_0 = P(t=0) + \lambda \partial J_{\text{cost}}/\partial C_0 \quad 4.12$$

Le calcul du gradient de la fonction de coût se ramène à la détermination de l'état adjoint  $P(t)$  à  $t = 0$  par intégration du modèle adjoint .

#### 4.7.1 Le modèle adjoint

Compte tenu de la forme de la fonction de coût, le système d'équations continues adjointes du système d'équations continues directes 4.1-3 est :

$$- \partial P / \partial t = \nabla(k \nabla P) + \nabla(v P) \quad 4.13$$

$$P(t=12) = C_0 - C^{\text{obs}} \quad 4.14$$

Le modèle adjoint doit être intégré à rebours à partir de l'état connu à  $t = 12$  heures jusqu'à l'instant initial  $t = 0$ , où sa valeur fournit le gradient de la fonction de coût par rapport à l'état initial.

#### 4.7.2 Algorithme de descente du gradient

Une fois obtenu le gradient de la fonction de coût, un algorithme de descente dans la direction du gradient est applicable. Aux premiers temps du développement du code sur mini-calculateur, nous n'avions pas accès à une librairie mathématique et donc nous avons écrit une

fonction de minimisation d'après l'algorithme de la plus grande pente du gradient (Souly et Zolésio 1991). Lors de simulations sur supercalculateur, il était possible d'utiliser des minimiseurs très efficaces, basés sur l'algorithme Quasi-Newton (Navon and Legler 1987). Néanmoins cela aurait demandé d'importantes adaptations du code existant et, de plus, nous n'aurions pas pu ajouter des contraintes de type inégalité lors de la minimisation. Nous verrons, dans le chapitre suivant, la nécessité de ce type de contrainte dans le problème que nous nous posons.

#### 4.7.3 *Le point de départ de la minimisation*

Le processus de minimisation est itératif et doit donc être démarré à partir d'un certain état initial, appelé *point de départ* de l'assimilation. Bien évidemment, lorsque le minimum de la fonction de coût est unique, la minimisation, si elle est correctement menée, convergera toujours vers ce minimum quelque soit son point de départ, et donc dans ces cas idéaux le choix du point de départ n'a mathématiquement pas d'importance. Dans les cas moins favorables, par contre ce choix a une influence sur la vitesse de convergence (nombre d'itérations de la minimisation) et peut aussi mener à l'obtention de solutions différentes. C'est pourquoi, il est important de considérer le point de départ comme une source d'information au même titre que les observations, et d'essayer de proposer à la procédure de minimisation un point de départ aussi proche que possible de la solution finale attendue.

Toutefois pour nos expérimentations nous avons prétendu ne pas avoir d'information sur la solution. Le point de départ est l'état de repos complet, c'est-à-dire que pour la première intégration du modèle direct, la concentration  $C_0$  est nulle en tous les sommets du maillage. Nous verrons, dans le chapitre suivant, les conséquences de ce choix dans les résultats numériques de l'assimilation.

**Chapitre 5 :**

**ASSIMILATION DANS UN MODELE D'ADVECTION-DIFFUSION :  
RESULTATS NUMERIQUES**

Nous présentons les résultats numériques obtenus par application de la procédure d'assimilation détaillée dans le chapitre 4. L'efficacité des techniques de pénalisation dans la convergence de la procédure d'assimilation est mise en évidence sur certains problèmes que pose l'emploi de l'Equation d'Advection-Diffusion. Finalement, quelques applications possibles de cette étude académique sont évoquées.

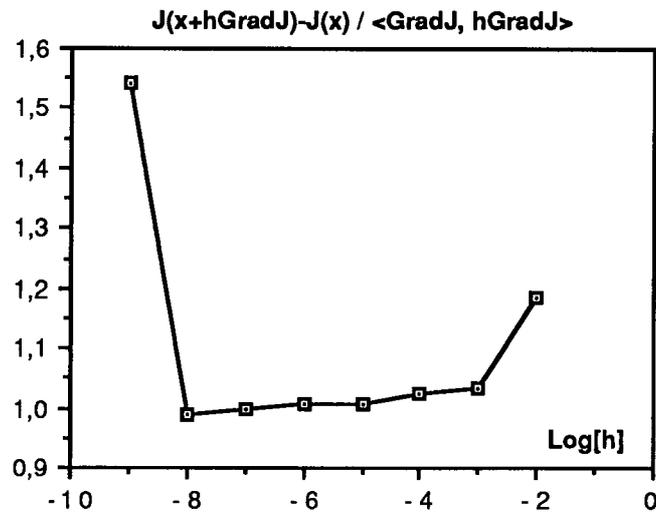
### 5.1 Test de la procédure d'assimilation

Avant de commencer les expériences d'assimilation, nous avons testé la validité du modèle adjoint. En effet, le développement de la procédure d'assimilation s'effectue en trois phases successives et il est impérieux de tester le code avant de passer d'une phase à la suivante.

Aussi à chaque étape, nous avons effectué les tests préconisés par Courtier *et al.* (1990). Ces tests transitoires ayant tous été vérifiés, nous présentons seulement l'ultime test qui nous a permis de valider le code. Ce test consiste à s'assurer que la formule de Taylor est vérifiée dans la direction du gradient calculé par le modèle adjoint. Plus précisément, il s'agit de vérifier l'égalité :

$$1 = \lim_{h \rightarrow 0} \frac{J(x + h \text{ Grad}J) - J(x)}{\langle \text{Grad}J, h \text{ Grad}J \rangle} = 1 \quad 5.1$$

En fait, la direction peut être choisie arbitrairement, mais celle du gradient offre un bon conditionnement des différentes variables. Nous avons tracé (Graphique 5.1) les variations de ce rapport en fonction du logarithme de  $h$ , pour  $h$  variant de  $10^{-9}$  à  $10^{-1}$ . Cette figure montre, que pour  $h$  décroissant de  $10^{-1}$  à  $10^{-8}$ , la limite définie par la formule 5.1 converge effectivement vers 1, cependant, pour  $h$  inférieur à  $10^{-8}$ , ce rapport augmente subitement car, pour de si faibles valeurs, les erreurs d'arrondi ne sont plus négligeables et le terme de limite n'a plus de signification. Nous avons par ailleurs vérifié que l'erreur résiduelle  $(1 - 1)$  convergeait linéairement vers 0 à la précision de la machine près.



**Graphique 5.1 :**  
Test du modèle adjoint (voir § 5.1).

## 5.2 Présentation générale des résultats

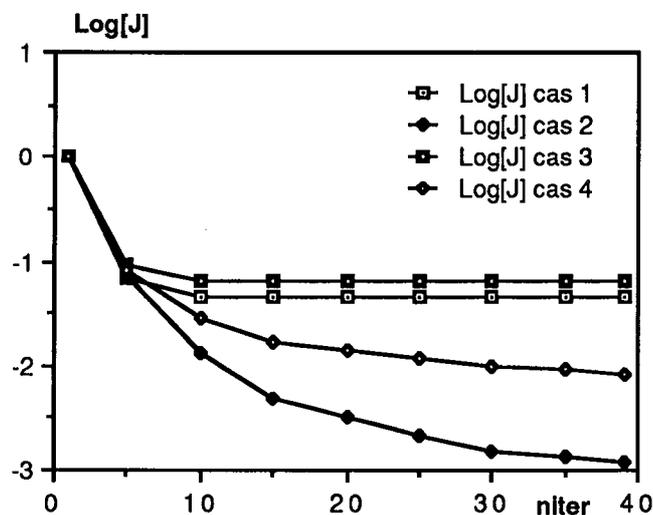
### 5.21 *Les cas étudiés*

Afin d'évaluer, d'une part, l'impact de l'erreur d'observation et, d'autre part, l'efficacité du terme de pénalisation dans la fonction de coût, nous avons été amenés à envisager quatre cas :

- cas 1 : Assimilation des observations pures sans pénalisation,
- cas 2 : Assimilation des observations pures avec pénalisation,
- cas 3 : Assimilation des observations bruitées sans pénalisation,
- cas 4 : Assimilation des observations bruitées avec pénalisation.

Dans chaque cas, nous avons opéré de la même façon : la minimisation est arrêtée lorsque la norme du gradient est inférieure d'un certain facteur multiplicatif  $\varepsilon$  à la norme du gradient initial. Ce critère

d'arrêt n'implique pas que la convergence soit complète, mais au contraire signifie qu'une descente de plus dans la direction du gradient ne modifiera la solution que d'une valeur inférieure à  $\epsilon$ . Ici nous avons retenu comme critère d'arrêt  $\epsilon = 10^{-8}$ , les résultats, c'est-à-dire l'état initial qui a été retrouvé à ce stade la minimisation, sont alors examinés. Pour faciliter les comparaisons entre les différents cas étudiés, les 4 courbes de variation de la fonction de coût ( $J_{\text{obs}}$  uniquement, cf. chapitre 4) par rapport au nombre d'itérations de la descente du gradient sont tracées sur une même figure (Graphique 5.2).

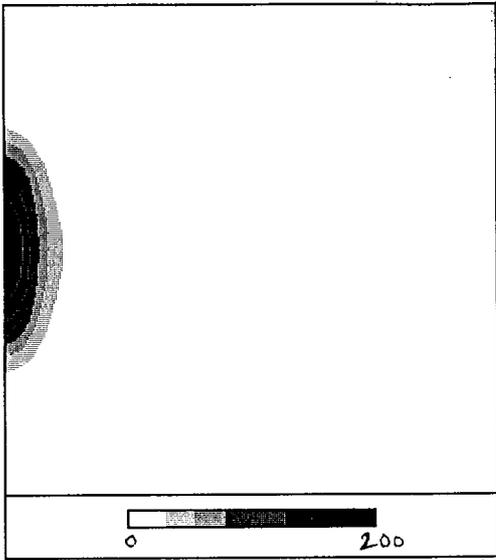


**Graphique 5.2 :**  
Variation du logarithme de la fonction de coût  $J$   
avec le nombre d'itérations de la descente du gradient.

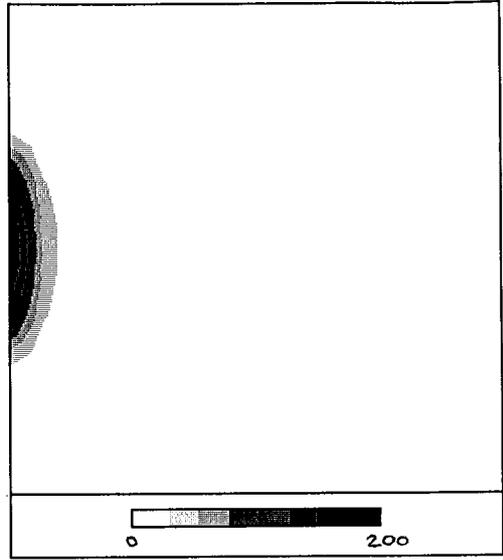
### 5.2.2 Visualisation

Pour chaque cas, une image satellitaire de l'état initial solution de la minimisation a été synthétisée en utilisant l'interpolation barycentrique. Ces images, représentées respectivement pour chaque cas en figure 5.1, 5.2, 5.3 et 5.4, donnent surtout l'aspect général des solutions car, du fait de l'erreur de troncature, les contrastes sont atténués. Ces images sont à comparer avec l'image de la figure 4.2 du chapitre 4 de l'état initial vrai qui est la solution attendue. Toutefois il

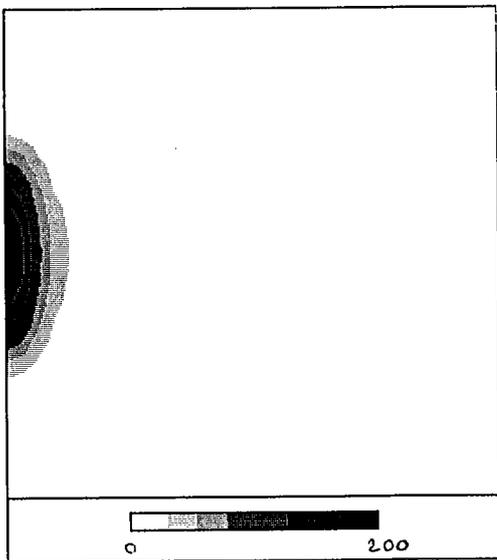
convient de faire attention au changement de l'échelle de couleur entre les images solutions et l'état initial vrai. Cette variation d'échelle, nécessaire pour une bonne visualisation, est par ailleurs révélatrice de la précision avec laquelle les solutions sont retrouvées. La visualisation étant un moyen par trop subjectif, nous avons cherché à comparer les résultats par un critère quantitatif.



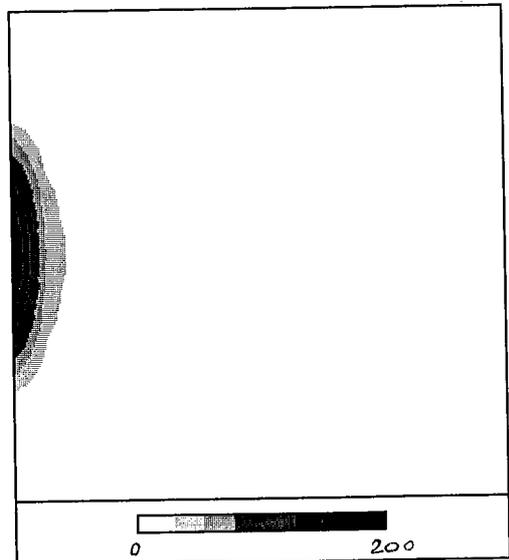
**Figure 5.1 :**  
Cas 1 : observations pures  
sans pénalisation.



**Figure 5.2 :**  
Cas 2 : observations pures  
avec pénalisation.



**Figure 5.3 :**  
Cas 3 : observations bruitées  
sans pénalisation.



**Figure 5.4 :**  
Cas 4 : observations bruitées  
avec pénalisation.

### 5.2.3 L'erreur résiduelle

Puisque le but premier de l'assimilation est de minimiser la fonction de coût, en toute rigueur, les performances de la minimisation doivent être jugées sur la valeur du rapport de la fonction de coût  $J_k$ , calculée à l'itération  $k$  de la descente, sur la valeur de  $J_0$ , calculée au point de départ. Néanmoins comme nous disposons de l'état initial vrai (cf. chapitre 4) qui est une solution dans les cas 1 et 2, et qui est aussi l'état dont nous voudrions voir la solution s'approcher dans les cas 3 et 4, il est beaucoup plus intéressant de suivre, avec le nombre d'itération, l'évolution de l'écart entre l'état initial solution de la minimisation et l'état initial vrai.

Pour quantifier cette erreur nous avons défini l'erreur résiduelle (ERR0) à l'instant initial par la formule suivante :

$$ERR0 = \sqrt{[\sum_D | C^k(t=0) - C^v(t=0) |^2 / \sum_D | C^v(t=0) |^2]} \quad 5.2$$

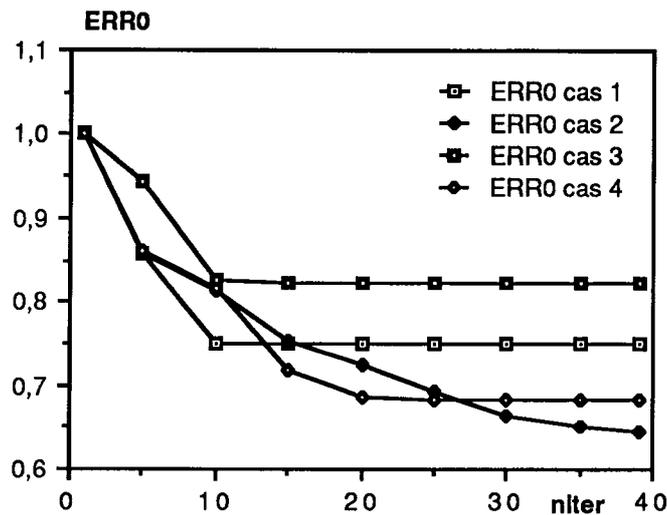
où  $C^k(t=0)$  est l'état initial solution après  $k$  itérations de la descente du gradient,  $C^v(t=0)$  est l'état initial vrai et  $D$  est le bassin, soit le domaine constitué des sommets du maillage.

De même, pour comparer les solutions de l'assimilations à 12 heures avec les observations, nous utiliserons en plus de la fonction de coût, l'erreur résiduelle à  $t = 12$  heures soit la fonction :

$$ERR12 = \sqrt{[\sum_D | C^k(t=12h) - C^k |^2 / \sum_D | C^{obs} |^2]} \quad 5.3$$

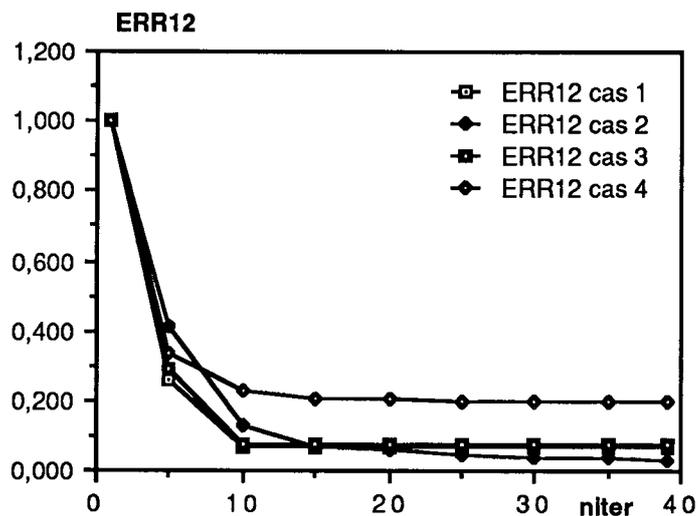
Les notations étant similaires aux précédentes,  $C^{obs}$  représente les observations qui ont été assimilées.

Comme précédemment, pour faciliter les comparaisons entre les différents cas étudiés, les 4 courbes de variation de ERR0 (respectivement ERR12) sont tracées sur une même figure (respectivement Graphique 5.3 et Graphique 5.4).



**Graphique 5.3 :**

Variation de l'erreur résiduelle initiale ERR0 avec le nombre d'itérations de la descente du gradient.



**Graphique 5.5 :**

Variation de l'erreur résiduelle finale ERR12 avec le nombre d'itérations de la descente du gradient.

### 5.3 Commentaires

D'après le chapitre 4 nous savons que, dans chaque cas, il y a une solution unique au problème de minimisation. Comme nous savons, de plus, que l'état initial vrai est solution de l'assimilation dans les cas 1 et 2, puisqu'il a donné naissance aux observations assimilées, nous connaissons la solution dans ces deux cas et allons pouvoir évaluer le degré d'approximation avec laquelle elles sont retrouvées lors de la minimisation. En ce qui concerne les jeux d'observations bruitées (cas 3 et 4), il n'est pas du tout évident que l'état initial vrai soit la solution puisque cela va dépendre de l'écart qui a été introduit lors du bruitage entre les observations et la trajectoire du modèle dans l'espace des phases. Il est clair que, pour des applications de télédétection, un modèle assimilateur robuste est préférable et c'est pourquoi, nous aimerions aussi dans ces cas retrouver l'état initial vrai comme solution de la minimisation.

De plus, le fait de connaître la solution dans les cas 1 et 2 va nous permettre d'identifier la solution dans les cas 3 et 4. En effet le processus est itératif et donc la solution n'est qu'approchée à un degré qui dépend du taux de convergence. Pour identifier l'état initial nous avons principalement deux critères, l'un visuel en synthétisant l'image de l'état initial solution, et l'autre, numérique à travers l'évolution de l'erreur résiduelle  $ERR_0$ . La décroissance de la fonction de coût ainsi que celle de l'erreur résiduelle finale permettent surtout de s'assurer que le processus de minimisation s'effectue correctement.

Quoiqu'il en soit, l'examen des évolutions de ces quatre paramètres sont très semblables d'un cas à l'autre. Si donc, nous admettons conformément à la théorie, que la solution dans les cas 1 et 2 tend vers l'état initial vrai, nous sommes en mesure de dire que c'est aussi vers cette limite que tend la solution de la minimisation dans les cas 3 et 4. Il est à noter que la précision avec laquelle l'état initial vrai a été retrouvé est faible et n'est guère affectée par le bruitage des observations.

Nous avons aussi voulu vérifier que la solution finale n'est pas liée au choix du point de départ et nous avons recommencé les expériences en utilisant comme point de départ un état dont la concentration en chaque sommet est choisie de façon aléatoire. Après quelques itérations, les concentrations initiales solution de la minimisation deviennent toutes nulles retombant ainsi sur l'état de repos qui est le point de départ que nous avons adopté, ensuite la minimisation devient en tous points similaires à la descente décrite dans le paragraphe précédent.

