N. Daget[1,*]

Revue des méthodes d'assimilation

14 novembre 2007

[1]CERFACS, 42 avenue G. Coriolis, 31057 Toulouse cedex 01, Toulouse, France

[*] Corresponding author.
CERFACS, 42 avenue G. Coriolis, 31057 Toulouse cedex 01, Toulouse, France, daget@cerfacs.fr

Technical Report
TR/CMGC/07/113
GLOBC - CERFACS

Table des matières

Chapitre 1

Préambule

> Plus le message est probable moins il fournit d'information : les clichés et les lieux communs éclairent moins que les grands poèmes.
>
> *Norbert Wiener*

Historique

La recherche scientifique repose sur deux sources d'information et d'étude différentes, les observations d'une part, et les modèles d'autre part. Au cours de l'histoire des Sciences, les observations furent d'abord qualitatives avant de devenir quantitatives. Parallèlement, les modèles descriptifs s'améliorèrent grâce aux techniques mathématiques, et particulièrement à l'analyse numérique.

La charnière du XVIII^{e} et XIX^{e} siècle vit ainsi apparaître l'assimilation de données. Cette technique vise à estimer l'état d'un système dynamique en utilisant toutes les sources d'information. Elle appartient au domaine de l'estimation statistique et de la théorie du contrôle (Gelb, 1974 et Lions, 1968). Johann Tobias Mayer (1723-1762), astronome allemand, calcula les mouvements de la lune avec une admirable précision, et mérita, par ses *Tables de la Lune*, le grand prix décerné par le Bureau des longitudes de Londres (1755). En effet, il évalua les erreurs dues aux imperfections des réglages des instruments de mesure et eut le premier l'idée de répéter la mesure des angles pour atténuer les erreurs de mesure. À cette époque, de nombreux mathématiciens et astronomes développèrent des méthodes proches de la *Méthode aux moindres carrés*. Adrien Marie Legendre (1752-1834), professeur de mathématiques à l'école militaire de Paris, exposa, dans un traité sur les orbites des comètes en 1805, cette méthode d'ajustement dite *des moindres carrés*. Elle fut cependant attribuée à Karl Friedrich Gauss (1777-1855) qui publia en 1809 son travail sur les mouvements des corps célestes contenant la *méthode aux moindres carrés* et permettant de calculer l'orbite de Cérès. Il affirma, pour sa défense, qu'il utilisait cette méthode depuis 1795. Cependant, l'astronome, physicien et mathématicien Pierre Simon Laplace revendiqua aussi la paternité de cette méthode dans ses travaux sur la stabilité mécanique du système solaire dans lesquels il développa en 1783 une méthode proche de celle aux moindres carrés visant à résoudre un système d'équations sous-déterminé.

Les astronomes furent donc les premiers à travailler sur la théorie de l'estimation avant que les mathématiciens ne reprennent la thématique. C'est au XX^{e} siècle que Sir Ronald Aylmer Fisher (1890-1962) apporta une contribution majeure au domaine de l'estimation. Il publia en 1912 (Fisher, 1912) un article sur les fonctions de densité de probabilité en utilisant, sans le citer, l'*estimateur du maximum de vraisemblance*. Il publia ensuite en 1925 l'ensemble de ses travaux sur l'estimation (Fisher, 1925) qui devint un best-seller de la statistique (14 éditions et traduit en 6 langues). Il faut ensuite attendre les années 40 pour que Norbert Wiener (1894-1964), philosophe et mathématicien, en se fondant sur la théorie des processus aléatoires, présente une approche du filtrage optimal adaptée aux problèmes spectraux (Wiener, 1949). Cette technique nommée *filtre de Wiener* traite les problèmes continus dans le temps à l'aide de fonctions de corrélation et se limite aux processus stationnaires. À la même époque, Andreï Nikolaïevich Kolmogorov (1903-1987) traite le problème discret dans le temps : ce que les historiens des sciences appellent une double découverte. Durant les années suivantes, le travail de Wiener fut étendu aux cas non-stationnaires. Dans les années 60, Rudolf Emil Kalman (1930-) unifia d'abord le problème continu et discret dans le temps avant de développer le filtrage optimal récursif (Kalman, 1960 et Kalman et Bucy, 1961), plus connu sous le nom de *filtre de Kalman*. À la différence du *filtre de Kalman*, le *filtre de Wiener* n'a pas cette bonne propriété de récursivité. Le *filtre de Kalman* est maintenant utilisé dans beaucoup de domaines tels que la détermination d'orbites satellitales, les systèmes de guidage, le traitement d'images, les sciences de la terre ou l'économétrie. Vers la même époque, Yoshi Kazu Sasaki, qui travaillait sur la prévision des trajectoires des ouragans, proposa une approche variationnelle de l'estimation (Sasaki, 1958 et Sasaki, 1970) qui, sous certaines hypothèses, conduit à des résultats semblables à ceux du filtrage optimal. La formulation tri-dimensionnelle est connue sous le nom de 3D-Var tandis que celle quadri-dimensionnelle est ordinairement nommée 4D-Var.