Il semble donc que l'état de repos joue un rôle particulier pour ce modèle d'Advection-Diffusion. Ce rôle sera certainement mieux cerné, si nous n'oublions pas de préciser, qu'au cours de la minimisation, nous avons dû imposer à chaque pas de la descente une contrainte de positivité. En effet, au cours de la minimisation, il est apparu que certains états initiaux solutions comportaient des valeurs de concentration négatives, faisant ensuite diverger le modèle lors de l'intégration. Concrètement pour remédier à ce problème, nous avons mis à 0 toutes les concentrations initiales qui apparaissaient négatives à chaque pas de la descente. Ce traitement particulier a donc fort tendance à annuler les concentrations.

Nous pouvons remarquer par ailleurs que ce procédé revient à ajouter, par des considérations purement physiques, une contrainte supplémentaire au problème de minimisation, puisque nous restreignons l'espace des contrôles admissibles aux valeurs positives. Or il est montré (Wahba 1989) que le fait d'ajouter des contraintes de type inégalité dans les problèmes de minimisation induit par l'assimilation de données améliore fortement la convergence de la procédure.

Nous devons reconnaître aussi qu'au cours des premiers essais nous avons été très surpris par l'apparition de ces concentrations négatives dans le modèle adjoint ; alors que le modèle direct n'utilisait que des valeurs positives. En fait si nous nous référons à la dualité, l'apparition de

valeurs négatives dans la solution est normale compte tenu de l'état initial vrai qui a été choisi (Wunsch 1988). En effet par la théorie du lisseur de Kalman nous savons que cette solution est donnée avec une certaine erreur caractérisée par la matrice de variance-covariance  $P'_0$ . Or cette matrice dépend fortement de la matrice de covariance  $P_0$  qui accompagne l'erreur introduite par l'usage du point de départ à la place l'état initial vrai qui est la solution exacte. Dans le cas de nos expérimentations cette erreur peut atteindre 1000 (100 %) pour les concentrations situées dans l'entourage du point du rejet. Cela signifie qu'au cours de la minimisation, certaines concentrations peuvent varier dans un intervalle d'approximativement -500 à +500 autour des valeurs de l'état initial vrai. Comme ce dernier contient beaucoup de valeurs nulles, il est clair que des valeurs négatives sont permises pour ces concentrations. Finalement, ajouter une contrainte de positivité revient à réduire l'erreur qui est faite en utilisant le point de départ, au lieu de la solution exacte au début de la minimisation. Toutefois les mécanismes de l'action de ces contraintes du type inégalité ne sont pas encore bien clairs, et il serait intéressant à ce sujet de reprendre l'étude en estimant non plus la concentration, mais son logarithme.

Nous pouvons aussi remarquer que l'amélioration de la convergence apportée par la technique de pénalisation est importante, surtout dans le cas où la solution est exacte, ce qui est très encourageant. Néanmoins, le fait que malgré l'addition du terme de pénalisation, dans les cas 2 et 4, la fonction de coût cesse rapidement de décroître, sans que la solution exacte ne soit retrouvée, montre que cette amélioration du conditionnement n'est pas suffisante. Nous devons imputer cette faiblesse à l'algorithme de la plus grande pente du gradient utilisé pour la descente, qui ne permet pas de préconditionner le système. A cette fin, la détermination de la matrice Hessienne serait bienvenue.

Il convient aussi de remarquer que dans les 4 cas, la solution finale est quand même assez éloignée de la solution attendue, puisque l'erreur résiduelle initiale (ERR0) demeure supérieure à 65 %. Cette forte valeur est à opposer à la faible valeur de l'erreur résiduelle finale (ERR12), qui est

inférieure à 10 % dans les cas favorables de données non bruitées. Cette différence souligne la relative insensibilité de l'équation d'Advection-Diffusion par rapport aux conditions initiales, comme cela est par ailleurs révélé par les faibles valeurs de la norme de l'état adjoint  $P(t)$  à  $t = 0$  (puisque'elle représente le gradient de la fonction de coût par rapport aux conditions initiales) qui ne dépasse jamais le dixième de la valeur de la fonction de coût. Ainsi une erreur d'une unité sur l'état initial solution ne répercutera qu'une erreur d'un dixième d'unité sur l'ajustement du modèle aux observations. C'est la raison pour laquelle il n'y a que peu d'écart entre les résultats obtenus avec les observations bruitées et ceux obtenues avec des observations non bruitées. Ceci explique aussi pourquoi la même valeur du paramètre de lissage a été obtenue pour les deux types d'observations (cf. § 4.6.4).

L'état initial n'est donc pas une variable bien adaptée au contrôle de l'Equation d'Advection-Diffusion. Malheureusement des contrôles menés indépendamment sur le coefficient de diffusion, ou la vitesse, ne peuvent avoir que des actions sur leur terme respectif dans l'équation. Quant au contrôle simultané des deux variables, il est permis de s'interroger sur la signification physique d'un problème où la vitesse et le coefficient de diffusion seraient simultanément contrôlés. Il n'y a donc guère que le contrôle sur les conditions aux bords qui pourrait avoir un intérêt pour l'assimilation de données et il serait intéressant de reprendre cette étude en utilisant le contrôle sur ces variables.

Néanmoins nous pouvons déjà prédire que, s'il y a une amélioration, elle sera certainement limitée, et cela à cause du modèle lui-même. En effet, le terme de diffusion du modèle direct se retrouve sans changement dans le modèle adjoint, ainsi toutes les structures contenues dans les observations qui sont introduites à 12 heures sont lissées par ce terme de diffusion durant l'intégration rétrograde du modèle adjoint. C'est donc une information passablement dégradée qui parvient à l'instant initial, d'où la difficulté de reconstruire l'état du système à cet instant là.

#### 5.4 Applications possibles

Comme cela avait été dit, l'étude qui a été présentée au cours des chapitres 4 et 5 est illustrative et n'a pas été développée pour une application particulière. Pourtant des applications à des problèmes d'environnement sont envisageables. En effet le comportement d'un bon nombre de traceurs peut être assez fidèlement modélisé par l'Equation d'Advection-Diffusion. Aussi le modèle peut être directement utilisé, il suffit pour cela de déterminer le coefficient de diffusion à partir de la nature chimique du traceur et des caractéristiques du milieu dans lequel ce dernier évolue, ainsi que d'obtenir le champ des vitesses à partir de connaissances sur la dynamique de ce même milieu. Le maillage permet ensuite d'adapter le modèle à des régions naturelles de géométrie irrégulière.

Bien évidemment, les applications ne sont possibles que si nous disposons de données capable d'identifier le traceur dans son milieu. Ces données peuvent être fournies par des observations de télédétection qui permettent de détecter certains effets caractéristiques des traceurs :

- La couleur, dans le cas des produits chimiques colorés, comme la rodhamine, qui est observable par les capteurs SPOT HRV (High Resolution Visible) avec une résolution spatiale de 20 m, Landsat TM (Thematic Mapper) avec une résolution de 30 m, CZCS (Costal Zone Color Scanner) avec une résolution de 800 m ainsi que par certains capteurs à venir comme MERIS avec une résolution de 250 m ;
- Les modifications de l'état de surface, dues par exemple à une nappe d'hydrocarbure, qui peuvent être repérées par les radars imageurs comme les SAR (synthetic Aperture Radar) embarqués sur le satellite européen ERS-1 et japonais JERS-1, ainsi que par l'imagerie visible précédemment citée ;

- La température (rejet d'eau domestique ou industrielle) qui est mesurée par les capteurs infrarouge AVHRR (Advanced Very High Resolution Radiometer) à bord des satellite de la NOAA (National Oceanic Atmospheric Administration) ainsi que du satellite Landsat TM.

Les applications sont donc nombreuses, et dans le cadre de la lutte pour la sauvegarde de l'environnement, nous devons espérer que l'utilisation des techniques d'assimilation par contrôle optimal va se généraliser dans les années à venir.

### 5.5 Conclusion

A première vue, la faible précision des états retrouvés montre certaines limites à l'assimilation de données dans le modèle d'Advection-Diffusion au moyen des méthodes adjointes. Il convient toutefois de souligner que, d'une part, ce modèle numérique est particulièrement défavorable pour ce genre de technique d'assimilation et, d'autre part, qu'un examen plus approfondi des résultats peut donner lieu à une autre interprétation.

En effet, si nous nous limitons au problème initial qui consistait à retrouver dans un bassin le point de rejet d'un traceur grâce à des observations ultérieures, alors nous pouvons dire que l'assimilation de données variationnelle est un outil satisfaisant pour ce problème de détection. En effet dans le but de caractériser l'exactitude des solutions nous avons calculé les deux quantités suivantes :

- le rapport  $M_s/M_v$ , égale à la masse totale de traceur contenue dans l'état initial solution divisée par la masse totale contenue dans l'état initial vrai,
- la corrélation entre l'état initial solution et l'état initial vrai.

Dans les 4 cas étudiés, nous avons obtenu un rapport  $M_s/M_v$  supérieur à 90 % et une corrélation d'environ 0.8. Il faudrait bien entendu connaître plus en détail les caractéristiques (probabilité de

fausse alarme et probabilité de non détection) de cette méthode de détection par assimilation de données, mais elle semble intéressante car elle pourrait être facilement intégrée dans une procédure de détection automatisée.

Cela dit, l'objectif premier de ces travaux était d'étudier les conditions d'utilisation des méthodes adjointes pour des applications de télédétection. Force est de constater que le modèle qui a été mis à contribution n'était pas vraiment approprié. Néanmoins, et cela représente un intérêt considérable, la théorie du contrôle optimal dont sont extraites les méthodes adjointes nous a permis de prévoir et d'expliquer les raisons de la faiblesse du modèle, et surtout propose des recettes (pénalisation) afin d'éviter les écueils. Les propriétés mathématiques des méthodes adjointes sont bien établies et là réside aussi un de leurs intérêts lors d'application à la physique.

Si donc nous faisons abstraction des difficultés dues au cas particulier du modèle, il résulte de cette étude, que l'implémentation des méthodes adjointes est tout à fait possible pour les modèles de la taille de ceux qui sont couramment employés en télédétection, et que les résultats obtenus avec des données artificielles permettent d'envisager l'assimilation d'observations satellitaires réelles. C'est ce que nous ferons dans le chapitre suivant.

**Chapitre 6 :**

**ASSIMILATION D'OBSERVATIONS D'UN DIFFUSIOMETRE-VENT  
SPATIOPORTE**

L'une des préoccupations de l'un de nos partenaires industriels est la connaissance en temps réel du champ de vent à la surface de l'océan et son évolution à court terme. Or les mesures de capteurs embarqués tels que les diffusiomètres-vent fournissent des estimations du vecteur vent en certains points de l'océan et donc des analyses, ainsi que des prévisions, peuvent être déduites en tout point par l'emploi des méthodes d'assimilation variationnelles. Pour ce faire, nous nous sommes appuyés sur les résultats qui ont pu être dégagés dans les chapitres précédents. Néanmoins, compte tenu de la nature applicative de l'étude, nous avons été amenés lors du développement de la procédure d'assimilation à faire certains compromis entre la théorie et les contraintes d'ingénierie

Ce chapitre décrit les solutions adoptées pour mener à bien cette étude dans le cadre de cette thèse en soulignant les compromis qui ont été faits. Il pourra être trouvé dans l'annexe 9.4 le résumé d'une communication faite sur ces travaux et qui peut être lue indépendamment des chapitres précédents.

## 6.1 Les données

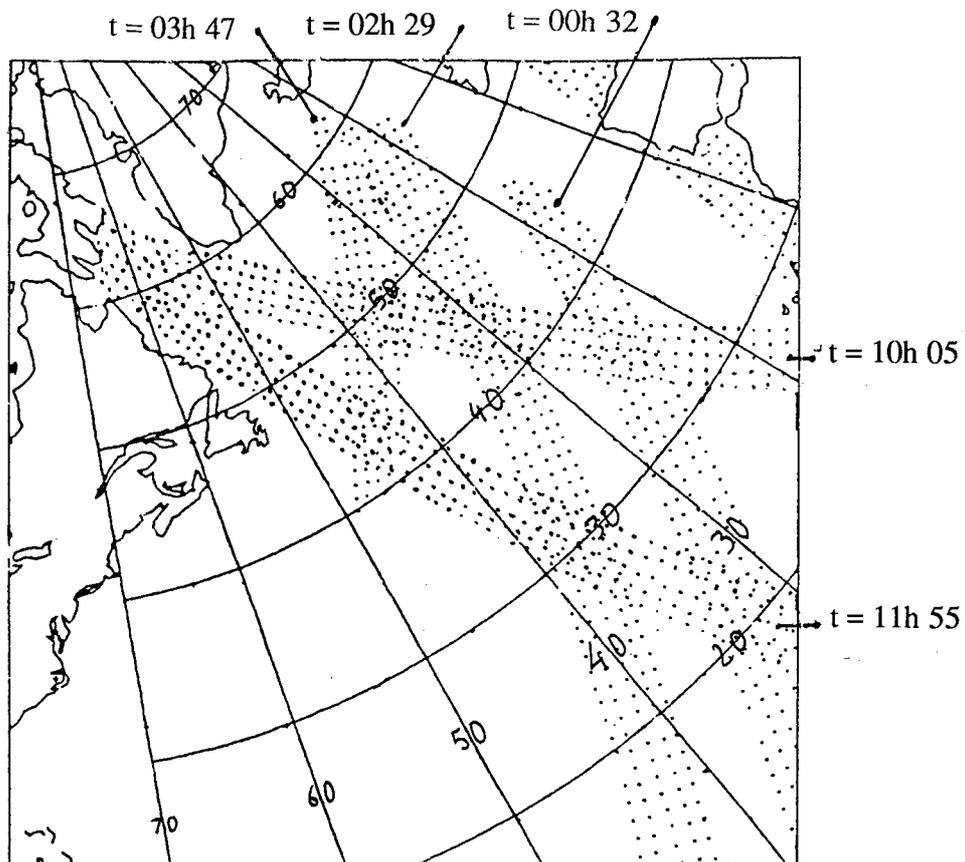
### 6.1.1 *Les observations satellitaires*

Nous utilisons les données du diffusiomètre A du satellite Seasat (données SASS) qui consiste en mesures, notées  $\sigma_0$ , de la réflexion par la surface de la mer d'un signal radar émis par l'appareil. Ces mesures sont ensuite inversées lors d'un pré-traitement pour obtenir des vitesses de vents. L'obtention du module du vecteur vitesse ne pose pas de problème majeur. Mais du fait de la non linéarité des équations, jusqu'à 4 directions sont possibles par valeur de  $\sigma_0$ . Afin de lever l'ambiguïté (opération de *dealiasing*) dans la détermination de la direction, il est nécessaire de faire appel à des sources d'information complémentaire (continuité des champs, climatologie, ...), ainsi que l'ont proposé plusieurs algorithmes, incluant des méthodes variationnelles (Hoffman 1984).

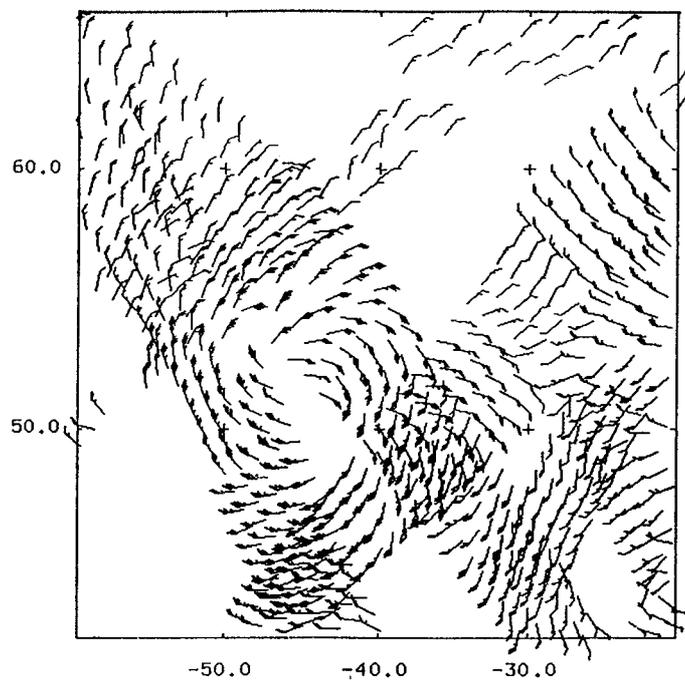
Dans cette étude nous ne nous sommes pas intéressés au problème de la levée d'ambiguïté, et n'avons considéré que des données déjà traitées selon l'algorithme de Wentz (1986). Nous disposons ainsi d'observations de vitesse de vent à la surface de la mer d'une résolution de 100 km, d'une précision de 10 % en module et 20 % en direction. Nous avons retenu une période de 12 heures le 11 septembre 1978, car ce jour là le satellite survola le centre d'une tempête (baptisée *Queen Elizabeth II*, ou *QE II*, du nom du navire qu'elle endommagea ce même jour) qui s'est abattue sur l'océan Atlantique nord avec une intensité sous-estimée dans les prévisions de la plupart des centres météorologiques nationaux. Ces données ont été sous-échantillonnées pour obtenir une résolution de 3° compatible avec les grilles de calcul. Nous obtenons ainsi 164 observations de vitesse réparties sous 5 orbites du satellite correspondant aux traversées de l'océan Atlantique nord lors de ses révolutions 1087, 1088, 1090, 1093 et 1094 effectuées respectivement à 0h 32, 2h 29, 3h 47, 10h 05 et 11h 55 GMT. La figure 6.1 montre l'emplacement géographique des observations selon l'instant de passage et la figure 6.2 montre les vitesses correspondantes ramenées sur la grille de calcul.

### 6.1.2 *Les analyses météorologiques*

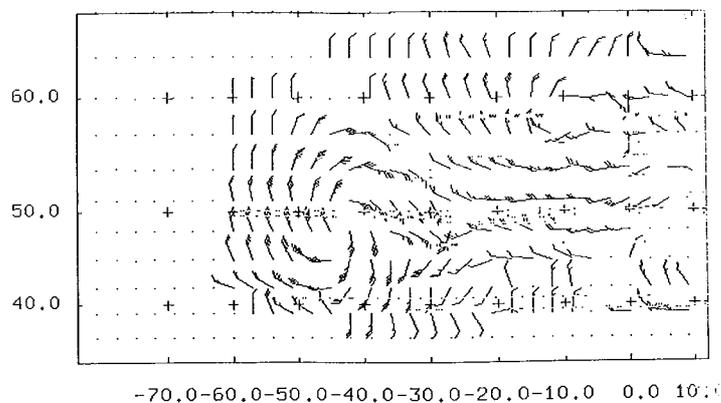
Dans les données, sont ajoutés aux observations satellitaires les champs de pression à la surface de l'océan issues d'analyses météorologiques publiées dans le *European Meteorological Bulletin* du 11 septembre à 0 heure GMT et 12 heures GMT (figure 6.3 et figure 6.4 respectivement), c'est-à-dire aux extrémités de la période d'assimilation. Le vent géostrophique (Harlan and O'Brien 1986) déduit de l'analyse de la situation à 0 heure est pris comme point de départ de la minimisation, alors que celui obtenu à partir de l'analyse à 12 heures sert de comparaison à la prévision résultant de l'assimilation variationnelle. Il faut toutefois être prudent avec ces deux données car le vent géostrophique n'est qu'une approximation du vent réel soufflant à la surface de l'océan et, de plus, les analyses dont il est issu sont inexactes, justement par manque d'observations satellitaires.



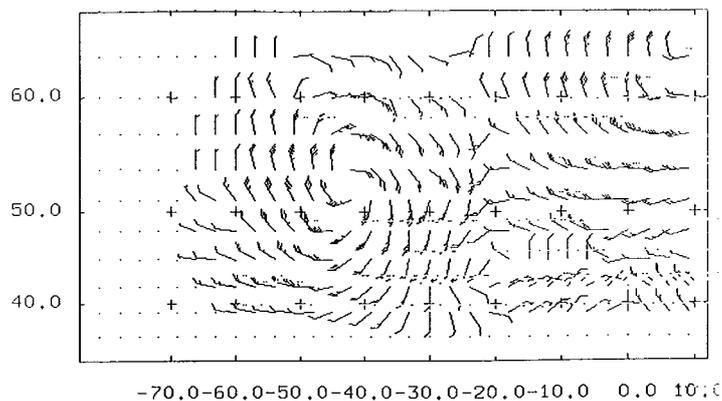
**Figure 6.1 :**  
Lieu des observations par orbite.



**Figure 6.2 :**  
Vecteurs vitesse vent issus des observations faites au cours  
des révolutions 1087, 1088, 1090, 1093 et 1094 du satellite Seasat.  
Une barre représente une vitesse de 5 m/s.



**Figure 6.3 :**  
 Vent géopotentiel déduit des analyses du  
 champ de pression en surface de l'océan  
 le 11 septembre 1978 à 0 heure GMT.  
 (European Meteorological Bulletin).