L'assimilation de données en météorologie...

Des sciences de la terre, la météorologie s'empara la première de l'assimilation de données. Gandin développa une analyse objective des champs de précipitations totales dès 1963 (Gandin, 1963). Au cours des dernières décennies, les progrès en météorologie ont été rendues possible par la conjonction systématique des observations et des apports de la théorie. La motivation et l'enjeu essentiel de ces différentes sources d'informations étaient et sont encore le besoin de prévisions météorologiques numériquement calculables. Très rapidement s'est imposée l'idée que les modèles dynamiques utilisés devaient rendre compte de la propagation des informations dans le temps et l'espace. Cette notion de propagation tente de compenser la disparité et l'imprécision des observations en permettant de construire une image consistante et quadri-dimensionnelle de l'atmosphère. Convaincus de l'importance cruciale de l'assimilation de données, de nombreux centres de recherche ou de prévision opérationnels tels que Météo-France ou le CEPMMT[1] Centre Européen pour les Prévisions Météorologiques à Moyen-Terme (ECMWF)(Rabier *et al.*, 2000) participent à des projets internationaux ayant pour objectif l'amélioration des prévisions opérationnelles, mais aussi l'étude de l'impact anthropique sur l'évolution du climat et l'amélioration de la connaissance de nombreux phénomènes climatiques régissant ce système. Ceci devient possible par une amélioration, qualitative et quantitative, des sources d'informations, grâce notamment aux nouvelles générations de satellites d'observation (ER2, Jason, Envisat), et par une incessante augmentation des capacités de calcul.

...puis en océanographie

L'essor de l'océanographie physique est plus récent. Cependant l'intérêt de plus en plus marqué pour l'étude de la dynamique de l'océan, ainsi que l'amélioration des modèles numériques (Gent et McWilliams, 1990 ; Holloway, 1992 ; Large *et al.*, 1994 et Griffies *et al.*, 2000) ont motivé l'émergence de l'assimilation de données dans le domaine de la prévision océanique (Koblinsky et Smith, 2001). Le déploiement de réseaux d'observation des océans avec des bouées ou des flotteurs (données *in situ*) et l'utilisation des données satellitaires (altimétriques, radiométriques,...) ont soutenu l'émergence et le développement de l'assimilation de données appliquée à l'océanographie. Ses objectifs sont de satisfaire, d'une part, aux besoins de l'étude de la dynamique océanique en effectuant la meilleure estimation possible de l'état de l'océan au cours du temps. Cette estimation permet d'obtenir un état initial le plus réaliste possible pour la prévision océanique et pour la prévision saisonnière qui utilise des systèmes couplés océan-atmosphère. D'autre part, l'assimilation de données permet d'estimer, voire d'améliorer, les paramètres du modèle numérique d'océan (Smedstad et O'Brien, 1991), les forçages (Stammer *et al.*, 2002 et Vossepoel et Behringer, 2000) ou les conditions aux frontières (Bennet et McIntosh, 1982 et Deltel, 2002). La capacité du système à s'ajuster vers les observations permet d'identifier certains biais ou dérives du modèle numérique d'océan (Bennett *et al.*, 1998). De la même manière, des écarts trop importants identifiés par

[1]Centre Européen pour les Prévisions Météorologiques à Moyen-Terme

l'assimilation entre les observations et le modèle peuvent mettre en évidence une incohérence dans les observations (Holland, 1989). Enfin, l'assimilation peut aussi être utilisée pour évaluer un système d'observations (Miller, 1990 et Carton *et al.*, 1996).