**Figure 6.4 :**  
 Idem Figure 6.3  
 le 11 septembre 1978 à 12 heure GMT.  
 Les figures 6.3-4 sont reprises avec les figures 6.5-6 pour comparaison.

## 6.2 Le modèle numérique

### 6.2.1 *L'Equation de la Vorticité*

Afin de simuler l'évolution de l'écoulement général de l'air à la surface de l'océan, nous avons d'abord fait l'hypothèse qu'il n'y a pas de mouvements verticaux et que ce fluide est incompressible, cela se traduit mathématiquement par la relation :

$$\nabla \cdot \mathbf{v} = 0 \quad 6.1$$

où  $\mathbf{v}(x, y)$  est le vecteur vitesse de l'air dans le plan horizontal  $(x, y)$  de l'écoulement et  $\nabla$  est l'opérateur gradient. Dans ces conditions, l'évolution du rotationnel de la vitesse, ou vorticité relative  $\xi$ , obéit à l'équation dite *Equation de la Vorticité* :

$$\partial \xi / \partial t + \mathbf{v} \cdot \nabla (\xi + f) = 0 \quad 6.2$$

où  $f$  est le paramètre de Coriolis. Compte tenu de la relation 6.1, il est possible de définir une fonction de courant  $\Psi$  et nous disposons alors du jeu de relations suivant :

$$\mathbf{v} = \nabla \times (\Psi \mathbf{k}) \quad 6.3$$

$$\xi = \mathbf{k} \cdot \nabla \times \mathbf{v} \quad 6.4$$

$$\xi = -\Delta \Psi \quad 6.5$$

où  $\mathbf{k}$  est le vecteur unitaire orthogonal au plan  $(x, y)$  de l'écoulement et  $\Delta$  est l'opérateur laplacien.

L'équation 6.2 peut être alors réécrite de la façon suivante

$$\partial \xi / \partial t = -\nabla \times (\Psi \mathbf{k}) \cdot \nabla (\xi + f) = -\mathbf{Jac}(\xi + f, \Psi) \quad 6.6$$

où  $Jac$  est le déterminant de la matrice jacobienne par rapport aux coordonnées rectangulaires  $x$  et  $y$  du plan de mouvement.

*Remarques sur les équations du modèle numérique*

L'Equation de la Vorticité qui est utilisée pour le modèle numérique ne modélise pas parfaitement les phénomènes qui ont lieu à l'interface l'océan-atmosphère, en particulier d'importants phénomènes convectifs sont totalement ignorés. Néanmoins nous avons retenu ce modèle principalement pour deux raisons :

- ne disposant pas de modèle numérique, il nous fallait trouver une équation pour laquelle nous puissions, dans le cadre de ces travaux, écrire le modèle direct et le modèle adjoint,
- la procédure est appelée à tourner en temps réel sur une station de travail du type Sun 4, ce qui imposait certaines limites au temps de calcul et donc un modèle assez simple.

L'Equation de la Vorticité en tant que modèle assimilateur a fait l'objet d'une étude très complète (Talagrand et Courtier 1987) et malgré sa non linéarité reste relativement simple à intégrer. C'est pourquoi nous avons retenu cette équation, tout en étant averti de l'erreur de modélisation qu'elle peut induire. Afin de pallier la faiblesse du modèle nous avons accordé une plus grande importance aux autres sources d'information que sont le first guess et les observations. Finalement les résultats obtenus se sont montrés satisfaisants pour l'usage ultérieur qui en est fait.

*6.2.2 Discrétisation, intégration et conditions limites.*

L'équation 6.6 est discrétisée par un schéma de différences finies et intégrée sur un domaine rectangulaire  $D$  en approximation  $\beta$ -plan. La résolution est de  $3^\circ$  (environ 333 km), ce qui nécessite 12 points en latitude et 33 points en longitude pour couvrir l'ensemble de l'Atlantique nord (respectivement  $33^\circ / 66^\circ$  Nord, et  $84^\circ$  Ouest /  $12^\circ$  Est).

L'algorithme utilisé pour l'intégration de l'équation est le suivant :

- a) calcul de  $\Psi_n$  par inversion de la relation  $\xi = -\Delta\Psi$ ,
- b) calcul de  $v_n$  par la relation  $v = k \times \nabla\Psi$ ,
- c) calcul de  $Jac(\xi_n + f, \Psi_n)$ ,
- d) calcul de  $\xi_{n+1}$  par la relation  $\partial\xi / \partial t = -Jac(\xi + f, \Psi)$ ,
- e)  $n = n+1$  et retour en a).

Cet algorithme suppose la connaissance de la vortacité  $\xi_0$  à l'étape initiale  $n = 0$ . En fait cette donnée n'est nécessaire qu'au tout début de la minimisation, et est acquise par la relation  $\xi_0 = k \cdot \nabla \times v_0$ . La vitesse initiale  $v_0$  étant fournie par l'analyse météorologique initiale qui sert de point de départ à la descente du gradient. Ici l'appellation de point de départ pour la vitesse est impropre dans la mesure où la vitesse n'est pas la variable d'état du modèle et, qui plus est, un champ de vitesse initiale  $v_0(x, y)$  ne définit pas sur la période d'intégration  $[0, T]$  une solution unique à l'équation 6.6. En effet, si  $v_0(x, y)$  définit une solution  $\xi(t)$ , alors  $v_0(x, y) + v$ , où  $v$  est une constante (champ uniforme) définit aussi cette même solution.

#### *Remarques sur la discrétisation*

L'inversion du laplacien à l'étape a) est opérée par la méthode de Gauss-Seigel en maintenant  $\Psi$  nulle sur les bords  $\partial D$  du domaine  $D$ . L'étape b) n'est pas nécessaire à l'intégration proprement dite mais elle sera utilisée aux instants d'assimilation et permet de visualiser les résultats en terme de vitesse. Pour le calcul du jacobien de l'étape c), nous avons adopté le schéma d'Arakawa (1966), ce dernier conserve l'antisymétrie, l'énergie cinétique et l'enstrophie (carré de la vortacité). A l'étape d) l'intégration temporelle est effectuée par pas de temps de 30 minutes suivant un schéma du type leap-frog, nécessitant donc 24 itérations pour intégrer le modèle entre 0 et 12 heures.

La discrétisation de l'équation en différences finies sur un domaine rectangulaire, pose le problème des conditions limites. En effet le système est du type hyperbolique et les conditions ne peuvent être imposées sur les bords. Nous avons songé à modifier le modèle en ajoutant à l'Equation de la Vorticité un terme de fermeture sous la forme d'un laplacien au carré qui aurait pour effet de modifier sa nature hyperbolique. Mais d'après les résultats du chapitre 4, nous savons que numériquement les termes de diffusion se prêtent mal à l'assimilation variationnelle.

Une autre possibilité serait d'incorporer les conditions aux bords ensemble avec les conditions initiales dans les variables de contrôle (Sokolowski and Zolésio 1990). Mais le contrôle simultané des deux variables est très délicat et surtout augmente le nombre du conditionnement de la matrice Hessienne (qui varie comme  $O(n^{2/d})$ , où  $d$  est la dimension du vecteur d'état, et  $n$  est la dimension de l'espace des contrôles), résultant en un taux de convergence plus faible que celui obtenu avec le contrôle uniquement effectué sur l'état initial (Zou *et al.* 1991). Aussi nous avons préféré fixer des valeurs aux bords qui, bien qu'inexactes, soient suffisamment réalistes, et c'est pourquoi dans cette étude nous avons réduit le nombre de calculs en imposant à la fonction de courant d'être constante et nulle sur les 4 bords du domaine. Physiquement cela signifie que le long des bords du domaine le vent y souffle parallèlement ; hypothèse qui est acceptable sur les bords sud et nord (tropicque et pôle) du domaine mais quelque peu exagérée pour les bords est et ouest (côte). Un compromis entre physique et calcul pourrait être de fixer de telles conditions là où elles sont réalistes (bords sud et nord) et de contrôler les conditions limites sur les deux autres bords.

### 6.3 La fonction de coût

Dans cette étude nous disposons d'observations de vitesse alors que la variable du modèle est la vorticité. L'opérateur qui permet de relier ces deux quantités est la combinaison des étapes a)

et b) décrites ci-dessus et que nous schématiserons dans le formalisme variationnelle générale par la matrice  $C = k \times \nabla \otimes \Delta^{-1}$ . Avec cette notation la fonction de coût adoptée est la suivante :

$$J = 1/2 \sum_{D, t_i} [ C\xi(t_i) - V_{obs} ]^t R^{-1} [ C\xi(t_i) - V(t_i)_{obs} ] \Delta t_i + \lambda/2 \sum_D [ \Delta\xi(t_0) ]^2 \quad 6.7$$

où  $D$  est le domaine spatial d'intégration,  $t_i = 0h\ 32, 2h\ 29, 3h\ 47, 10h\ 05, 11h\ 55$  sont les instants où les observations de vitesse  $V_{obs}$  sont disponibles,  $R$  est la matrice de variance-covariance des erreurs d'observation (donc positive et symétrique),  $^t$  dénote l'opération de transposition,  $\Delta t_i = t_{i+1} - t_i$  est le pas d'intégration temporelle et  $\lambda$  un coefficient de pondération entre les deux termes de la fonction de coût.

La fonction de coût définie dans cette étude diffère quelque peu de celle étudiée dans les chapitres précédents par son terme de pénalisation sous forme de carré du laplacien de la vorticit  initiale. En effet nous p nalisons ici uniquement les grandes valeurs du laplacien mais pas le terme de d riv es crois es. Concr tement cela n'apporte que peu de diff rence sur les effets de lissage cr es par ces deux types de p nalisation, alors que le codage du mod le se pr te beaucoup mieux   l' valuation du laplacien de la vorticit . Nous avons aussi remarqu , par ailleurs, que le terme de p nalisation n'a que peu d'influence sur le taux de convergence de la fonction de co t et   quelques it rations pr s la m me minimisation aurait pu  tre r alis e sans p nalisation. De m me l'influence de la matrice  $R$  est tr s faible et pratiquement les exp rimentations ont  t  men es en rempla ant  $R$  par la matrice identit .

## 6.4 Assimilation variationnelle

### 6.4.1 *Le modèle adjoint*

Compte tenu de la définition de la fonction de coût, l'opérateur dont nous cherchons à calculer le gradient est l'opérateur qui, à un champ de vorticité initiale  $\xi_0$  fait correspondre à chaque instant  $t$  un champ de vitesse  $v(t)$ . Du fait de la non linéarité des équations le calcul de l'adjoint de cet opérateur s'avère complexe (Talagrand and Courtier 1987) et nous donnons seulement les résultats. L'état adjoint  $P$  est solution du système d'équation :

$$-\partial P / \partial t = \Delta \text{Jac}(\Delta^{-1}P, \Delta^{-1}\xi) + \text{Jac}(\xi + f, \Delta^{-1}P) + C^t A(C\xi - V_{\text{obs}}) \quad 6.8$$

$$P(t=12h) = 0 \quad 6.9$$

Le système adjoint doit être intégré de façon rétrograde à partir de l'état de repos à 12 h, le terme de forcing n'étant en fait présent qu'aux instants où les observations sont disponibles. A la fin de l'intégration nous obtenons le gradient de la fonction de coût  $J$  par rapport à la vorticité initiale  $\xi_0$  :

$$\text{Grad}_{\xi_0}[J] = P(0) + \lambda \sum_D \Delta \xi(t_0) \quad 6.10$$

Comme nous l'avons déjà souligné les équations 6.8-9 sont les équations adjointes du système continu, néanmoins le schéma de discrétisation utilisé n'introduit pas d'erreur (voir § 3.6 ou Courtier and Talagrand 1987) et nous pouvons donc appliquer aux équations adjointes continues le schéma de discrétisation des équation directes.

### 6.4.2 *Minimisation et unicité de la solution*

Du fait de son application industrielle, la procédure d'assimilation doit être facilement transportable et il n'était donc pas question de faire appel à une fonction de minimisation d'une librairie scientifique, c'est pourquoi une nouvelle fois encore nous avons programmé

l'algorithme de la plus grande pente du gradient. Le minimiseur obtenu s'est révélé d'une bonne efficacité.

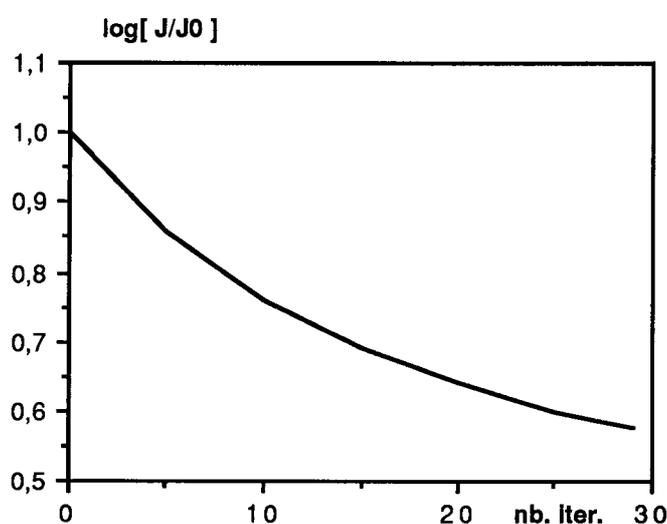
Il n'est pas non plus question ici de calculer la matrice Hessienne afin de nous renseigner sur l'unicité du minimum. Comme nous l'avons vu au paragraphe 6.2.2, il est évident que le minimum de la fonction de coût n'est pas unique si l'on considère le système par rapport à la variable vitesse initiale. Par contre, la démonstration de l'unicité de la fonction de coût par rapport à la vortacité initiale n'est pas facile et nous ne l'avons pas menée à son terme. En effet le faible nombre d'observations, 164, relativement au nombre de degrés de liberté, 384, pourrait indiquer l'existence de plusieurs minimas ; en réalité il semble qu'il existe un minimum local que nous retrouvons empiriquement pour une très grande plage de valeurs de vortacité initiale. Pour nous en assurer, nous avons répété la procédure d'assimilation pour différents points de départ, dont l'état de repos, et à chaque expérience nous avons retrouvé le même minimum.

## 6.5 L'assimilation séquentielle

Afin de comparer les deux procédures d'assimilation, variationnelle et séquentielle, nous avons comparé nos résultats à ceux obtenus par l'assimilation des mêmes données au moyen d'une méthode séquentielle (Cauneau et Chossat 1991). La technique utilise un filtre de Kalman passablement dégradé ; puisque l'erreur du modèle est modélisée par un bruit blanc gaussien et, lors de l'étape de prédiction, le calcul de la variance de l'erreur (diagonale de la matrice de variance-covariance) se réduit à une multiplication par un coefficient (3 pour une prévision à 12 heures), alors que les corrélations entre les différents points sont modélisées par des fonctions d'influence gaussiennes d'une longueur caractéristique de 400 km. La qualité des analyses des vitesses obtenues par cette assimilation est croissante durant la période d'assimilation, et suffisante pour la prévision en fin de période (Cauneau and Chossat 1991).

## 6.6 Résultats numériques

La procédure d'assimilation variationnelle des données SASS, telle que nous l'avons décrite, a été mise en oeuvre sur une station de travail. La descente du gradient est arrêtée aux premières apparitions d'oscillations ; ce qui, avec l'algorithme utilisé, a lieu après environ 30 itérations. La décroissance de la fonction de coût est tracée sur le graphique 6.1.

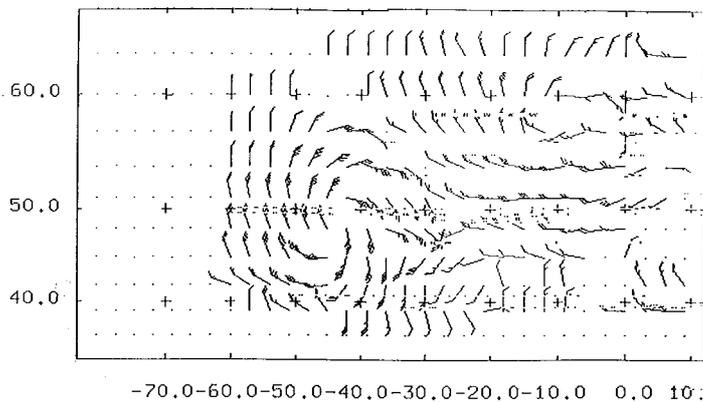


**Graphique 6.1 :**  
évolution du logarithme de la fonction de coût avec  
le nombre d'itération de la descente du gradient

L'état initial solution à ce stade de la minimisation est représenté sur la figure 6.5, aux côtés du point de départ (Figure 6.3). A partir de cette état initial solution, par intégration du modèle aux instants 0h 32, 2h 29, 3h 47, 10h 05 et 11h 55, nous avons produit les 5 analyses immédiatement disponibles après chaque passage du satellite. En fait des analyses peuvent être produites à n'importe quel instant  $t$ , il suffit pour cela d'intégrer l'état initial solution sur l'intervalle de temps  $[0, t]$  ; néanmoins nous avons retenu ces dates particulières car elles correspondent à l'arrivée de

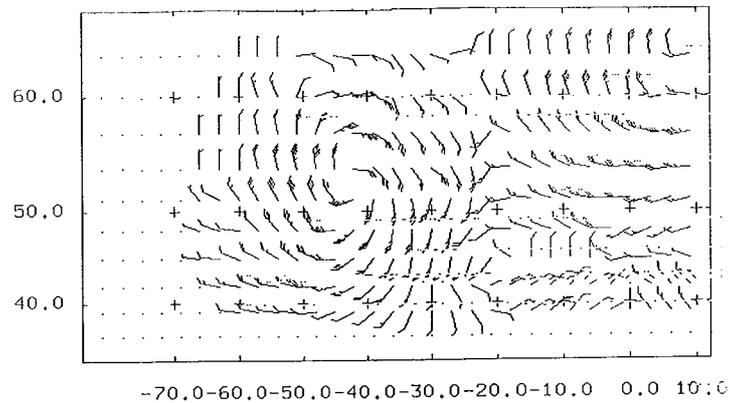
nouvelles informations, provoquant d'importantes modifications dans les analyses produites par l'assimilation séquentielle.

Nous avons aussi comparé la prévision issue du modèle à 12 heures (figure 6.6) avec l'analyse météorologique correspondante (figure 6.4). La comparaison n'est pas facile car l'analyse météorologiques n'est pas très fiable, et la forte erreur résiduelle de plus de 100 % (voir graphique 6.2) doit être impartie plus à cette analyse qu'à la prévision. De plus, dans le cas particulier de cette étude où des observations sont disponibles 5 minutes avant la fin de la période d'assimilation, la prévision normalement effectuée à 12 heures ne résulte que de quelques minutes d'intégration. Dans ces conditions, il est difficile de conclure sur les capacités de prévison du modèle assimilateur.



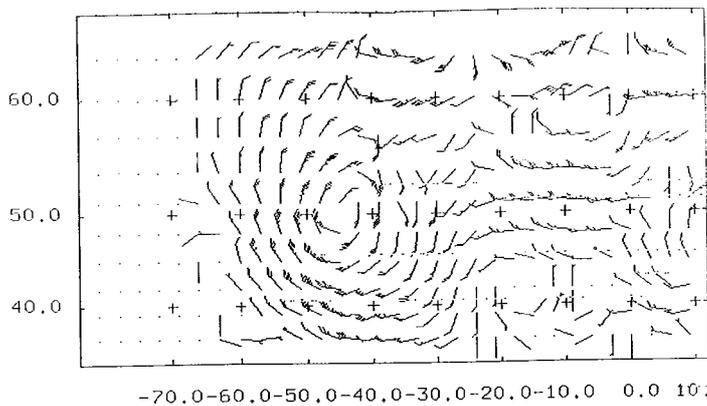
**Figure 6.3 :**

Vent géopotential déduit de l'analyse du champ de pression en surface de l'océan le 11 septembre 1978 à 0 heure GMT. (European Meteorological Bulletin).



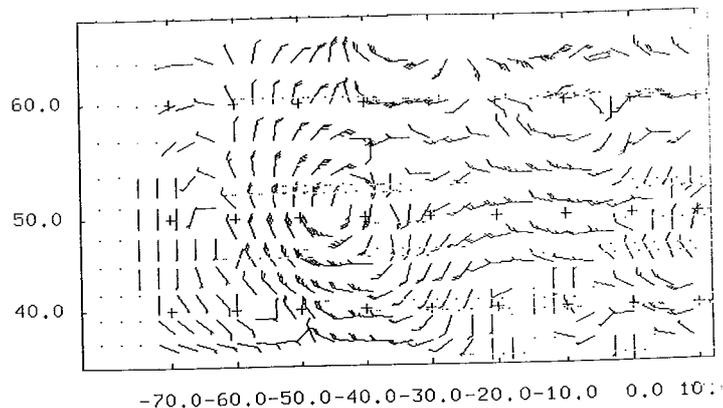
**Figure 6.4 :**

Vent géopotential déduit de l'analyse du champ de pression en surface de l'océan le 11 septembre 1978 à 12 heures GMT. (European Meteorological Bulletin).



**Figure 6.5 :**

Etat initial solution de la minimisation qui constitue l'analyse des champs de vent à  $t = 0$  heure.



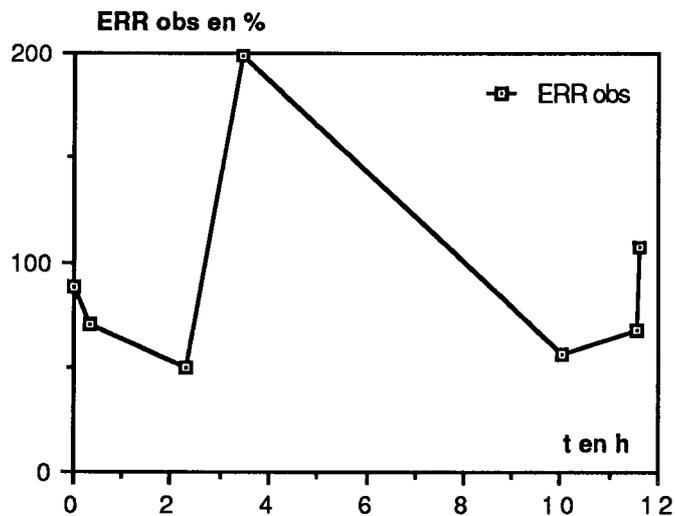
**Figure 6.6 :**

Prévision du vent de surface à  $t = 12$  heures obtenue par intégration de l'état initial solution de la minimisation.

D'une façon similaire à celle utilisé dans le chapitre 5, nous avons cherché à évaluer l'impact des observations sur les champs analysés. Nous avons calculé aux différents instants d'assimilation l'erreur résiduelle  $ERR_{obs}$  entre respectivement la solution du modèle et les observations, l'état initial et le first guess, ainsi qu'entre l'état final et l'analyse météorologique à 12 heures. Pour rappel cette valeur est calculée de la façon suivante :

$$ERR_{obs}(t_i) = \frac{\sum_D [V_{sol}(t_i) - V_{obs}(t_i)]^2}{\sum_D [V_{obs}(t_i)]^2} \quad 6.11$$

où  $t_i = 0$  heure, 0h 32, 2h 29, 3h 47, 10h 05, 11h 55 et 12 heures (instant initial, dates d'assimilation et instant final). Nous avons tracé sur le graphique 6.2 les variations de cette erreur résiduelle au cours du temps.



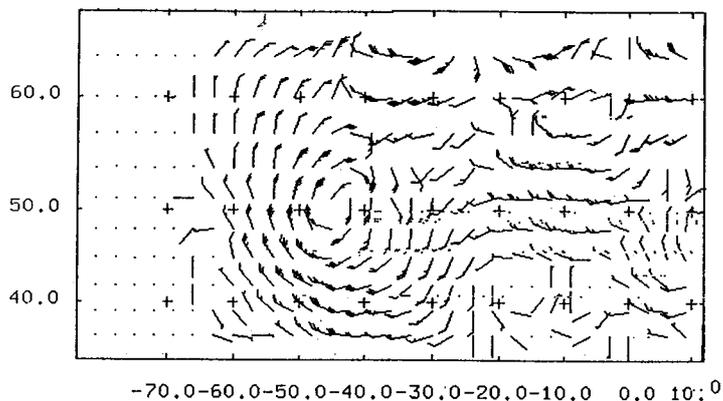
**Graphique 6.2 :**  
évolution dans le temps de l'erreur résiduelle ERR obs

## 6.7 Commentaires

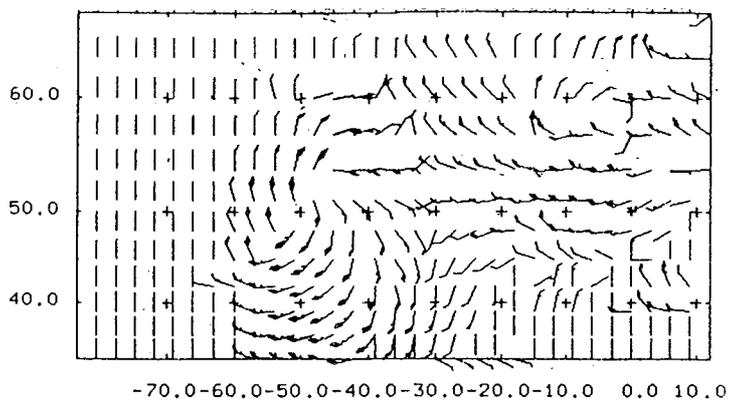
L'examen de la solution sur la figure 6.5 permet de déceler déjà à l'instant initial l'influence des observations qui lui sont pourtant postérieures. L'intégration rétrograde a effectivement propagé vers l'état initial l'information apportée par les observations des instants ultérieurs. Cette influence devient encore plus évidente si nous comparons l'analyse des champs de vitesse produite à 2h 09 par l'assimilation variationnelle avec celle obtenue par la méthode séquentielle (respectivement figures 6.7 et 6.8). Avec l'assimilation séquentielle, il est impossible de détecter la structure du vortex qui est contenu dans les jeux d'observations à venir ; alors que dans la solution variationnelle, le vortex apparaît clairement en position et en intensité. Nous retrouvons ainsi une propriété caractéristique de l'assimilation de données variationnelle qui est de distribuer l'information dans le temps.

Toutes les valeurs d'erreur résiduelle sont importantes (graphique 6.2) et ceci peut s'expliquer par la faiblesse de notre modélisation. Toutefois la forte erreur résiduelle qui apparaît à 3h 47 sur le graphique 6.2 étonne d'abord par la valeur qu'elle atteint, plus de 200 %, ensuite par sa contradiction avec l'allure générale de ces courbes. En effet l'erreur résiduelle qui découle d'une assimilation variationnelle est normalement la plus faible au milieu de la période d'assimilation (Talagrand 1992). Compte tenu de nos conditions limites, il est possible que cette forte valeur de l'erreur résiduelle soit due à un effet de bord (reflexion d'onde par exemple) qui se manifeste à cet instant là. Mais il est plus probable que ce soit le jeu d'observation qui soit erroné (vortex trop au nord et d'intensité insuffisante), comme nous pouvons le soupçonner d'après d'autres études réalisées sur cette tempête (Duffy et Atlas 1986, Gyakum 1983).

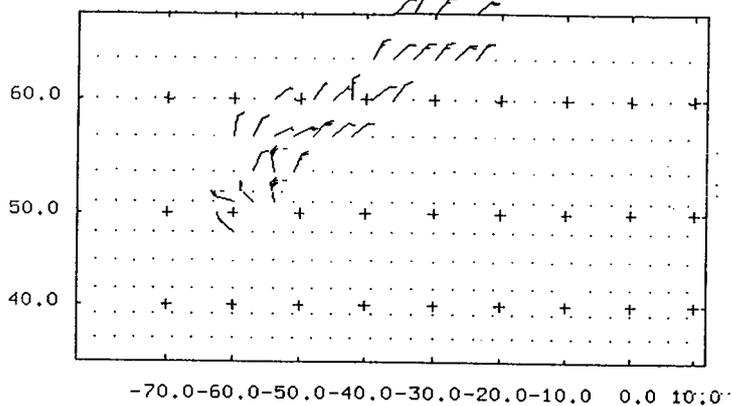
Afin d'examiner comment le modèle assimilateur répercute l'erreur d'observation, nous avons représenté la partie du champ analysé à 3h 47 (figure 6.10) qui correspond au lieu géographique des observations qui ont été assimilées à cette date là (figure 6.9).



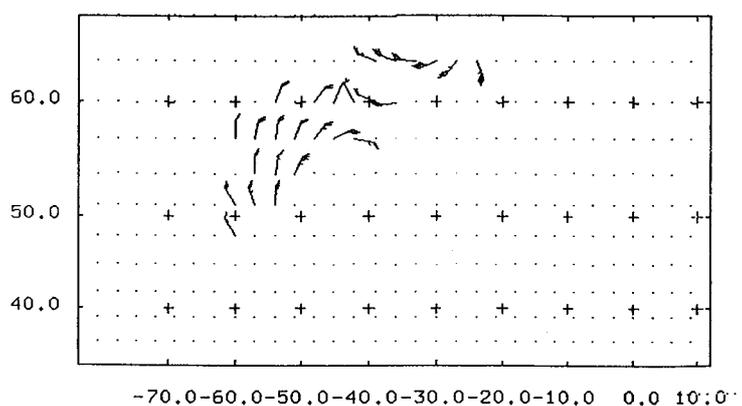
**Figure 6.7 :**  
Analyse des vents de surface par assimilation  
de données **variationnelle**  
à t = 2h 09.



**Figure 6.8 :**  
Analyse des vents de surface par assimilation  
de données **séquentielle**  
à t = 2h 09.



**Figure 6.9 :**  
Observations du satellite  
à t = 3h 47.



**Figure 6.10 :**  
Analyse des vents de surface par assimilation  
de données **variationnelle** à t = 3h 47.  
(vecteurs correspondants aux observations)

Après assimilation, nous pouvons remarquer que la direction des vecteurs vitesse a été fortement modifiée ; si bien que nous pouvons déjà deviner la forme du vortex, de plus l'intensité a été amplifiée surtout sur le quart sud-est des observations qui est plus proche du centre de la tempête. Sur ce jeu d'observation, nous découvrons un aspect important de l'assimilation quadri-dimensionnelle (espace + temps) : les observations ne sont pas introduites telles quelles dans le modèle, mais subissent des corrections afin de rester cohérente à la fois avec la dynamique du modèle et avec l'ensemble des observations restantes.

Des observations erronées peuvent venir principalement de deux causes : une défectuosité passagère du capteur à bord du satellite ou une mauvaise interprétation lors de la levée de l'ambiguïté (dealiasing) sur la direction de la vitesse, inhérente au traitement du signal diffusiométrique. Dans le premier cas il n'y a pratiquement rien à espérer des observations ; par contre dans le deuxième cas un nouveau traitement effectué en incorporant cette nouvelle information pourrait être efficace. Ainsi, ce genre d'assimilation de données pourrait être un moyen d'évaluer, voire de valider, le prétraitement des observations.

Une alternative serait aussi de chercher à corriger la levée de l'ambiguïté pendant l'assimilation en incorporant la direction du vecteur vent dans les variables de contrôle pour ce jeu d'observations particulier. Cela serait parfaitement réalisable compte tenu du fait que le nombre d'observations disponibles à cette date est relativement limité et qu'il n'y, en fait, que 4 directions possibles. Vu les imperfections de notre modèle, nous n'avons pas tenté l'expérience ; mais néanmoins le développement de procédure d'assimilation variationnelle intégrant le dealiasing dans le modèle afin d'assimiler directement le signal radar  $\sigma_0$  doit être envisagé, si nous voulons exploiter toutes les potentialités qu'offrent le diffusiomètre-vent maintenant embarqué sur le satellite européen ERS-1.

## 6.8 Conclusion

En vue d'applications industrielles, nous avons cherché à évaluer l'apport des observations issues d'un diffusiomètre-vent spatioporté dans la cartographie des champs de vent marin. A cette fin nous avons utilisé les méthodes adjointes pour assimiler directement des vitesses de vents déduites des observations d'un diffusiomètre dans un modèle numérique d'évolution de la vorticit  relative. L'Equation de la Vorticit  sous-tendant le mod le a  t  discr tis e sur un domaine rectangulaire en approximation  $\beta$ -plan et les observations de vitesse sont assimil es afin de produire des champs des vents r guli rement r partis sur une grille de  $3^\circ$  de r solution couvrant approximativement l'oc an Atlantique nord. Nous comparons les r sultats obtenus avec deux m thodes d'assimilation, variationnelle et s quentielle, sur des mesures de vitesse des vents de surface issues de 5 passages successifs effectu s par le satellite Seasat au dessus de l'oc an Atlantique nord le 11 septembre 1978 entre 0 heure et 12 heures GMT.

Si nous nous r f rons   l'objectif premier de cette  tude, nous pouvons conclure qu'il est possible   partir d'observations satellitaires diffusiom triques d'obtenir en temps r el des cartes des champs de vent marin avec une fiabilit  suffisante pour l'application industrielle pr vue (essentiellement forcing de mod les r gionaux de houle).

Mais au del  de l'application industrielle directe, ces travaux montrent la faisabilit  num rique des m thodes d'assimilation variationnelles. En d pit d'une mod lisation relativement simple, il est possible de minimiser une fonction de co t mesurant la distance entre les solutions du mod le et des observations r elles distribu es dans l'espace et dans le temps. Cette  tude montre de plus que les propri t s th oriques que nous  tions en droit d'attendre des m thodes variationnelles, sont effectivement conserv es lors de l'emploi de donn es satellitaires r elles. En effet, d'une part la comparaison de nos r sultats avec ceux obtenus dans des conditions similaires par une m thode s quentielle met en  vidence la capacit  qu'ont les m thodes variationnelles   r partir l'information

sur toute la durée de la période d'assimilation. D'autre part le couplage dynamique qui se crée lors de l'assimilation entre le modèle numérique et les observations permet de modérer l'influence des données dont les caractéristiques diffèrent exagérément de l'ensemble.

Ces deux qualités sont bienvenues lors de l'assimilation d'observations du diffusiomètre-vent embarqué. En effet aucun satellite ne peut, en un seul passage, couvrir avec une résolution suffisante une région aussi vaste que l'océan Atlantique nord. Une telle région est donc fractionnée en plusieurs zones, d'une superficie à peu près égale au champ d'observation du capteur diffusiométrique, qui seront successivement observées lors de leur survol par le satellite. Ainsi l'opérateur au sol reçoit des données issues de l'observation de chacune des zones tour à tour observées avec un décalage temporel, qui peut être significatif dans l'évolution des phénomènes physiques. L'assimilation séquentielle est capable de tenir compte de ces décalages temporels, mais uniquement pour les phénomènes physiques existant dans la zone observée. Seule l'assimilation variationnelle est capable de répercuter ce décalage globalement sur tous les phénomènes de la région.

De plus, comme nous l'avons vu, les données diffusiométriques nécessitent un pré-traitement qui peut parfois comporter certaines erreurs d'interprétation. Le couplage créé par les méthodes variationnelles entre la dynamique et les observations révèle les incohérences que peuvent introduire ce genre d'erreur.

Enfin, en raison des spécificités de cette étude, nous n'avons pas pu étudier les qualités prévisionnelles de la méthode d'assimilation variationnelles développée pour cette application. Néanmoins il est à attendre que la faiblesse de notre modélisation constituera une gêne importante. Pour étudier ces caractéristiques il nous faudrait disposer d'un modèle plus complet, mais donc, plus long à développer et à exécuter. Cette approche n'est toutefois pas immédiate, car les phénomènes qui existent dans la couche limite atmosphère-océan sont complexes et difficilement

modélisables numériquement. Bien évidemment le développement d'une procédure d'assimilation variationnelle pour un tel modèle, s'il était mis au point en serait considérablement alourdi et demanderait particulièrement d'attention. Cela dit, cette étude est aussi un premier pas dans ce sens.

**Chapitre 7 :**

**CONCLUSION**

En associant modélisation numérique et observation, l'assimilation de données est un moyen de traiter l'information géophysique d'une façon systématique. L'élaboration de nouvelles méthodes d'assimilation a ainsi permis à la météorologie d'accomplir d'importants progrès dans le domaine de la prévision. Cet exemple suscite beaucoup d'espoirs parmi les utilisateurs des données de télédétection. Il est vrai que la grande hétérogénéité de leur distribution spatio-temporelle est une spécificité des observations de télédétection qui complique leur exploitation. En effet, l'association de la modélisation numérique dans le traitement de ces données nécessite une procédure d'assimilation dite quadri-dimensionnelle, c'est-à-dire capable de traiter les données simultanément en espace et en temps.

Dans cette thèse, nous nous sommes intéressés aux méthodes qui permettent d'assimiler les données de télédétection. Elles relèvent de deux types principaux, les méthodes séquentielles issues du filtrage de Kalman et les méthodes variationnelles basées sur les équations adjointes du contrôle optimal. Au contraire des méthodes d'assimilation séquentielles, dont l'usage est courant, les méthodes variationnelles sont très récentes et ne sont pas encore utilisées de façon opérationnelle. C'est pourquoi nous les avons examinées afin de déterminer quelle pourrait être leur efficacité pour des applications de télédétection.

Une première étude pratique menée sur un modèle numérique linéaire d'Advection-Diffusion a montré la faisabilité numérique : il est possible de minimiser une fonction de coût mesurant la distance entre la trajectoire du modèle et des observations réparties à la fois dans le temps et dans l'espace. Cette minimisation est réalisée grâce à l'introduction des équations adjointes qui permettent de réduire le nombre de calculs nécessaires à l'évaluation du gradient de la fonction de coût par rapport aux variables d'entrée du modèle. Dans cette étude nous maîtrisons tous les paramètres, ce qui nous a permis de dégager les principales difficultés liées à la méthode et de les interpréter afin de les surmonter en vue d'applications à des cas réels.

La deuxième application, réaliste cette fois, utilise l'assimilation de données diffusiométriques comme méthode d'analyse en vue de cartographier les champs de vent à la surface des océans. Ces observations consistent en mesures de l'état de surface de l'océan et fournissent après un traitement approprié, une estimation de la vitesse des vents en cette surface. Les données utilisées proviennent du satellite Seasat qui, le 11 septembre 1978, a observé en plusieurs passages, une tempête particulièrement forte dans l'océan Atlantique nord. Sur une période de douze heures contenant cinq passages du satellite, nous avons assimilés les observations dans un modèle numérique non linéaire sur un domaine rectangulaire représentant approximativement l'océan Atlantique nord. Les champs de vents analysés par l'assimilation variationnelle ont été comparés à des résultats antérieurs obtenus sur la même situation par une assimilation séquentielle.

Cette seconde étude montre que le formalisme mathématique des méthodes d'assimilation variationnelles est non seulement bien établi mais qu'il peut de plus s'adapter aux contraintes de fonctionnement industriel, dont en particulier les limitations des ressources informatiques. La comparaison avec les méthodes d'assimilation séquentielles met en évidence la propriété de rétro-propagation de l'information dans le temps que possèdent les équations adjointes. Il en résulte des analyses des champs de vent sensiblement meilleures, surtout en début de la période d'assimilation et une grande cohérence dynamique entre les champs analysés et le modèle numérique.

La supériorité des méthodes variationnelles dans les analyses du début de la période d'assimilation s'explique par le fait qu'elles fournissent un état initial calculé à partir de l'ensemble des données. C'est grâce à cette qualité, qui est par ailleurs propre aux équations adjointes, que nous avons pu résoudre le problème qui était posé par le modèle d'Advection-Diffusion. En effet la procédure présentée s'apparente à une inversion, puisque, pour retrouver l'état initial de notre système, nous avons en quelque sorte remonté de douze heures dans le temps alors que paradoxalement l'assimilation de données a été originellement introduite pour les besoins de la

prévision. Ces deux applications, presque contradictoires, soulignent la souplesse des méthodes d'assimilation variationnelles.

Les propriétés que nous venons d'énumérer semblent faire des méthodes adjointes un outil bien adapté au traitement des données de télédétection. Toutefois quelques remarques s'imposent.

Le coût numérique est élevé et, pour de gros systèmes, il est impérieux d'utiliser des algorithmes de descente qui soient capables de reconditionner le système. Ceci implique la connaissance de la matrice Hessienne. Normalement le calcul de cette matrice n'est pas aisé et nécessite également de nombreuses opérations. Toutefois de récents travaux ont montré qu'il était possible de se contenter d'estimer cette matrice sans la calculer explicitement (Gauthier 1992, Rabier et Courtier 1992). Ceci devrait faciliter la mise en oeuvre des méthodes adjointes dans le futur.

Bien que peu rencontrée dans la pratique, la possibilité de minima multiples existe et pour lever cette ambiguïté la connaissance (ou l'estimation) de la matrice hessienne est là aussi nécessaire.

Contrairement au filtre de Kalman, le formalisme variationnel ne permet pas de prendre en compte les imperfections du modèle numérique ni ne produit la matrice de variance-covariance d'erreur qui accompagne l'estimation des champs.

Si donc les méthode adjointes offrent des possibilités intéressantes, il ne faut pas pour autant négliger les méthodes d'estimation statistiques, car vraisemblablement le système d'assimilation le plus performant sera celui qui combinera ces deux approches (Thépaut 1992).





**Chapitre 8 :**

**REFERENCES BIBLIOGRAPHIQUES**

Arakawa, A., 1966. Computational design for long term numerical integration of the equations of fluid motion: two dimensional incompressible flow. Part I.-*Journal of Computational Physics*, 1, p. 119-143.