Les principaux objectifs

L'assimilation de données passées disponibles sur de longues périodes se nomme ré-analyse. Le projet européen ENACT[2]ENhanced ocean data Assimilation and Climate predicTion avait ainsi pour objectif de réaliser, entre autre, des ré-analyses sur la période 1962-2001. La compréhension de la circulation océanique, de la variabilité du système et des mécanismes qui engendrent les phénomènes physiques majeurs de l'océan (Stammer, 2002) peut être améliorée avec les produits issus de ces ré-analyses.

Comme en météorologie, la connaissance d'un état initial le plus juste possible permet de produire des prévisions les plus fiables. L'assimilation de données a ainsi pour objectif de construire cette condition initiale afin d'améliorer les prévisions météorologiques ou océaniques. C'est pourquoi elle est au cœur de projets d'océanographie opérationnelle comme MERCATORProjet d'océanographie opérationnelle - http ://www.mercator-ocean.fr. Les prévisions obtenues permettent de connaître au mieux la dynamique globale des courants et masses d'eau, ainsi que l'équilibre biologique régnant dans l'océan qui en découle. Tous ces produits sont directement utiles dans les domaines de la pêche, du transport maritime, de la protection des espèces marines, de la défense et, plus anecdotiquement, dans celui des courses au large. La qualité des systèmes permet aujourd'hui une analyse de la circulation océanique réelle (Leetmaa et Ji, 1989) ainsi que la représentation de la physique méso-échelle globale (Fu et Smith, 1996). Un système d'océanographie opérationnel performant doit ainsi être composé d'un modèle océanique réaliste, d'une méthode d'assimilation efficace et d'observations nombreuses et de qualité.

Pour la prévision saisonnière, la connaissance la plus réaliste de l'état de l'océan est primordiale. En effet, la qualité du couplage océan-atmosphère est conditionnée fortement par la composante océanique qui représente la mémoire du système : la capacité de stockage de l'océan est 1200 fois supérieure à celle de l'atmosphère. L'assimilation de données permet de déterminer des conditions initiales suffisamment réalistes à l'interface océan-atmosphère. En supposant que le couplage soit réalisé parfaitement entre un modèle d'océan et un modèle d'atmosphère tout deux parfaits, l'information contenue dans l'état initial de l'océan est préservée et propagée par le modèle d'océan, puis transmise correctement au modèle d'atmosphère. Une des hypothèses fortes des modèles parfaits est d'être non biaisée. Cependant, et malgré ces hypothèses fortes, la capacité des méthodes d'assimilation à améliorer la qualité des conditions initiales océaniques et des prévisions saisonnières n'est plus à démontrer. La mise en évidence d'un impact positif sur l'ensemble du globe est cependant plus compliquée. Ces difficultés sont dues notamment aux techniques de couplage, par exemple l'échange des flux entre les modèles d'océan et d'atmosphère, et à la qualité des modèles toujours perfectibles. Cependant, pour certaines régions du globe où la dynamique est spécifique, ou dans un contexte d'étude particulier, l'assimilation de données a pu clairement montrer son intérêt. L'exemple de la région Pacifique tropical avec l'assimilation des données TAO[3] Tropical At-

[2]ENhanced ocean data Assimilation and Climate predicTion
[3]Tropical Atmosphère Océan

mosphère Océan améliore clairement la prévision des événements ENSO[4] El Niño Southern Oscillation (Ji *et al.*, 1997 ; Segschneider *et al.*, 2000 et Alves *et al.*, 2004). Ces prévisions saisonnières sont produites de manière opérationnelle par le NCEP[5] National Centers for Environmental Prediction (Behringer *et al.*, 1998 et Ji *et al.*, 1998) et par le CEPMMT (Segschneider *et al.*, 2000 et Alves *et al.*, 2004).