Balgovin, R., A. Dalcher, M. Ghil and E. Kelnay, 1983. A stochastic-dynamic model for the spatial structure error statistics.-*Mon. Wea. Rev.*, 115, p. 273-296.

Bates, D. M., M.J Lindstrom, G. Wahba and B. S. Yandel, 1986. *GCVPACK routines for generalised cross validation*.-Technical report n 775, departement of statistics, University of Wisconsin, Madison WI, USA.

Bennett, A. F. and R. N. Miller, 1990. Weighting initial conditions in variational Assimilation Schemes.-*Mon. Wea. Rev.*, 119, p. 1098-1102.

Bertsekas, D. P., 1982. *Constrained optimization and Lagrange multiplier methods*. Academic press, London, 395 p..

Budgell, N.P., 1986. Non linear data assimilation for shallow water equations in branched channels.-*J. Geophys. Res.*, 91, p. 10633-10644.

Cacuci, D. G. and M. C. G. Hall, 1984. Efficient estimation of feedback effects with application to climate models.-*J. Atmos. Sci.*, 41, p. 2063-2068.

Cauneau, F. and P. Chossat, 1991. Atmospheric mesoscale vortices detection and tracking with wind scatterometers.-Proceedings of the 11th de l'EARSel Symposium, Graz, Autriche.

Cirillo, M. C. and A. A. Poli, 1992. An intercomparison of semiempirical diffusion models under low wind speed, stable conditions.-*Atmospheric Environment*, Vol. 26A, No 5, p. 765-774.

Cohn, S. E. and D. F. Parrish, 1991. The behavior of forecast error covariances for a Kalman filter in two dimensions.-*Mon. Wea. Rev.*, 119, p. 1757-1785.

Courtier, Ph., 1987. *Application du contrôle optimal à la prévision numérique en météorologie*.-Thèse de doctorat de l'Université Pierre et Marie Curie, Paris, 275 p..

Courtier, Ph. and O. Talagrand, 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation, II: numerical results.-*Tellus*, 42A, p. 1329-1347.

Courtier, Ph. and O. Talagrand, 1990b. Variational assimilation of meteorological observations with the adjoint shallow water equation.-*Tellus*, 42A, p. 531-549.

Courtier, P. and O. Talagrand, 1990. Assimilation of meteorological observation, a review of present problems.-In *Proceedings of International Symposium on Assimilation of Observations in Meteorology and Oceanography*, World Meteorological Organization, Geneva, Switzerland, p. 12-15.

Courtier, Ph., J. N. Thépault. and O. Talagrand, 1990. 4 dimensional data assimilation using the adjoint of a primitive equation model.-In *Proceedings of International Symposium on Assimilation of Observations in Meteorology and Oceanography*. World Meteorological Organization, Geneva, Switzerland, p. 337-340.

Cressman, G. P., 1959. An operational objective analysis system.-*Mon. Wea. Rev.* 87, p. 367-375.

De Larminat, P. and Y. Thomas, 1977. *Automatiques des systèmes linéaires : 2 identification*. Flamarion Sciences, Paris, 227 p..

- Duffy, D. G. and R. Atlas, 1986 The impact of Seasat-A scatterometer Data on the Numerical Prediction of the Queen Elizabeth II storm.-*Journal of Geophysical Research*, 91-C2, p. 2241-2248.
- Faure P., 1986. *Notes d'optimisation*. Cours de l'Ecole Polytechnique, Paris, 175 p..
- Gauthier, P., 1992. Chaos and quadri-dimensional data assimilation: a study based on the Lorenz model.-*Tellus*, 44A, p. 2-17.
- Ghil M., S. Cohn, J. Tavantzis, K. Bube and E Isaacson, 1981. Applications of estimation theory to numerical weather prediction. In *Dynamic Meteorology: Data Assimilation Methods*, L. Bengtsson, M. Ghil and E. Kallen eds. Springer-Verlag, New York, p. 139-124.
- Ghil, M. and P. Malanotte-Rizzoli, 1991. Data assimilation in meteorology and oceanography.-*Adv. in Geophys.*, 33, p.141-226.
- Gill, A. E., W. Murray and M. H. Wright, 1982. *Practical optimization*.-Academic press, London.
- Gyakum, J. R., 1983. On the Evolution of the QE II Storm. I: Synoptic Aspects.-*Month. Weather. Rev.*, Vol 111, No 6, p. 1137-1155.
- Harlan, J. Jr. and J. O'Brien, 1986. Assimilation of scatterometer winds into surface pressure fields using a variational method.-*J. Geophys. Res.*, 11, p. 7816-7836.
- Hoffman, R. N., 1984. SSAS wind ambiguity removal by direct minimization. Part II: use of smoothness and dynamic constraints.-*Mon. Wea. Rev.*, 112, p. 1829-1852.
- Hoffman, R. N., 1986. A four dimensional analysis exactly satisfying equations of motion, *Mon. Wea. Rev.*, 114, p. 388-397.

- Johnson D. R., Bates J. R., Brasseur G. P., Hollingsworth A., Miyakoda K., Rasmusson E., Sarachik E. S. & Warner T. T., 1991. *Four dimensional model assimilation of data: a strategy for the earth system sciences*. Rapport of panel on model assimilated data sets for atmospheric and oceanic research. National Research Council. 78 p..
- Kalman, R. E., 1960. A new approach to linear filtering and prediction problem.-*Trans. ASME, J of Basic Eng.*, 82D, p. 34-45.
- Lacarra, J. F. and O. Talagrand, 1988. Short range evolution of small perturbations in a barotropic model.-*Tellus*, 40A, p. 81-95.
- Le Dimet, F. X. and O. Talagrand, 1986. Variational algorithm for analysis and assimilation of meteorological observations : theoretical aspects.-*Tellus*, 38A, p. 97-110.
- Le Dimet, F. X., I. M. Navon, and X. Zou, 1991. *Incomplete observations and control of gravity waves in variational data assimilation part I: theoretical aspects*.-Report FSU-SCRI-91-91, The Florida State University, Tallahassee, Florida, USA, 26 p..
- Lewis, J. M. and J. C. Derber, 1985. The use of adjoint equations to solve a variational adjustment problem with convective constraints.-*Tellus* 37A, p. 309-322.
- Lions, J. L., 1969. *Contrôle optimal des systèmes gouvernés par des équations aux dérivées partielles*. Dunod, Paris, France, 426 p..
- Lions, J. L., 1984. *Analyse numérique*. Cours de l'Ecole Polytechnique, Paris, 179 p..
- Long, B. R., W. C. Thacker and S. W. Hwang, 1990. Fitting a general ocean circulation model to observations: a progress report.-In *Proceedings of International Symposium on Assimilation*

*of Observations in Meteorology and Oceanography*. World Meteorological Organization, Geneva, Switzerland, p. 65-70.

Lorenc, A., 1986. Analysis methods for numerical weather prediction.-*Q. J. R. Meteorol. Soc.*, 114, p. 1174-1194.

Martchuk, G. I., 1974. Formulation of the theory of perturbation for complicated model. *Appl. Math. Optimization*, 2, p.1-33.

Méaille, R. and L. Wald, 1990. Using geographical system and satellite imagery within a numerical simulation of an regional urban growth.-*Int. J. Geographical Information Systems*, Vol. 4, No. 4, p. 445-456.

Miller R. N., 1986. Toward the application of the Kalman filter to the regional open ocean model. *J. Phys. Oceanogr.* , 16, p. 72-86.

Navon, I. M. and D. Legler, 1987. Conjugate gradient methods for large scale minimization in meteorology.-*Mon. Wea. Rev.*, 15, p. 1479-1502.

Petri, H. A. Ebel, H. Hass, H. Jakobs and M. Memmesheier, 1990. Assimilation of Environmental data.-*In proceedings of international symposium on assimilation of observations in meteorology and oceanography*, WMO, Geneves, Swizerland, p. 196-198.

Press, W. H., B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, 1990. *Numerical Recipes in C* - Cambridge University Press, Cambrige, MA, USA.

Rabier, F. and Ph. Courtier, 1992. Four-dimensional assimilation in the presence of baroclinic instability.-*Q. J. R. Meteorol. Soc.*, 118, p. 649-672.

Shiau, J. J. G. Wahba and D. R. Johnson, 1986. Partial spline models inclusion of tropopause and frontal boundary information in otherwise smooth two and three dimensional objective analysis. -*Journal of Atmosphere and Oceanic Technology*, vol 3, no 4, decembre 1986, p. 714-725.

Sarmiento, J. L., 1983. A simulation of bomb tritium entry onto the Atlantic Ocean.-*J. Phys. Oceanogr.*, 13, p. 1924-1939.

Sokolowski, J. and J. P. Zolésio, 1990. *Introduction to shape optimization and shape sensitivity analysis*.-Springer-Verlag, 250 p..

Souli, M. et J. P. Zolésio, 1991. Le contrôle de domaine dans le problème de résistance de vague non linéaire en hydrodynamique- *Ann. sc. math. Québec*, 15 (2), p. 203-214.

Talagrand, O. and Ph. Courtier, 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation I: Theory.-*Q. J. R. Meteorol. Soc.*, 113, p. 1311-1328.

Talagrand, O., 1988. Four dimensional variational assimilation.-*Proceedings of the Seminar on Data Assimilation and the Use of Satellite Data*, Volume 2, p. 1-30. Centre européen de prévisions météorologiques à moyen terme, Reading, Grande-Bretagne.

Talagrand, O., 1991. The use of adjoint equations in numerical modeling of the atmospheric circulation.-In *Proceedings of the Workshop on Automatic Differentiation Algorithms: Theory, Implementation and Applications*. Society for Industrial and Applied Mathematics, Breckenridge, CO, USA.

Talagrand, O., 1992. Data Assimilation Problems. A paraître dans *Tellus A*.

Thacker, W. C., 1988. Fitting models to data by enforcing spatial and temporal smoothness.-*J. Geophys. Res.*, 93, p. 1655-1665.

Thacker, W. C. and R. B. Long, 1988. Fitting dynamics to data.-*J. Geophys. Res.*, 93, p. 1227-1240.

Thépaut, J. N., 1992. *Application des méthodes variationnelles pour l'assimilation quadridimensionnelle des observations météorologiques*.-Thèse de Doctorat, Université de Paris 6, Paris, 171 p..

Thépaut, J. N. and Ph. Courtier, 1992. Four dimensional variational data assimilation using the adjoint of a multilevel primitive equation model.-*Q. J. R. Meteorol. Soc.*, 117, p. 1225-1254.

Villalobos, M. and G. Wahba, 1987. Inequality constrained multivariate smoothing splines with application to the estimation of the posterior probabilities.-*Journal of the American Statistical Society*, Vol. 82, no 397, p. 239-248.

Wahba, G. and J. Wandelberger, 1980. Some news methods for variational objective analysis using splines and cross validation.-*Mon. Wea. Rev.*, 108, p. 1121-1143.

Wahba, G., 1982. Variational methods in simultaneous optimum interpolation and initialization.-in *The interaction between objective analysis and initialization*. Proceedings of the 14th Stansted seminar. Ed Wiliamson D. (Publication in Meteorology, 127). Mc Gilll University Montreal, p. 178-185.

Wahba, G., 1987. Three topics in ill-posed problems.-*Inverse Ill-Posed Problems*, p. 37-51, Academic press, London, England.

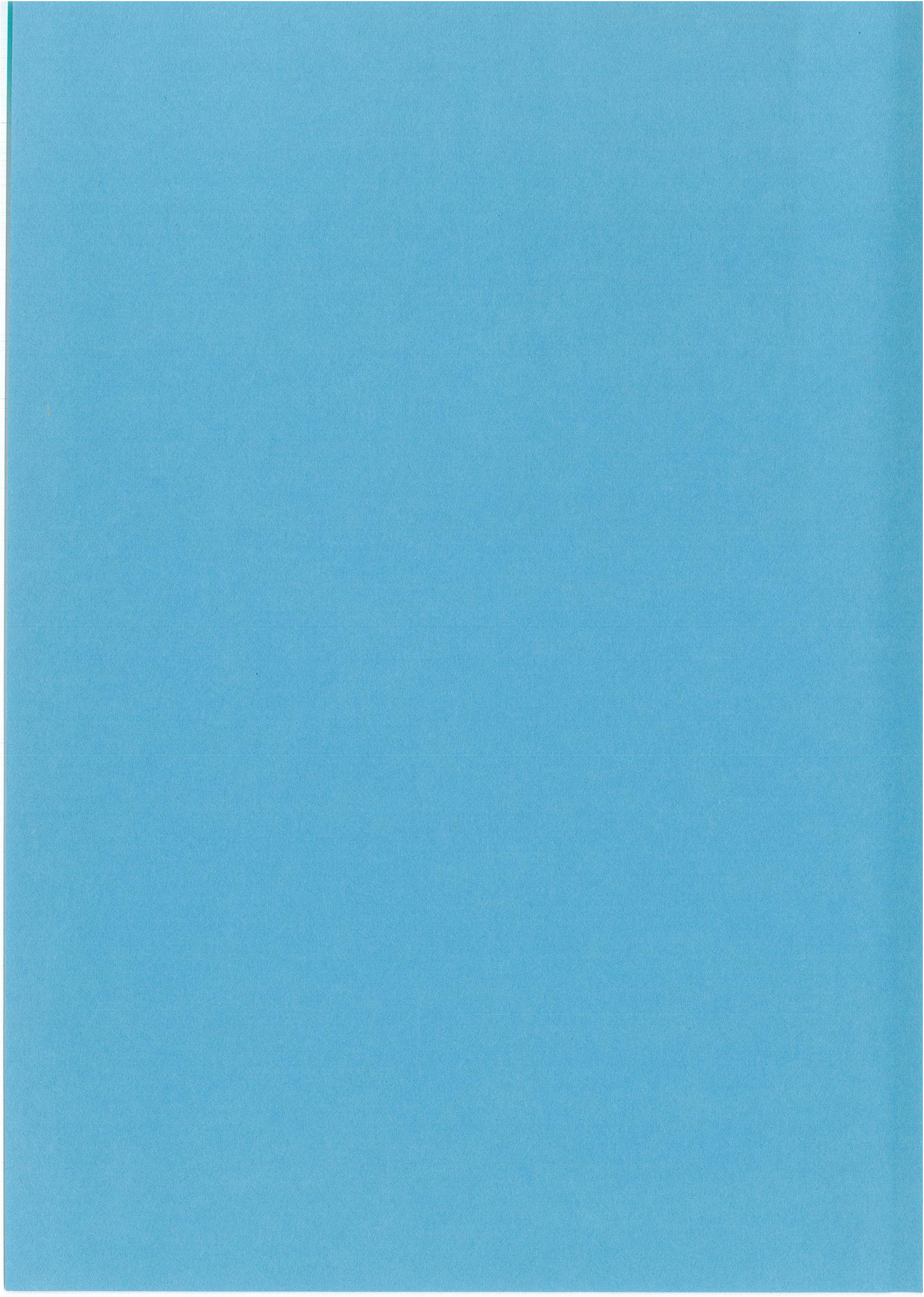
Wahba, G., 1989. Regularization and cross validation methods for non linear, implicit, ill-posed inverse problem.-*Proceedings of the 7th International Mathematica Geophysics Seminar*, Free Universitat of Berlin, Bjorn Ursin Ed., p. 3-13.

Wentz, F. J. and L. A. Mattox, 1986. New algorithms for microwave measurements of ocean winds : application to Seasat and the special Sensor Microwave Imager.-*Journal of Geophysical Research*, 91-C2, p. 2289-2307.

Wunsch C., 1987. Using transient tracers: the regularisation problem.-*Tellus*, 39B, p. 477-492.

Wunsch, C., 1988. Transient tracer as a problem in control theory.-*J. of Geophys. Res.*, vol. 97, C7, p. 8099-8110.

Zou, X., I. M. Navon, and F. X. Le Dimet, 1991. *Incomplete observations and control of gravity waves in variational data assimilation part II: applications and numerical results*.-Report FSU-SCRI-91-92, The Florida State University, Tallahassee, Florida, USA, 43 p..



**Chapitre 9 :**

**ANNEXES**

### 9.1 Méthodes adjointes

Supposons un système dont les équations d'évolution aient été discrétisées par rapport aux variables d'espace, permettant ainsi de décrire sur un intervalle de temps  $[0, T]$  l'état de ce système comme solution d'une équation différentielle ordinaire :

$$dX/dt = F(X, V) \quad 9.1$$

$$X(0) = U \quad 9.2$$

où  $X$  représente l'état du système et appartient à un espace vectoriel  $\mathbf{X}$  de dimension  $n$ ,  $F$  est une fonction (linéaire ou non linéaire) de  $\mathbf{X}$  dans  $\mathbf{X}$ ,  $V$  est l'état initial et  $U$  est un ensemble de paramètres ajustables du modèle. Nous supposons qu'à chaque couple  $(U, V)$ ,  $F$  fait correspondre une unique solution  $X(t)$  sur l'intervalle  $[0, T]$ .

Le système est observé sur la période  $[0, T]$ . Soit  $X_{\text{obs}}(t)$  le vecteur constitué des observations. Il appartient à l'espace  $\mathbf{O}$  de dimension  $m$ . Les espaces  $\mathbf{X}$  et  $\mathbf{O}$  peuvent être différents car il existe des points d'observations qui ne coïncident pas avec les points de la grille de calcul du modèle, mais aussi parce que les variables du système peuvent ne pas être directement observables. Néanmoins nous supposons qu'il existe un opérateur linéaire  $C$  de  $\mathbf{X}$  dans  $\mathbf{O}$  qui permet de comparer les observations à la solution du modèle au travers d'une fonction de coût  $J$  définie par

$$J(U, V) = 1/2 \int_{[0, T]} H(X, t) dt \quad 9.3$$

$H$  est une fonction de  $X$  et de  $t$  et donc une fonction implicite de  $U$  et  $V$ . Dans le cas particulier de la formule 9.3, où  $H$  est une fonction quadratique de l'erreur entre les observations et le modèle,  $J$  devient :

$$J(U, V) = 1/2 \int_{[0, T]} H(X, t) = 1/2 \int_{[0, T]} \| W(C X(U, V, t) - X(t)_{\text{obs}}) \|^2 dt \quad 9.4$$

Ici nous avons choisi d'écrire explicitement dans l'expression de la fonction de coût la matrice  $W$  qui pondère les différences entre elles. Idéalement  $W$  devrait être l'inverse de la matrice  $R$  de variance-covariance des observations et  $W$  peut aussi varier dans le temps. Comme  $W$  est une matrice positive, elle peut être incorporée dans la norme  $\| \cdot \|$  qui est associée au produit scalaire  $\| X \|^2 = \langle X, X \rangle = X^T W X$  qui est défini sur l'espace des observations  $O$ .

Le problème de contrôle induit par l'assimilation de données consiste à déterminer les valeurs  $U^*$  et  $V^*$  des variables de contrôle telles que la solution des équations 9.1-2 minimise la fonction de coût  $J$ .  $U^*$  et  $V^*$  doivent vérifier :

$$J(U^*, V^*) \leq J(U, V) \quad \forall (U, V) \in V \quad 9.5$$

$V$  est l'espace des contrôles admissibles, c'est à dire l'ensemble des valeurs permises pour les variables de contrôle.

Si  $F$  est différentiable par rapport à  $X$ , il existe des méthodes d'optimisation performantes, basées sur le gradient de la fonction de coût. Il est donc nécessaire de calculer les gradients  $J_U(U, V)$  et  $J_V(U, V)$  de fonction de coût par rapport aux variables de contrôle  $U$  et  $V$ . Pour cela, calculons les dérivées de Gateaux de  $J$  dans la direction  $H = (H_U, H_V)$ . La définition de la dérivée de Gateaux est :

$$J'(U, V, H) = \lim_{\theta \rightarrow 0} 1/\theta \{ J(U + \theta H_U, V + \theta H_V) - J(U, V) \} \quad 9.6$$

On obtient :

$$J'(U,V,H) = \int_{[0, T]} \langle CX(U,V) - X_{\text{obs}}, CY(H) \rangle dt \quad 9.7$$

Dans cette expression, Y est la dérivée de Gâteaux dans la direction H de la fonction qui, à (U, V) associe X(U, V). Y peut être obtenu comme solution du système différentiel suivant :

$$dY/dt = [\partial F/\partial X]Y + [\partial F/\partial U]H_u \quad 9.8$$

$$Y(0) = H_v \quad 9.9$$

où  $[\partial F/\partial \cdot]$  représente le Jacobien de F par rapport à la variable indiquée.