Les contraintes

Le développement des méthodes d'observation et de la modélisation de l'océan ont fait évoluer, au cours des dernières décennies, les techniques d'assimilation de données du système océanique. D'autre part, le développement de méthodes d'assimilation de plus en plus performantes dans le domaine de l'atmosphère a profité au domaine océanique. Cependant, quelle que soit la méthode d'assimilation de données choisie, les problèmes sont globalement les mêmes, essentiellement celui de la taille du système, qui comprend généralement plus de 10^7 degrés de liberté, et qui pose certains soucis de temps de calcul et d'espace mémoire. Aucun progrès notable n'aurait pu voir le jour sans la mise à disposition de moyens de calculs importants. Aujourd'hui, grâce à des super-calculateurs, les centres de prévisions atmosphériques opérationnels utilisent des méthodes d'assimilation performantes. Néanmoins, la puissance de calcul reste le facteur limitant pour le développement de méthodes d'assimilation dans le domaine des sciences de la terre. Les objectifs actuels de la communauté scientifique sont donc les suivants : réduire la taille des systèmes étudiés sans perdre d'information, améliorer la connaissance des paramètres physiques, la statistique des erreurs de mesure, prendre en compte l'erreur modèle sans trop pénaliser la résolution numérique. . .

[4]El Niño Southern Oscillation
[5]National Centers for Environmental Prediction

Chapitre 2

Introduction

La connaissance progresse en intégrant en elle l'incertitude, non en l'exorcisant.

Edgar Morin in La Méthode (1977), La Vie de la vie

2.1 Un exemple simple : la météorologie

La première question que l'on peut se poser est : « A quoi sert l'assimilation de données ? ». Une question bien peu scientifique car ne répondant pas à la question « Comment ? ». Et pourtant, c'est souvent la première question posée lorsque l'assimilation de données fait irruption dans une discussion. Certes, c'est un sujet peu abordé d'ordinaire, mais que certains chercheurs ou doctorants travaillant dans le domaine rencontrent fréquemment.

Alors, à quoi peut bien servir l'assimilation de données ?

Un exemple à la fois simple et historique est celui de la météorologie. Prenons un exemple concret tel que la prévision de la température à Toulouse. La première méthode consiste à faire régulièrement des mesures et à utiliser cette série de mesures pour prévoir la température. Plusieurs méthodes sont disponibles comme celle de dire que la température du lendemain sera identique à celle du jour précédent. Ou alors celle supposant que la température du lendemain sera fonction de la tendance des derniers jours. En d'autres termes, la démarche consiste à extrapoler les mesures en se fondant sur la régularité de la courbe statistique ou bien sur une base statistique événementielle. Cette méthode était celle utilisée avant l'arrivée des ordinateurs.

La seconde méthode consiste à étudier le système globalement pour comprendre quelles sont les lois qui gouvernent la dynamique météorologique. Armé de ces lois, il est ensuite possible de développer un modèle informatique de prévision météorologique. En lui prescrivant un état initial construit grâce aux mesures effectuées, ce modèle peut alors prévoir la température sur Toulouse. Cependant, la complexité du modèle, la taille du système et les observations permettant d'initialiser le modèle sont telles que les prévisions ne sont pas toujours très précises.

La solution parait donc évidente : il faut utiliser le meilleur des deux méthodes. C'est-à-dire combiner au mieux la connaissance théorique du système avec les observations effectuées sur ce système. Dans le cas des prévisions météorologiques, cela consiste à initialiser la dynamique du modèle à l'aide de toutes les observations passées. C'est-à-dire que le la dynamique du modèle est contrainte par les observations jusqu'à atteindre l'instant présent où une prévision, à proprement parlé, est lancée.

2.2 Un exemple plus théorique

Une autre façon de penser l'assimilation de données est de la comparer aux méthodes traditionnelles de la recherche scientifique.

En général, face à un système inconnu, la première démarche est d'essayer de comprendre comment il fonctionne. Ceci se traduit par le développement d'un modèle qui va tenter de simuler toutes les composantes internes du système. Ce modèle, aussi perfectionné soit-il, nécessite des paramètres d'entrée. Une fois le modèle abouti, il est ensuite comparé à des résultats expérimentaux. En supposant que le modèle est adapté, les comparaisons entre les sorties du modèle et les mesures expérimentales permettent d'améliorer les entrées du modèle. Il s'agit donc d'un problème inverse.