Pour obtenir les gradients  $J_u$  et  $J_v$  à partir de l'équation 9.7, nous introduisons l'état adjoint P dans  $X'$ , espace dual de X. Effectuons le produit scalaire de P avec 9.8 puis intégrons par parties de 0 à T :

$$\int_{[0, T]} \langle dY/dt, P \rangle dt = \int_{[0, T]} \langle [\partial F/\partial X]Y, P \rangle + \langle [\partial F/\partial U]H_u, P \rangle dt \quad 9.10$$

Soit

$$\begin{aligned} \langle Y(T), P(T) \rangle - \langle Y(0), P(0) \rangle - \int_{[0, T]} \langle Y, dP/dt \rangle dt \\ = \int_{[0, T]} \langle Y, [\partial F/\partial X]^t P \rangle dt + \int_{[0, T]} \langle H_u, [\partial F/\partial U]^t P \rangle dt \end{aligned} \quad 9.11$$

Ou encore :

$$\langle Y(T), P(T) \rangle - \langle H_v, P(0) \rangle = \int_{[0, T]} \langle Y, dP/dt + [\partial F/\partial X]^t P \rangle dt$$

$$+ \int_{[0, T]} \langle H_u, [\partial F / \partial U]^t P \rangle dt \quad 9.12$$

Définissons alors l'état adjoint comme solution de l'équation différentielle suivante

$$- dP/dt = [\partial F / \partial X]^t P + C^t(CX - X_{obs}) \quad 9.13$$

$$P(T) = 0 \quad 9.14$$

En reportant 9.13-14 dans 9.12 nous déduisons :

$$\begin{aligned} - \langle H_v, P(0) \rangle = & - \int_{[0, T]} \langle Y, C^t X(U, V) - X_{obs} \rangle dt \\ & + \int_{[0, T]} \langle H_u, [\partial F / \partial U]^t P \rangle dt \end{aligned} \quad 9.15$$

En comparant 9.15 avec 9.7 nous reconnaissons dans le premier terme intégral l'expression  $J'(U, V, H)$ , soit

$$- \langle H_v, P(0) \rangle = - J'(U, V, H) + \int_{[0, T]} \langle H_u, [\partial F / \partial U]^t P \rangle dt \quad 9.16$$

Ce qui, d'après la définition du gradient, implique :

$$J_v(U, V) = P(0) \quad 9.17$$

$$J_u(U, V) = \int_{[0, T]} [\partial F / \partial U]^t P dt \quad 9.18$$

Le gradient de la fonction est obtenu par intégration rétrograde du système 9.17-18.

## 9.2 Les Multiplieurs de Lagrange

Dans notre cas où nous voulons minimiser une fonctionnelle  $J$  vérifiant les équations d'un modèle numérique de matrice  $S$ , le Lagrangien se construit de la façon suivante. Si, en reprenant les notations du chapitre 2,  $X = (x_1, \dots, x_N)$  est un vecteur constitué des  $N$  états ( $N$  vecteurs de dimension  $n$ ) du système qui sont tous contraints de vérifier les équations du modèle numérique  $S$ , la contrainte s'écrit alors  $F(X) = x_k - Sx_{k-1}$ ,  $1 \leq k \leq N$ . Et si, de plus, nous disposons d'observations aux instants  $k = 1, \dots, N$ , il est possible d'associer  $N$  vecteurs  $\Lambda_k$  de dimension  $n$ , appelés multiplicateurs de Lagrange pour construire le Lagrangien associé au problème qui s'écrit :

$$L(X, \Lambda) = J(X) + \sum_{1 \leq k \leq N} \Lambda_k^t [x_k - Sx_{k-1}]$$

Il est démontré que le minimum du problème d'optimisation ci-dessus correspond à un point stationnaire du Lagrangien  $L(x, \Lambda)$  considéré comme fonction de  $x_k$  et de  $\Lambda_k$ ,  $k = 1, \dots, n$ .

Ainsi lorsque nous écrivons que les  $2N + 1$  dérivées partielles du Lagrangien sont nulles nous retrouvons les  $N$  équations du système direct ainsi que  $N$  équations adjointes permettant de déterminer les valeurs de  $X$  et  $\Lambda$  qui correspondent au point stationnaire du Lagrangien. Ce point stationnaire permet de déduire le minimum de la fonction de coût  $J$  sous la contrainte  $F$ .

## Variational assimilation of remote sensing data for the mapping of pollutant sources in lakes

F. Van den Berghe

Centre d'Energétique - Groupe Télédétection & Modélisation, Ecole des Mines de Paris  
B.P. 207, F-06904 Sophia-Antipolis cedex, France

**ABSTRACT:** A punctual and instaneous release of a pollutant in a lake is simulated by a numerical model of advection-diffusion. The resulting state 12 hours later is used as remote sensing observations. A random noise is added to those observations in order create data of variable accuracy. Those data are then assimilated using the adjoint technics to invert the model and locate on the lake the geographical point of the initial release. The results show that, even with a degradation of 20% in the accuracy of the observations, the variational assimilation procedure is able to detect the source. However the convergence is slow and several numerical problems inherent to the advection-diffusion model are addressed in order to improve the detection accuracy.

### 1. INTRODUCTION

When some chemical products are released in a closed water body like a lake, they rapidly diffuse and the water becomes contaminated. If the emission of pollutant is sporadic, after a certain lapse of time, the concentration of the chemical products within the water is more or less homogeneous and therefore it is difficult, and sometimes impossible, to locate the source responsible of the pollutant emission. Similar questions arise with continuous emissions if they are released from different geographical sources. In such situations the problem of the local authorities can be expressed as an estimation problem: considering a polluted water body at time  $t$ , to determine the location of the source that has polluted at time  $t = 0$ . This problem was simulated using a numerical model of diffusion, and solved by a variational data assimilation procedure.

### 2. VARIATIONAL DATA ASSIMILATION

This environmental problem of estimation can be efficiently treated under a four dimensional data assimilation point of view (Lorenç 1986). This approach is used in meteorology and oceanography where the evolution of physical phenomena has to be described (Ghil and Manalotte-Rizzoli 1991). Atmosphere and oceans are dynamical systems far too large to be completely observed. So their temporal evolution can be only estimated from the available information, that is observations (sonds, satellite ...) and the knowledge on the physics of the phenomena, usually contained in a numerical model. The data assimilation is the procedure that processes

all the information to produce a time sequence of the systems states.

The last decade saw the elaboration of variational algorithms, particularly efficient when applied on data assimilation (LeDimet and Talagrand 1986). Observations are compared with model solution by the means of a cost function, which thus measures the misfit of the model to the actual state of the system, and some parameters of the model are tuned in order to minimize this misfit (Thacker and Long 1988). The values of the parameters that produce the minimum are then set up and the states of the system computed with those values are the best estimates of its actual evolution. The algorithms carried out for our purpose is a particular case of the variational data assimilation when observations are only available on the final state and the tuning parameters are the initial state.

### 3. NUMERICAL SIMULATION OF POLLUTANT DIFFUSION

The release of the pollutant into the water is simulated by the advection-diffusion Equation:

$$\frac{\partial C}{\partial t} = \nabla k \nabla C - \nabla \cdot VC \quad (1)$$

$$C(t=0) = C_0 \quad (2)$$

where:  $C(x,y,t)$  is the pollutant concentration,  $(x,y)$  is the plan of motion,  $t$  is the time,  $V(x,y)$  is the velocity vector of the water stream,  $k(x,y)$  is a constant,  $\nabla$  is the gradient in cartesian coordinates.

This equation has been discretized in finite elements, the mesh contained for the presented

simulation  $n_s = 961$  points on  $n_t = 1800$  triangles. Newman boundary conditions are adopted along the boundary and the speed vector is orthogonal to the normal vector. The temporal integration uses the Euler Forward scheme, and the program has been implemented on the Cray-YMP2E computer of Institut Méditerranéen de Technologie in Marseille.

For the purpose of the simulation the spatial domain  $\Omega$  is supposed to be a square of  $30 \times 30 \text{ km}^2$ . The emission source is located in the middle of the left side. We suppose that a water stream of  $1 \text{ km/h}$  speed is taking the diffusing pollutant from left to right, as if the pollutant source were the mouth of a river (see the arrow on Fig. 1). The diffusion coefficient is set to  $1 \text{ km}^2/\text{h}$ , so that the characteristic number of the flow  $Re = V/k$  is equal to 1.

#### 4. REMOTE SENSING OBSERVATIONS

The method used in the simulation is based on the twin experiments idea (Courtier *et al.* 1990). Between  $t = 0$  and  $t = 7.5 \text{ mn}$  (sporadic emission), the pollutant concentration is set to 1 at the geographical location of the source (in the middle of the left side of the lake see Fig. 1), while the other points are kept to 0. From this initial state (hereafter referred as *true initial state*) the model is integrated on a 12 hours period which requires 96 iterations of 7.5 mn each. The results of the integration represent the state of the lake 12 hours after the pollutant release.

From those results two cases of assimilation are reviewed:

case(a): the exact results of the integrations are used as observations, they are hence solution of the model (see Fig. 2).

case (b): a random noise is added to those results in order to simulate a remote sensor of 20 % accuracy, this noise consists in a multiplication of the exact observations of case (a) by a coefficient randomly equal to 0.8 or 1.2. It has a zero value mean so that the total mass of pollutant contained in the lake is not affected by its addition (see Fig. 4).

The values at the points of the mesh are interpolated, using the triangular barycentric coordinates, in order to produce homogeneous fields on a square image of  $512 \times 512$  pixels which stands for the remote sensing observations. Only observations on the final state used for the assimilation which hence turns to be a model inversion.

#### 5. THE COST FUNCTION

From the simulated remote sensed image (Fig. 2 and Fig. 4) on the state of the lake at  $t = 12$  hours, one wants to know where was located the source that had released the pollutant at  $t = 0$  pretending beside the observations we have no information about this source. So, from any initial state, called *the starting*

*point* of the minimization, the model is integrated and the solutions produced at  $t = 12 \text{ h}$  are compared to the observations derived from the image. The comparison is evaluated through a cost function (Talagrand and Courtier 1987):

$$J_{\text{obs}} = \frac{1}{2} \sum_{i \in I} |C_{\text{sol}}^i(t=12) - C_{\text{obs}}^i|^2 \quad (3)$$

where:  $I$  is the part of the mesh where the observations are available, here  $I$  is covering all the domain ( $I = \Omega$ ).

The data assimilation consists in minimizing this cost function by successive corrections of the initial state moving it from the starting point to the minimum of the cost function.

When the functional  $J_{\text{obs}}$  defined by (3) has a unique minimum, the minimization procedure will normally always reach the corresponding minimizing initial state (the one which by integration leads to the minimum value of the cost function), no matter the starting point is. However in some cases it can happen that the cost function defined by (3) has several minima and then in such conditions different starting points can lead to several solution initial states.

Those multiple solutions arise because the minimization problem is mathematically ill-posed in the sense that the observations are not enough to determine uniquely the initial state of the model. In order to supply the model with information, some adding constraints can be imposed to the initial state candidate to be solution. One can expect that the retrieved initial state has a limited amplitude (Bennett and Miller 1988), this will be expressed by adding to the cost function the following term:

$$J_{\text{amp}} = \frac{1}{2} \sum_{i \in \Omega} |C_{\text{sol}}^i(t=0)|^2 \quad (4)$$

Or one can impose the solution to be smooth and to give by integration smooth states at every time step  $t_k$  of the time period. This leads to add the corresponding smoothing penalty term:

$$J_{\text{smo}} = \frac{1}{2} \sum_{i \in I} \left[ \left( \frac{\partial^2 C_{\text{sol}}^i(t=0)}{\partial^2 x} \right)^2 + \left( \frac{\partial^2 C_{\text{sol}}^i(t=0)}{\partial^2 y} \right)^2 + 2 \left( \frac{\partial^2 C_{\text{sol}}^i(t=0)}{\partial y \partial x} \right)^2 \right] \quad (5)$$

A way to know if the minimum of a cost function is unique is to determine its Hessian matrix with respect to the control variables. If this matrix is definite positive (Le Dimet *et al.* 1992), then one can assume that the searched minimum is unique.

Actually in the special case of complete final observations of our experiments, it can be easily

deduced from the observability condition that the cost function defined by (3) has an unique minimum.

However in order to get a better condition number of the system during the minimization, we used the mentioned penalty terms in the cost function.

This led to a cost function which is a weighted sum of the three terms :

$$\mathbf{J} = \mathbf{J}_{\text{obs}} + r \mathbf{J}_{\text{amp}} + \lambda \mathbf{J}_{\text{smo}} \quad (6)$$

Where  $r$  is a parameter moderating the amplitude penalization in the final cost function and the coefficient  $\lambda$  sets the degrees of smoothing required to the solution. Practically those two coefficient can influence the convergence speed of the minimization procedure.

As we assume having no information on the initial sources, we have set the starting point to be the null state in which the concentration is null in every point of the domain  $\Omega$ .

## 6. ESTIMATION OF THE INITIAL STATE FROM FINAL OBSERVATIONS

In order to find the initial state which minimizes the cost function  $\mathbf{J}$ , its gradient with respect to  $\mathbf{C}(t=0)$  has to be computed. A way to compute the gradient is to introduce the adjoint state  $\mathbf{P}(t)$  since there is a result of Optimal Control theory (Lions 1969) which relays it to the gradient:

$$\text{Grad } \mathbf{J}_{/\mathbf{C}(t=0)} = \mathbf{P}(0) + r \mathbf{C}(0) + \lambda \frac{\partial \mathbf{J}_{\text{smo}}}{\partial \mathbf{C}(0)} \quad (7)$$

In order to compute the adjoint state at  $t = 0$ , the adjoint equation is introduced. We must first evaluate the tangent operator of the operator defined by equation (1) and take its adjoint. This procedure leads to the adjoint equation which is verified by the adjoint state  $\mathbf{P}(t)$ . In the present case  $\mathbf{P}(t)$  is solution of:

$$-\frac{\partial \mathbf{P}}{\partial t} = \nabla \mathbf{k} \nabla \mathbf{P} + \mathbf{V} \cdot \nabla \mathbf{P} \quad (8)$$

but the adjoint equation must be integrated backward from the final time  $t = 12$  h to  $t = 0$  with what stands for initial conditions (Talagrand 1989):

$$\mathbf{P}(t=12) = \mathbf{C}_{\text{sol}}(t=12) - \mathbf{C}_{\text{obs}} \quad (9)$$

Then a gradient descent algorithm can be carried out (Navon and Legler 1987)

## 6. ALGORITHM

- From a starting point which consists in some initial conditions  $\mathbf{C}(0)$  ( $= 0$  in our experiment, see Fig.

3), integrate the direct model from  $t = 0$  to  $t = 12$ h and store all the results  $\mathbf{C}_{\text{sol}}(t)$ .

- With the observations, compute the cost function  $\mathbf{J}$  and the term given by formula (8),
- From the initial condition:  $\mathbf{P}(t=12)$  given by (9) integrate the adjoint equation from  $t = 12$ h to  $t = 0$
- The gradient of the cost function  $\mathbf{J}$  is deduced by the equation (7) and a descent procedure (steepest gradient in our experiment) can be applied on.

## 7. NUMERICAL RESULTS

In both cases the minimization is stopped after 30 iterations of the steepest gradient which requires approximatively 8 mn of CPU time. For both cases the decrease of the cost function is displayed on Graph. 1 and the corresponding retrieved initial state are shown on Fig. 5 and Fig. 6. Those figures give a general view of the principal features contained in the solutions, but no quantitative comparison can be made since each image has its own color scale.

In order to assess the accuracy with which the minimization procedure has retrieved the solutions, for both cases we have computed the normalised residual means square error between:

- the retrieved initial state and the true initial state (RMS0)
- the state resulting of the integration at  $t = 12$  h of the retrieved initial and the assimilated observations (RMS12).

Those values are given in Tab. 1 for 0 and 30 iterations of the gradient minimization.

From the similitudes of Fig. 5 and 6 with Fig. 1, one could say that the source was detected, however the important values of the error revealed by the residual mean square error suggest that this detection is mainly visual and an objective detection criterium is desirable.

Because the initial source is not completely punctual but spread on a little domain  $\omega$  of the total bassin  $\Omega$  (exactly 139 points of the mesh are not null), the retrieved sources given by the solutions will not be punctual but be gathered around this domain. Hence the quality of a solution can be evaluated through its ability to concentrate the more pollutant mass  $M_\omega$  in this domain relatively to its total mass  $M_\Omega$  contained in the lake. Therefore a solution can be characterised by the following quantity:

$$p = \frac{M_\omega}{M_\Omega} = \frac{\sum_{i \in \omega} C_{\text{sol}}^i(t=0)}{\sum_{i \in \Omega} C_{\text{sol}}^i(t=0)} \quad (10)$$

The variation of the  $p$  values with respect to the gradient iteration number are displayed for both cases on Graph. 2.

## 8. COMMENTS

The cost function has a unique minimum and since we know that the true initial state is solution in case (a), we can conclude that this is the unique solution in this case. Despite this is not theoretically obvious, the true initial state is also the solution we could reasonably expect in case (b).

From the very similar aspects of the decrease of the both cost functions there are preemptions to think that the true initial state is also the minimum in case (b).

In order to verify the uniqueness of the minimum, we had the starting point varied and everytime we have retrieved almost the same solution, that is a state "closed" to the true initial state.

However because of the presence of the laplacian term in the adjoint model (equation (8)), the convergence rate is very poor, and it was worth improving this rate by adding the penalty terms defines by (4) and (5) The determination of suitable weighting coefficient is not straightforward. There exists some methods to determine an optimal value of  $\lambda$  from the observations (Wahba and Wandelberger 1980). Unfortunately it was not possible to apply those methods and we had to proceed by trials and corrections. No appropriate values enabling to improve the convergence rate, at least for the first 30th iterations, could be found simultaneously for  $r$  and  $\lambda$ . It will be worth examining the systematic methods mentioned above as well as carrying out some more efficient descent algorithms such as the conjugated gradient or the quasi-newton algorithm.

We could also note in Tab. 1 the important divergence between the values of RMS0 and RMS12. This points out the smoothing effect of the advection-diffusion model which literally "cut off" the propagation of the initial error. The system (model + cost function) is not very sensitive to its initial state, as it is reflected by of the adjoint magnitude (since the adjoint state at  $t=0$  represents the cost function gradient with respect to the initial conditions) which is permanently approximatively the tenth of the cost function value.

We have to add that during the experiment some negative values for the initial concentration was reached by the minimization procedure and we have to set those values to 0 at each step of the gradient descent. Despite its physical nonsense, the appearance of negative concentration is not surprising, it can be well understood if we refer to the duality between the variational assimilation and the Kalman filtering. By setting the starting point to 0 we have introduced an initial error between the concentration of this initial state and those of the true initial state which can be at its maximum equal to 1000. Since the true initial state is almost null and the estimated solution resulting of the minimization approximates the true initial state at an accuracy

which is in the same order of the initial error (Wunsch 1988), the concentrations of the solution are allowed to vary between -500 and +500. Adding a constraint for positiveness results in reducing the error associated to the solution. From this point of view we can expect the addition of positive constraint to speed up the convergence (Villalobos and Wahba 1987).

## 9. CONCLUSION

The results obtained are very encouraging and demonstrate the potential of the variational data assimilation technics to enhance the usefulness of satellite or airborne remote sensing data. Many existing numerical models describe efficiently complex phenomena occurring in the environment, it would be worth modifying such models to take into account remote sensing data in many fields of application.

## 9. REFERENCES

- Bennett A. F. and R. N. Miller, 1990. Weighting initial conditions in variational Assimilation Schemes. *Mon. Wea. Rev.*, 119, pp 1098-1102.
- Courtier Ph., J. N. Thepaut and O. Talagrand, 1990. 4 dimensional assimilation using the adjoint of primitive equations model. In *proceedings of international Symposium on assimilation of observations in meteorology and oceanography*. World Meteorological Organisation, Geneva, Switzerland, pp 337-340.
- Ghil M. and P. Malanotte-Rizzoli, 1991. Data assimilation in meteorology and oceanography. *Adv. in Geophys.*, 33, pp 141-266.
- LeDimet F. X. and O. Talagrand, 1986. Variational algorithm for analysis and assimilation of meteorological observations : theoretical aspects, *Tellus*, 38A, pp 97-110.
- Lions J. L., 1969. *Contrôle optimal des systèmes gouvernés par des équations aux dérivées partielles*. Dunod, Paris, 426 pp.
- Lorenç A., 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, 114, pp 1174-1194.
- Navon I. M. and D. Legler, 1987. Conjugate gradient methods for large scale minimization in meteorology, *Mon. Wea. Rev.*, 15, pp 1479-1502.
- Talagrand O. and Ph. Courtier, 1987. Variational assimilation of meteorological observations with

the adjoint vorticity equation I: Theory. *Q. J. R. Meteorol. Soc.*, 113, pp 1311-1328.

Talagrand O., 1989. 4 dimensional variational assimilation. in *seminar proceedings on data assimilation and the use of satellite data*. ECMWRF, Reading, England, pp.1-31.

Thacker W. C. and R. B. Long, 1988. Fitting dynamics to data. *J. Geophys. Res.*, 93, 1227-1240.