Une autre approche, inspirée de l'automatisme, est de considérer le système inconnu comme une boîte noire. En faisant évoluer les paramètres d'entrées du système et en étudiant les variations des mesures expérimentales obtenues en sortie, il est possible de construire des lois appelées généralement des fonctions de transfert. Ces lois ne décrivent pas le fonctionnement interne du système mais lient seulement les entrées aux sorties. De ce fait, si le système est complexe et non-linéaire, il est alors difficile d'obtenir les sorties prévues. Comme dans l'optique de l'automatisme, l'objectif est de contrôler les entrées afin d'obtenir les sorties espérées, il suffit de faire une boucle de rétroaction qui compare les sorties obtenues aux sorties désirées en temps réel et de corriger en conséquence les entrées su système. Il s'agit donc encore une fois d'un problème inverse.

L'approche de l'assimilation de données est encore différente. Elle nécessite qu'un modèle paramétrique ait déjà été développé. Elle ne peut donc intervenir qu'après une étude sur le système inconnu. Ensuite, elle envisage la comparaison des sorties du modèle et des mesures expérimentales sous l'angle probabiliste afin d'estimer les paramètres d'entrée du modèle. Il s'agit donc aussi d'un problème inverse pour lequel les erreurs (c'est-à-dire les incertitudes) sur les sorties du modèle, sur les mesures expérimentales et sur le modèle lui-même doivent être estimées. Cette méthode consiste à obtenir le meilleur du modèle et des mesures expérimentales. En effet, le meilleur n'est pas toujours d'obtenir des sorties du modèles très proches des mesures expérimentales si cela se fait au détriment de la dynamique du système et des sorties du modèle qui n'ont pas d'équivalent en mesures expérimentales. En effet, aux erreurs de mesures s'ajoutent généralement des erreurs de représentativité dues au caractère discontinu des modèles numériques et des erreurs liées à la transformation des sorties du modèle afin d'obtenir un équivalent des observations (de la simple interpolation jusqu'à la transformation des variables des sorties du modèle pour obtenir des grandeurs comparables aux mesures expérimentales). La comparaison avec des mesures expérimentales est donc très différente de la comparaison avec un état « vrai ». L'assimilation de données est donc particulièrement adaptée aux systèmes de grandes tailles pour lesquels les modèles doivent être simplifiés (simplification des lois physiques et discrétisation ne permettant pas de résoudre des processus fins) ou pour les systèmes pour lesquels les observations sont parcellaires et inhomogènes. Elle reste cependant aussi adaptée à des problèmes plus « simples ».

L'**assimilation de données** est donc définie comme l'ensemble des techniques statistiques qui permettent d'améliorer la connaissance de l'état d'un système à partir de sa connaissance théorique et des observations expérimentales.

L'**interpolation statistique** est alors une technique permettant de trouver une solution à ce problème. Ces techniques sont souvent mathématiquement assez simple (équivalent à la méthode des moindres carrés), mais techniquement complexe du fait de la taille des systèmes.

Chapitre 3

Présentation du problème

> Il est très triste qu'il y aie si peu d'informations inutiles de nos jours.
>
> *Oscar Wilde*

3.1 Concepts de base

3.1.1 L'analyse

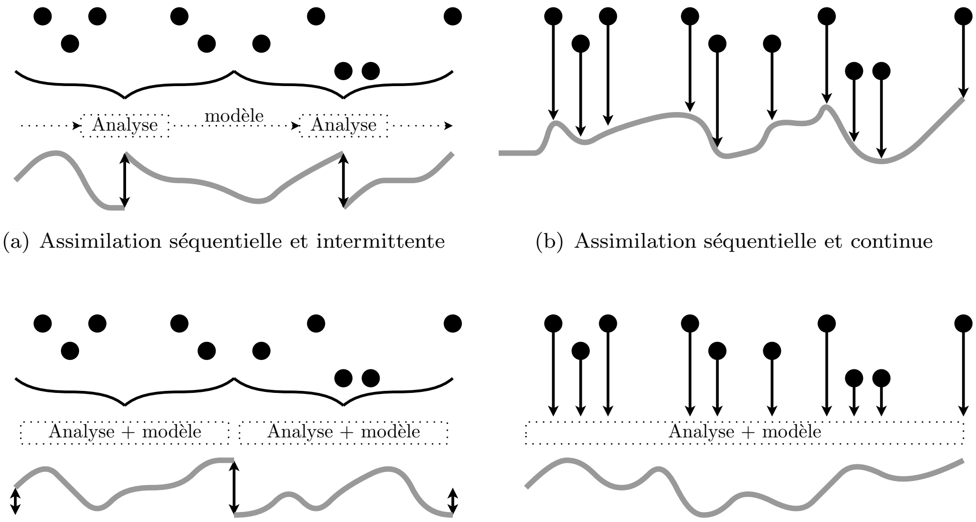
L'analyse est une description fiable de l'état vrai du système à un instant donné. Elle est déjà utile par elle même en tant que représentation globale et consistante du système étudié. Elle peut aussi servir comme état initial pour une prévision du système à l'aide du modèle ou comme pseudo-observation. Elle peut aussi servir de référence afin de vérifier la qualité des observations.

Pour obtenir l'état analysé, les seules informations objectives sont les mesures des observations effectuées sur l'état vrai. Le système peut parfois être surdéterminé. Dans ce cas, l'analyse se résume à un problème d'interpolation. Il est, en général, sous-déterminé car les observations sont clairsemées et pas toujours liées directement aux variables du modèle. Ce qui n'empêche pas d'avoir des régions où les observations sont très denses et où le système est ainsi sur-déterminé. Afin de bien poser le problème, il est nécessaire de disposer d'une ébauche de l'état du modèle (c'est-à-dire une estimation *a priori* de l'état du modèle). Des contraintes physiques peuvent aussi permettre de mieux déterminer le système. Cette ébauche peut aussi bien être une climatologie, un état quelconque ou état obtenu à partir de précédentes analyses. Dans ce cas, si le système efficace, l'information est sensée s'accumuler dans l'état du système et se propager entre les variables du modèle.

3.1.2 L'assimilation de données

L'assimilation de données est une technique d'analyse pour laquelle les informations apportées par les observations sont accumulées dans l'état du modèle grâce à des contraintes cohérentes avec les lois d'évolution temporelle et grâce aux propriétés physiques.

Les approches de l'assimilation de données peuvent être décrites, basiquement, de deux manières différentes : l'assimilation séquentielle et variationnelle. L'approche séquentielle suppose que toutes les observations proviennent du passé relativement à l'analyse. C'est une approche tout à fait appropriée pour les systèmes d'assimilation de données en temps réel. Elles s'appuient sur l'étude statistique des états du système afin de déterminer celui qui, statistiquement, est le plus adapté aux observations. En d'autres termes, ces méthodes permettent d'effectuer une analyse à chaque temps où une observation est disponible afin d'estimer l'état vrai du système à cet instant. L'approche variationnelle suppose que des observations provenant du futur par rapport à l'analyse sont aussi utilisables. Ce type de méthode est particulièrement adapté aux ré-analyses. Il est possible de décrire les méthodes d'assimilation de données en distinguant si elles sont intermittentes dans le temps ou continues. Les méthodes dites intermittentes découpent le temps en petites périodes sur lesquelles une analyse est effectuée. Ces méthodes sont très pratiques techniquement parlant. Les méthodes dites continues utilisent de très longues périodes sur lesquelles elles effectuent l'analyse. L'intérêt est d'obtenir un état analysé respectant mieux la dynamique physique et l'évolution temporelle du modèle. Ces différentes méthodes sont résumées succinctement dans la Fig. 3.1.



(a) Assimilation séquentielle et intermittente

(b) Assimilation séquentielle et continue

(c) Assimilation variationnelle et intermittente

(d) Assimilation variationnelle et continue

FIG. 3.1 – Représentation des quatre méthodes caractéristiques pour l'assimilation de données en fonction du temps. Les points noirs représentent les observations utilisées par la méthode d'assimilation pour obtenir une série d'états analysés, continus ou non, représentée par le trait gris.

3.1.3 Analyse de Cressman and Co

Il est possible de définir une méthode d'analyse telle que l'état analysé soit égale aux observations dans leur voisinage et égale à un état arbitraire partout ailleurs. Par exemple une climatologie ou une précédente prévision. Cette méthode s'apparente au schéma d'analyse de Cressman qui est souvent utilisé pour des systèmes d'assimilation simples.