Wahba G. and J. Wandelberger, 1980. Some news methods for variational objective analysis using splines and cross validation. *Mon. Wea. Rev.*, 108, pp 1121-1143.

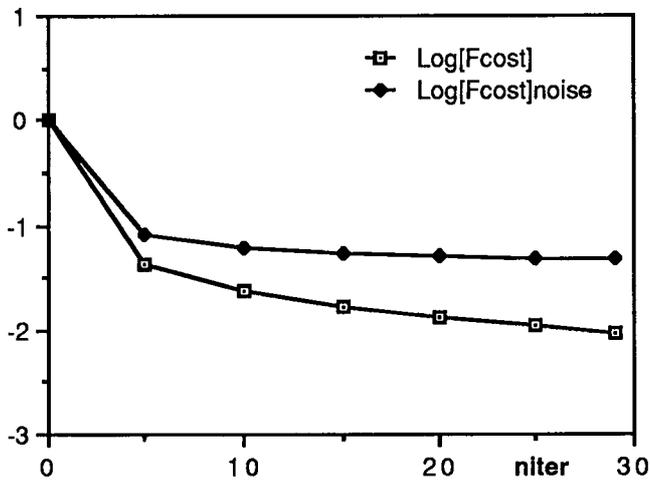
Villalobos M. and G. Wahba, 1987. Inequality constrained multivariate smoothing splines with applications to the estimation theory. *Journal of the American Statistical Association*, Vol 82, no 397, pp 239-248.

Wunsch C., 1988. Transient tracer as a problem in control theory. *J. of Geophys. Res.*, vol. 97, no. C7, pp 8099-8110.

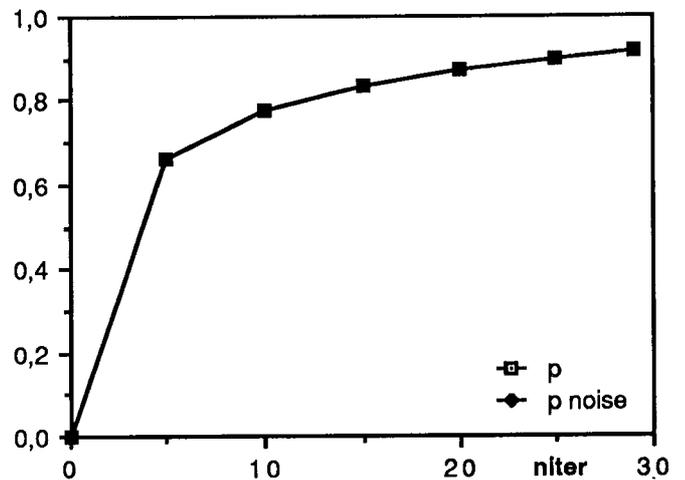
Gradient iteration number	RMS0		RMS12	
	Exact observations	Noisy observations	Exact observations	Noisy observations
0	1	1	1	1
30	0.73	0.77	0.09	0.21

Tab. 1: Normalised Residual Mean Square Error between:  
 - RMS0: the retrieved initial state and the true initial state,  
 - RMS12: the integration at t = 12h of the retrieved initial state and the observations, for 0 and 30 gradient iterations

-oOo-



Graph. 1: logarithm of the normalised cost function with respects to the gradient iteration number case (a): Fcost - case (b): Fcost noise



Graph. 2:  $p = M_{\omega} / M_{\Omega}$  with respect to the gradient iteration number boths plots are identical

DIFFUSION OF POLLUTANT IN A LAKE:  
INITIAL STATE, OBSERVATIONS AND STARTING POINT OF THE MINIMIZATION

$30 \times 30 \text{ km}^2 - V = 1 \text{ km/h} - k = 1 \text{ km}^2/\text{h} - T = 12 \text{ hours}$

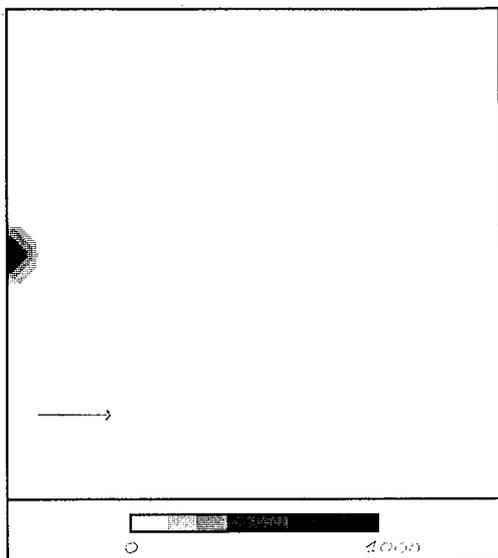


Fig. 1: initial source at  $t = 0$   
the arrow indicates the direction of  
the uniform speed field:  $V = 1 \text{ km/h}$

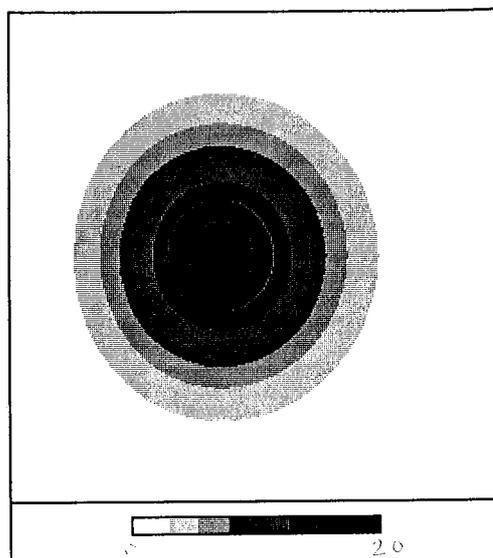


Fig. 2: integration by the model  
of the initial source, retained as  
observations at  $t = 12 \text{ h}$  in case (a)

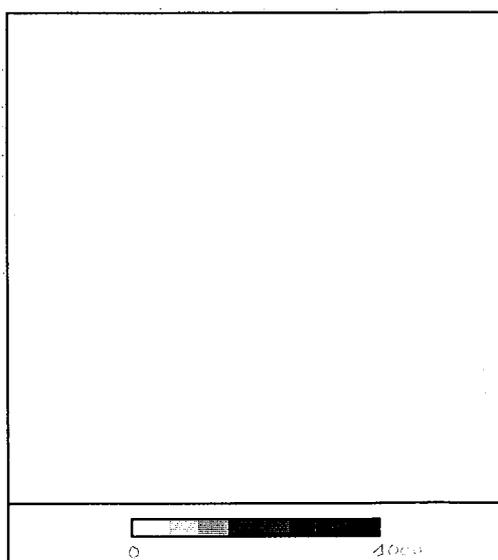


Fig. 3: starting point of  
the gradient minimization  
null state  $C = 0$  at  $t = 0 \text{ h}$

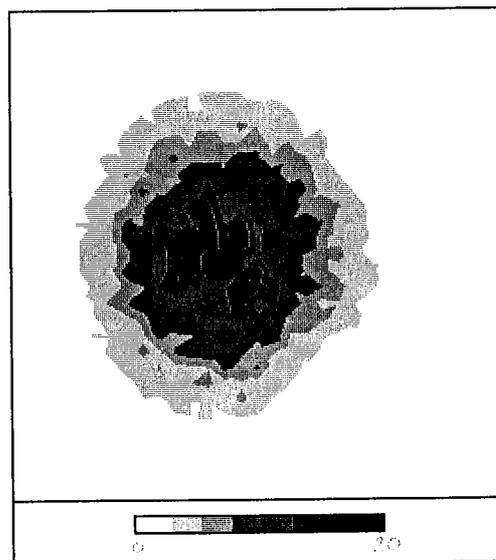


Fig. 4: as Fig. 2 with addition of  
random noise 20% of magnitude,  
retained as observations at  $t = 12 \text{ h}$  in case (b)

DIFFUSION OF POLLUTANT IN A LAKE:  
RESULTS OF THE MINIMIZATION

$30 \times 30 \text{ km}^2$  -  $V = 1 \text{ km/h}$  -  $k = 1 \text{ km}^2/\text{h}$  -  $T = 12 \text{ hours}$

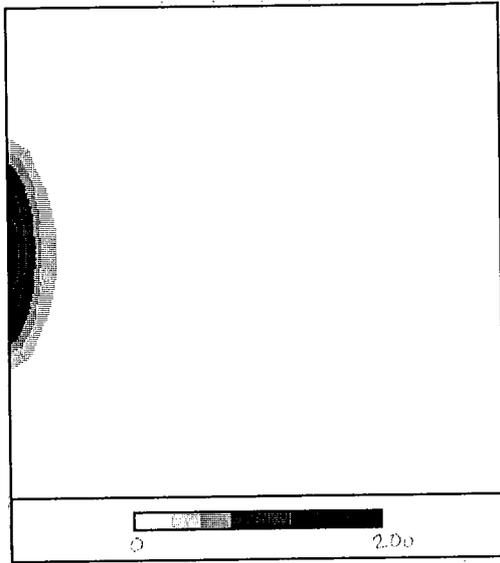


Fig. 5: retrieved initial source  
resulting of the minimization in  
case (a) at  $t = 0 \text{ h}$

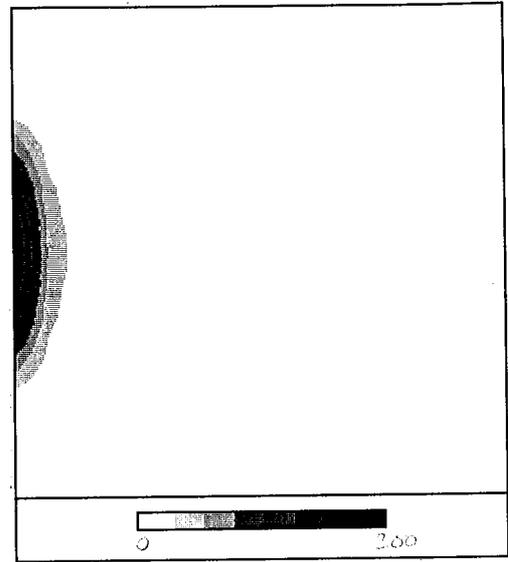


Fig. 6: retrieved initial source  
resulting of the minimization in  
case (b) at  $t = 0 \text{ h}$

## ASSIMILATION OF SATELLITE BORNE WIND SCATTEROMETER OBSERVATIONS

F. VAN DEN BERGHE &amp; F. CAUNEAU

Groupe Télédétection et Modélisation, Centre d'Energétique, Ecole des Mines de Paris, Sophia-Antipolis  
BP 207, F-06904 Sophia-Antipolis Cedex, France

## ABSTRACT

A variational formulation involving the adjoint technic is used to directly assimilate Seasat-A scatterometer satellite (SASS) surface wind measurements into a numerical model of the relative vorticity evolution. The discretized Vorticity Equation underlying the model simulates the global circulation on a rectangular basin in a  $\beta$ -plan approximation and satellite wind borne scatterometer observations are assimilated in order to produce sea surface wind fields. The variational formalism is described, and the adjoint equation is introduced. The variational assimilation procedure is then compared to Kalman filtering on real wind speed data issued from the successive passages of the Seasat satellite over the North Atlantic between September 11, 1978, 00 GMT and 12 GMT. This allows to track the evolution of the famous QE II severe storm in the global circulation of the North Atlantic ocean.

Keywords: variational assimilation, adjoint equation, wind scatterometer data, Vorticity Equation,

## 1. INTRODUCTION

The homogeneous definition of some physical fields over some spatial or temporal domains is often useful for various works or studies in Meteorology or Oceanography. They are prerequisite as initial conditions for Numerical Prediction. Unfortunately observations are made irregularly in space and time, and they are inexact. Observations are also incomplete in sense that not all the variables necessary for the numerical model are observed. So the problem of creating consistent fields on regular grid is often underdetermined. To resolve the underdeterminacy, additional information is required. The procedure that reconstructs fields from observations using prior knowledge such as statistical or balance relationships is named analysis (Lorenz Ref. 7). It becomes data assimilation when observations are distributed in time and the procedure uses an explicit dynamical model for the time evolution of the physical phenomenon inducing the studied fields.

Due to the paucity of conventional observations on sea, some regions of the globe are even void, defining consistent and accurate fields of wind or pressure over oceans is a particular challenge for data assimilation. However, despite the fact they are synoptic and distributed in time, satellite observations can supplement conventional data. Time repetitive wind speed measurements, sea surface radar images and significant waves heights, temporarily available during the 100 days of work of the Seasat satellite in 1978, will be soon available on large part of oceans and routinely delivered by the ERS-1 satellite. In the same time development of reasonably realistic models and efficient algorithms to assimilate those data tends to envisage oceanic forecasting (Thacker Ref. 9).

Our present work does not deal with forecasting but is to assess satellite wind speed measurements in order to reconstruct wind speed fields upon sea at meso-scale. Therefore the Vorticity Equation is discretized on a coarse scale ( $3^\circ$ ) and is integrated in  $\beta$ -approximation plan to simulate the global circulation over a rectangular basin which nearly corresponds to the North Atlantic ocean. The numerical model hence obtained does not realistically reflect the physical phenomena occurring at the surface of the sea, and so cannot be used for forecasting. Nevertheless some preliminary experiments have identified some good properties to reconstruct mesoscale vortices.

However, keeping in mind future applications to prediction, we have carried out one of the most promising assimilating scheme based on the adjoint equations. This variational algorithm which minimizes a distance between model solutions and observations is tested and compared to classical sequential estimation, such as Kalman filtering, on real data provided by Seasat-A scatterometer satellite (SASS) during the day of September 11, 1978.

Observations are described in section 2, the numerical model is presented in section 3, the variational formalism is derived in section 4, the results of the variational assimilation, opposed to results obtained with Kalman filtering are exposed in section 5 and commented in section 6.

## 2. THE DATA

2.1 Seasat-A Satellite Scatterometer data were available for approximately a hundred days in 1978. The Seasat scatterometer actually measured radar backscatter from surface. Backscatter measurements has to be inverted to obtain surface wind. Up to 4 wind directions (i.e. ambiguities or aliases) are possible and various algorithms, including variational method (Hoffman Ref. 3) was proposed to dealias the measurements. We used the method proposed by Wentz (Ref. 10), which provides wind speed vector with a resolution of 100 km and an accuracy of 10 % for wind speed and 20 % for wind direction. Since the rectangular grid used for the study has a spatial resolution of  $3^\circ$  ( $\approx 333$  km), the value of the closest observation is affected to the grid point.

The chosen data consists in observations made by satellite between September 11, 1978 00 GMT and 12 GMT. During this twelve hours period Seasat passed several times over the Atlantic ocean, we have retained revolutions 1087, 1088, 1090, 1093 and 1094, providing 164 observations (see Fig. 1). This particular period has been chosen since a severe storm (named QE II) was developing and seen by the satellite, while the analysis (without satellite data) of the meteo services failed to forecast it, in geographical location as well as in intensity.

**2.2 The meorological Analysis failure** is propice to asses the impact of the satellite and two analysis of the European Meteorological Bulletin during the same day at 00 GMT and 12 GMT are also used. Quasi geostrophic winds are derived from the sea surface pressure fields (Harlan & O'brien Ref. 2) given in those analysis at a resolution of 3°. The wind field obtained at 00 hour (Fig. 3) is used as first guess and then compared to the initial state retrived by the variational assimilation. The wind field obtained at 12 hours (Fig. 3) is compared to the forecast of the model after assimilation.

### 3. THE NUMERICAL MODEL

**3.1 The Vorticity Equation:** The evolution of the atmosphere is slow and an adequate approximate of the temporal evolution of the relative vorticity is given by the Vorticity Equation (Hoffman Ref. 3):

$$\xi_t(x,t) = \text{Jac}(\xi+f,\Psi) \quad 1$$

where where  $t$  is the time and  $\dot{t}$  the time derivative,  $x$  is the state vector in the plan of motion  $(x,y)$ ,  $\Psi$  the stream function and  $\xi$  the relative vorticity,  $f$  is the coriolis parameter and  $\text{Jac}$  is the jacobian operator with respect to the rectangular  $x$  and  $y$ . Some useful relationships can be derived for inviscid flow:

$$\nabla(x,t) = k \times \nabla\Psi, \xi(x,t) = k \cdot \nabla \times V(2), \xi(x,t) = \Delta\Psi \quad 2$$

where  $k$  is the unit vector normal to the plane of motion,  $V$  is the speed vector,  $\nabla$  is the two dimensional gradient operator and  $\Delta$  is the laplacian operator.

**3.2 Discretization:** The equation is discretized according to a finite difference scheme and integrated on a rectangular grid of 12 x 33 points, The field resolution is 3° (333 km) which would correspond approximately to the North Atlantic basin (from -73° to 12° East and from 33 to 66° North). The scheme adopted (Arakawa Ref. 11) to compute the Jacobian conserves the total kinetic energy and the total enstrophy. The leap-frog time integration method used integration interval of 30 mn at each iteration. The period of assimilation is 12 hours.

**3.3 boundary conditions** arise serious problems when inverting equations (2). The best solution is to put boundary conditions together with initial conditions in the control variables of the variational problem. This is succesful, if assimilated data contained information on the boundary, but in our case few observations was at the boudary of the grid and instead of complicating the numerical implementation, the boundary conditions from the initial state were kept during the integration. Since it is quiet far from the boundary, this does not much influence the evolution of the principal vortex in the studied fields, but, clearly some spurious effects appear near the boundary after 6 hours of integration.

## 4. VARIATIONAL ASSIMILATION WITH THE VORTICITY EQUATION

**4.1 Variational versus sequential assimilation:** The ingredients for data assimilation are heterogenous observations geographically spread up and distributed on a time period  $[0, T]$ , and a numerical model which is able to compute homogeneous fiels regulary spaced on a grid at each instant  $t$  of the time period provided that the initial state at  $t = 0$  is supplied. The question is how to merge observations and model solutions in order to create homogeneous fields on regular grid which are realistic. The traditional solution developed by Kalman filtering is sequential (Ghil & al Ref.

4): as news observations are available, they are blended with the time corresponding model state (forecast) so as to minimize the error covariance matrices. The resulting fields is called analysis. The forecast comes from the integration by the model of the analysed field obtained at the previous time step. If the process is just beginning then it must be guessed. As time passed, more and more data are assimilated and the start-up bias may gradually vanish due to the melding of forecasts and new data. The Kalman filtering is a recursive method, that is to say only past observations are taken into account in the present analysis fields.

**4.2 Variational method:** On the contrary, with variational assimilation, past, present and future data are taken into account in the present analysed fields. The basic idea is that a parameter of the model can be varied until a cost function which measures the misfit of the model to all the observations available in the assimilating time period is minimized. The parameter is then set to the value corresponding to this minimum and the fields computed with this value of the parameter are the "closest" state, among all the possible model solutions, to the observations.

Normaly the minimization occurs the whole state vector computed during the time period and is usually untracktable. By introducing adjoint equations of the model, optimal control theory gives an algorithm to efficiently perform the minimization (LeDimet & Talagrand Ref. 8). In our study control variables (the tuning parameter) are the initial state of the model.

**4.3 The cost function:** The distance between the model state and observations is evaluated through a quadratic function. This cost function is the sum extended to all the grid points of the square error between the vorticity obtained by integration of (1) and the wind speed measurements sent by the satellite at different instants  $t_i$ ,  $0 < t_i < T$ :

$$F_{\text{cost}} = 1/2 \sum_{D, t_i} [ C\xi(t_i) - V_{\text{obs}}(t_i) ]^t A [ C\xi(t_i) - V_{\text{obs}}(t_i) ] \cdot dt \quad 3$$

Where the subscript  $t$  denotes the transposed matrix,  $dt$  is the time period at each iteration of the integration,  $A$  is a positive matrix standing for the scalar product,  $D$  is the part of the model grid where observations are available and  $C$  is a mapper from the state space to the observation space.

Since the model state is a regulary gridded field of relative vorticity, whereas observations are made of wind speeds gathered in the satellite sweath,  $C$  first converts vorticity into speed and then matches the obtained gridded vector with the domain of observation  $D$ .

The  $A$  matrix moderates the influence of each observation in the cost function. The choice of their weight depends on the relative confidence we have in observation, ideally it should be the inverse of the error covariance matrix. Practically error measurements has been kept constant in our assimilation, and, in the presented experiments, we have set  $A$  to be diagonal and equal to 1.

**4.4 The minimization procedure:** The data assimilation problem is to find the initial state  $\xi(0)$  which by integration of equation (1) will give the smallest value of  $F_{\text{cost}}$ . This is a mathematical problem of unconstrained minimization. In order to perform such a minimization one must access the gradient of  $F_{\text{cost}}$  with respect to  $\xi(0)$ , this is possible thanks to a basic result of the optimal control theory (Lions Ref. 6), which exhibits a relationship between the searched gradient and the adjoint state  $P(0)$ , more precisely:

$$\text{Grad}_{\xi(0)}(F_{\text{cost}}) = P(0) \quad 4$$

**4.5 The adjoint model** gives an efficient means to compute the value of  $P(0)$ , which in our case is solution of (Talagrand & Courtier 1987) :

$$P_t(x,t) = \Delta Jac (\Delta^{-1}P, \Delta^{-1}\xi) + Jac (\xi, \Delta^{-1}P) - A C^\dagger (C\xi(t) - V_{obs}(t)) \quad 5$$

$$P(t=T) = 0 \quad 6$$

The adjoint equation should be integrated backward in time. The discretisation and integration scheme is the same for the direct and the adjoint model. Therefore an algorithm can be carried out.

**4.6 Assimilation algorithm:** The gradient of  $F_{cost}$  corresponding to some initial state  $\xi(0)$  can be explicitly obtained by performing the following operations :

- Starting from some  $\xi(0)$ , integrate the basic equations (1) from 0 to T and evaluate the cost function, store the values computed for each value of t which will be necessary to make up the adjoint equation.

- Starting from  $P(T) = 0$ , integrate the adjoint equation (5) backwards in time from T to 0.

The final result at time  $t = 0$  will give the gradient of  $F_{cost}$  with respect to  $\xi(0)$  which will allow to apply a descent minimizing procedure (steepest gradient, conjugate gradient).

**4.7 Remark on backward integration:** Information is inserted in the model every time an observation is available. This information is propagated by the direct model from the insertion instant to the final time of the integration. Hence in classical sequential estimation, only the information contained in the assimilated observations before time t resides in the model solution at t. When integrating backward, information is then propagated from the final instant to the initial instant. Since all the information is contained at the final instant, all the information is then redriven to the initial time and the initial conditions retrieved by the minimization procedure also contains the information brought by the future observations.

## 5. NUMERICAL RESULTS

Variational assimilation of the SASS data described in section 2 is implemented, the cost function is minimized using the steepest gradient procedure which is far from being the more computationally efficient but immediately implementable when no software library is available. As in conventional assimilation procedure, meteorological analysis at  $t = 0$  is used as first guess, that is to say is, the meteorological analysis (Fig. 2) is used to start up the minimization procedure. The cost function minimum is obtained after 30 iterations of the descent procedure (graph 1) which takes two hours and half on a limited power computer (5 MIPS).

Five wind speed fields are produced by the procedure resulting of the assimilation of observations issued from the satellite passages at 0h 32, 2h 09, 3h 47, 10h 05 and 11h 55. To assess the impact of data, at each assimilating date the Root Mean Square (RMS obs) error between observations and the corresponding solution is evaluated and divided by the mean value of the observations. Those solution fields are obtained by integration of the initial state which has minimised the cost function, so it is also worth examining it and to compare it to the meteorological analysis which was used as first guess. Hence the Root Mean Square (RMS meteo) error between the retrieved initial state and the meteorological analysis is also evaluated. Mathematically those RMS are computed as follows:

$$RMS\ obs = \sqrt{\frac{\sum_D [V_{sol}(t_i) - V_{obs}(t_i)]^2}{\sum_D [V_{obs}(t_i)]^2}} \quad 7$$

$$RMS\ meteo = \sqrt{\frac{\sum_D [V_{sol}(0) - V_{meteo}(0)]^2}{\sum_D [V_{meteo}(0)]^2}} \quad 8$$

Their values are given in table 1. The retrieved initial state is shown on Fig. 4.

A Similar assimilation has been previously performed using the Kalman Filtering (see Ghil & al in Ref. 4 for practical implementation) and the analysis fields obtained at  $t = 2h\ 09$  by each procedure are compared, they are presented on Fig. 5 for variational assimilation and Fig. 6 for Kalman filtering. As the observations available at  $t = 3h\ 47$  seem to be erroneous, namely they are locating the vortex too north and too low in intensity, we have exhibited them in Fig.7 and the corresponding part of the fields obtained after their assimilation in Fig. 8.

## 6. COMMENTS

Influence of observations can be recognised in the retrieved initial state on Fig 4. The backward integration has effectively reversely propagated the information brought by successive observations to the initial time. It is more obvious, when comparing analysed fields obtained at  $t = 2h\ 09$  produced by variational assimilation and sequential estimation (resp. Fig. 5 and Fig. 6). With the sequential analysed fields, the influence of the two first passages of the satellite can be noticed whereas in the variationally analysed fields the later observations (3, 10 and 11h) have already influenced the evolution of the vortex. This is important, since the storm is better described in the latest passages. To assess the relative importance of the model and the observations in the assimilating scheme, we can examine the part of the analysed field obtained after the assimilation at  $t = 3h\ 47$  which geographically corresponds to the observations made by the satellite at that time (Fig. 7 and Fig. 8). The satellite observations seem to be erroneous, as they represent the vortex too much north and as they are too low in intensity. After their assimilation, wind speed vectors which are geographically at the location of the observations are much more intense, which is more realistic and better located relative to the storm. The model has effectively corrected the observations while assimilating them. The value of the Root Mean Square error (Tab. 1) between the analysed field and observations indicate an error of 200 % instead of 80 % with other observations. This confirms the strong contradiction between the observations and the model. We can also notice the increase of the Root Mean Square error between the retrieved initial state and the meteorological analysed fields from 87% at  $t = 0$  and 106 % at  $t = 12h$ . This is certainly due to the simplicity of our model which does not filter noise and unrealistic features. Furthermore some important spurious boundary effects can be seen in the results of the model after 6 hours of integration.

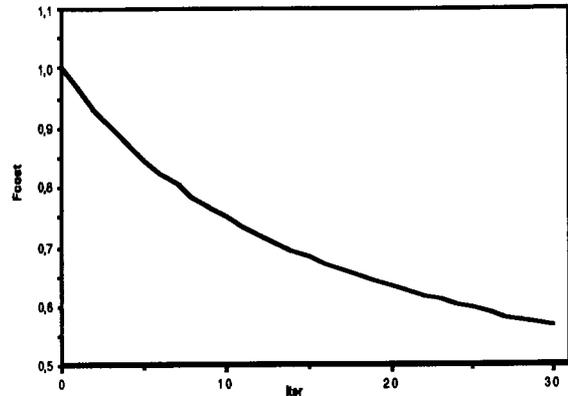
## 7. CONCLUSION

The benefit on the quality of analysed sea wind vector fields performed with a variational procedure is shown on an assimilation of SASS data on a twelve hours periods and is compared to the Kalman filtering which is easier to implement and requires less computing time. However the variational procedure is now being implemented on the Cray Y-MP2E of Institut Méditerranéen de Technologie in Marseille (France) and very important improvements have been obtained on computing time during the experiments.

## 8. REFERENCE

1. Arakawa A 1966, Computational Design for long term Numerical Integration of the Equations of Fluid Motion : Two dimensional Incompressible flow. Part 1. Journal of Computational Physics, 1, pp 119-143.
2. Harlan J Jr & O'Brien J 1986, Assimilation of Scatterometer Winds into Surface Pressure Fields using a Variational Method. Journal of Geophysical Research, Vol 11 pp 7816-7836.
3. Hoffman R 1984, SASS Wind Ambiguity Removal by Direct Minimization. Part II: Use of smoothness and dynamical constraints. Monthly Weather Review, 112, pp 1829-1852.
4. Ghil M & al, 1981 Application of estimation theory to numerical weather prediction, Dynamic Meteorology: Data Assimilation Methods, Springer-Verlag, New-York, pp 139-224.
5. LeDimet F X and Talagrand O 1986, Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspect. Tellus, 38 A, pp 97-110.
6. Lions, J L 1971, Optimal control system governed by partial differential equations. Springer-Verlag, Berlin, 396 pp.
7. Lorenc A C, 1981. Analysis methods for numerical weather prediction. Quart. J. R. Met Soc. 112, p 1177-1194.
8. Talagrand O. Courtier P. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. 1 : Theory. Quart. J. R. Met Soc. 113, pp 1311-1328.
9. Thacker W. C. and R. B. Long, 1986. Fitting Dynamics to Data. Journal Of Geophysical Research, Vol 93, NO. C2, pp 1227-1240.
10. Wentz F J & Mattox A, 1986, New algorithms for microwave measurements of ocean winds: Application to

-oOo-



Graph 1 : Cost function versus iteration number of the steepest gradient descent

Date	RMS obs
t = 00h 00	0.875
t = 00h 32	0.705
t = 02h 29	0.862
t = 03h 47	2.088
t = 10h 05	0.561
t = 11h 55	0.675
t = 12h 00	1.064

Tab 1: Relative Root Mean Square error between analysis and observations at each assimilating time step. For t = 0 and t = 12h RMS is computed with respect to the meteorological analysis.

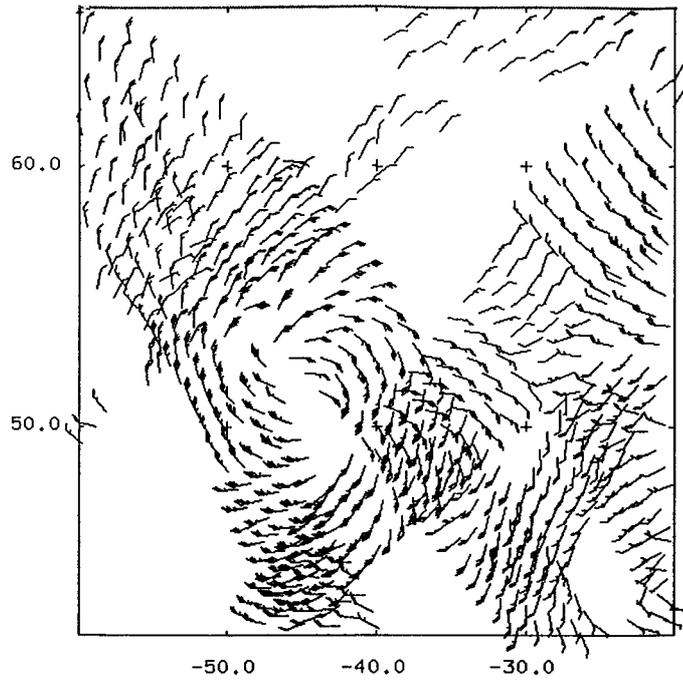


Fig. 1: Wind speed vector observations from Seasat revolutions 1087, 1088, 1090 and 1093, September 11, 1978. A bar indicates 5 m/s of magnitude.

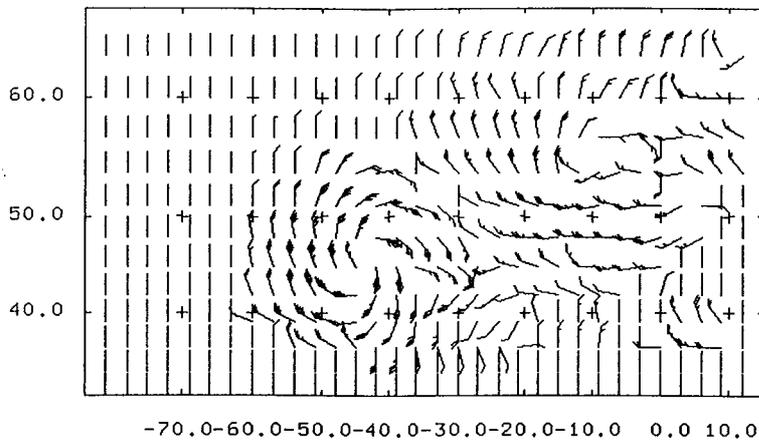


Fig. 2: geopotential winds derived from the sea surface pressure analysis on September 11, 1978 00 GMT (European Meteorological Bulletin)

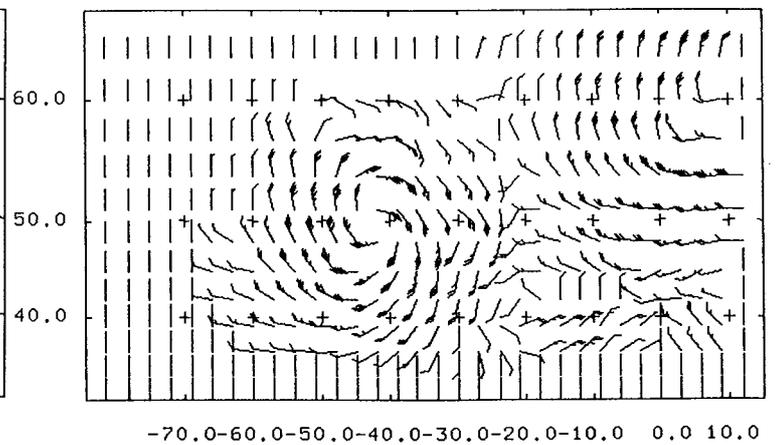


Fig. 3: geopotential winds derived from the sea surface pressure analysis on September 11, 1978 12 GMT (European Meteorological Bulletin)

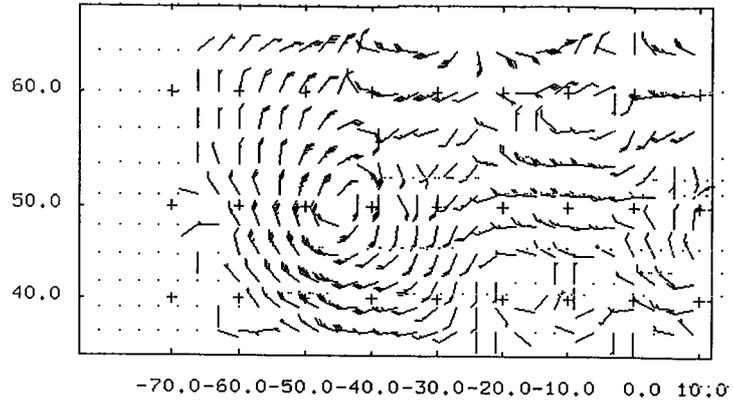


Fig. 4: sea surface wind speed retrieved by the variational assimilation at t = 00h

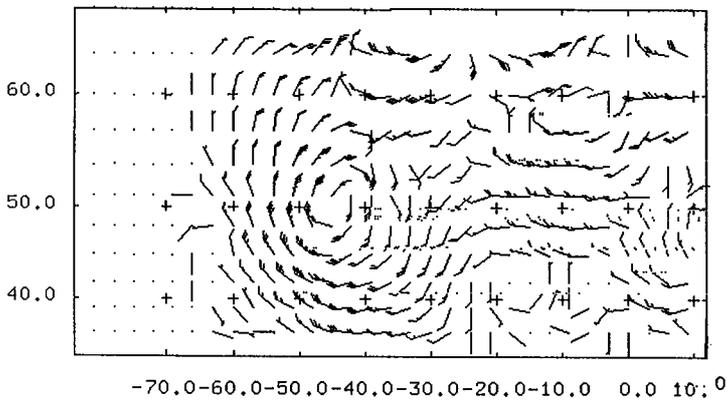


Fig. 5: sea surface wind speed retrieved by the **variational** assimilation at t = 2h 09

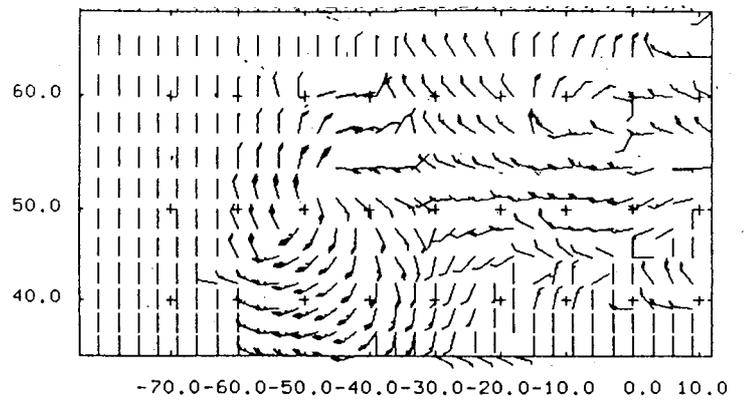


Fig. 6: sea surface wind speed retrieved by the **sequential** assimilation at t = 2h 09

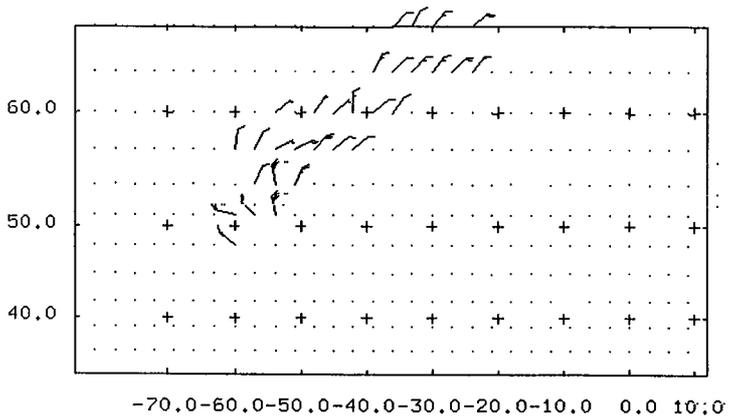


Fig. 7: Satellite observations at t = 3h 47

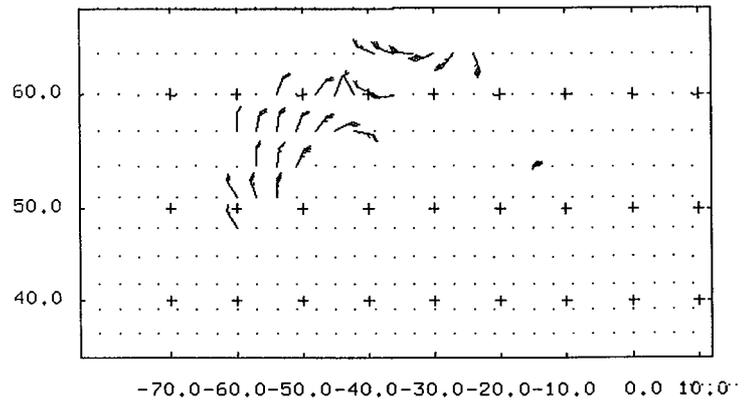


Fig. 8: analysed wind speed vector after assimilation at t = 3h 47. Vectors corresponding geographically with observations are represented

## RESUME

Nous nous intéressons aux méthodes permettant d'assimiler les données de télédétection. Il y a principalement deux types de technique : les méthodes séquentielles, issues du filtrage de Kalman et les méthodes variationnelles, basées sur les équations adjointes du contrôle optimal. Les méthodes variationnelles sont récentes et nous essayons de présenter leur intérêt pour la télédétection à l'aide de deux applications.

Une première étude est menée sur un modèle numérique d'advection-diffusion. Elle montre la faisabilité numérique de la méthode : il est possible de minimiser une fonction de coût mesurant la distance entre la trajectoire du modèle et des observations réparties à la fois dans le temps et dans l'espace. La deuxième application est réalisée sur des observations effectuées par le diffusiomètre-vent embarqué à bord du satellite Seasat. Son objet est la cartographie des champs de vent à la surface des océans. Sur une période de douze heures, nous assimilons les observations dans un modèle numérique non linéaire, discrétisant l'équation de la vorticité sur un domaine rectangulaire couvrant à peu près tout l'océan Atlantique nord.

Une comparaison entre les champs de vent, obtenus par l'assimilation variationnelle, et ceux produits par une assimilation séquentielle, met en évidence la propriété de rétro-propagation de l'information dans le temps que possèdent les équations adjointes. Il en résulte des analyses sensiblement meilleures, surtout en début de période d'assimilation, et une grande cohérence dynamique entre les champs analysés et le modèle numérique.

**Mots clés :** télédétection, assimilation de données, équations adjointes, filtre de Kalman, modèle numérique, mesures diffusiométriques.

## ABSTRACT

Satellite data assimilation methods are investigated. There are mainly two kinds of technics : the sequential methods, derived from the Kalman filter and the variational methods, based on the adjoint equations of the optimal control theory. Variational methods are more recent. This work attempts to asses their potentialities in remote sensing through two examples.

The first study, carried out with the advection-diffusion equation as a numerical model, demonstrates the feasibility of the variational methods. It is possible to minimize a cost function measuring the distance between the model trajectory and observations distributed in time and space. The second application is performed with actual observations of the Seasat satellite borne-wind scatterometer. Its aim is the mapping of the wind fields at the sea surface. Within a twelve hours time period, data are assimilated in a non linear numerical model. This model discretizes the vorticity equation on a rectangular domain covering approximatively the North Atlantic Ocean.

A comparison between the wind fields obtained with the variational assimilation and those produced by a sequential assimilation highlights the adjoint equations ability to propagate information backward in time. The resulting analyzes are better, particulary at the beginning of the assimilation time period, and the analyzed fieds are consistent with the model dynamics.

**Key words :** remote sensing, data assimilation, adjoint equations, Kalman filter, numerical model, scatterometer observations.