L'état du modèle est supposé être univarié et représenté par les valeurs aux points de grille. En définissant $\mathbf{x}^b$ comme une estimation *a priori* de l'état du modèle provenant d'une climatologie, d'une persistance ou d'une prévision antérieure et $\mathbf{y}_i^o$ comme une série de N observations d'un même paramètre, une simple analyse de Cressman permet d'obtenir un état analysé du modèle $\mathbf{x}^a$ défini en chaque point de grille j telle que :

$$\mathbf{x}_j^a = \mathbf{x}_j^b + \frac{\sum_{i=1}^N w_{i,j}(\mathbf{y}_i^o - \mathbf{x}_i^b)}{\sum_{i=1}^N w_{i,j}} \quad (3.1)$$

$$w_{i,j} = \max\left(0, \frac{R^2 - d_{i,j}^2}{R^2 + d_{i,j}^2}\right) \quad (3.2)$$

où $d_{i,j}$ est la distance entre les points i et j, $\mathbf{x}_i^b$ est l'état d'ébauche interpolé au point i et $w_{i,j}$ est une fonction de poids dont le maximum est égale à un quand le point de grille j est situé sur l'observation i et qui décroît en fonction de la distance entre i et j pour devenir nulle quand $d_{i,j} > R$. R est donc défini comme un rayon d'influence au-delà duquel les observations n'ont plus d'influence. Un exemple d'une analyse de Cressman mono-dimensionnelle est représenté Fig. 3.2.

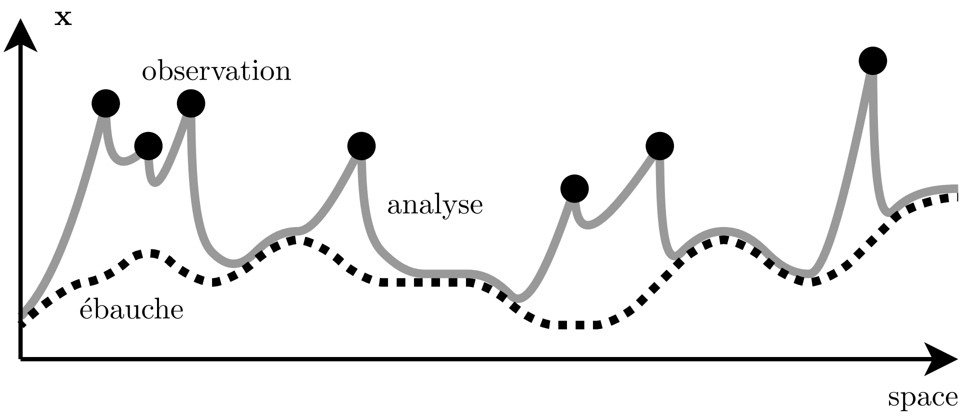


Fig. 3.2 – Exemple d'une analyse de Cressman mono-dimensionnelle. L'état de l'ébauche est représenté par la courbe noire pointillée, les observations par les points noirs et l'état analysé par la courbe grise.

D'autres variantes de la méthode de Cressman existent. La fonction de poids peut être, par exemple, redéfinie telle que $w_{i,j} = \exp(-d_{i,j}^2/2R^2)$. Une méthode particulièrement connue est l'*observation nudging* qui a plusieurs caractéristiques intéressantes. La fonction de poids est toujours inférieure à un même quand le point de grille i est superposé à l'observation j. Il s'agit alors d'une moyenne pondérée entre l'ébauche et les observations. Il est aussi possible de faire plusieurs mises à jour et donc d'utiliser l'Eq. 3.1 de manière itérative afin d'obtenir une correction de l'état analysé plus lisse.

3.1.4 Approche statistique

Malgré tout l'intérêt que la méthode de Cressman et ses dérivées peuvent avoir, elles restent trop limitées. En effet, si la première estimation de l'analyse (l'ébauche) est de bonne qualité tandis que les observations sont, quant à elles, de moins bonnes qualités, le remplacement de cette bonne estimation par les observations n'est pas profitable. D'autre part, il est difficile de définir la fonction de rappel w vers les observations car il n'y a aucune raison objective de choisir une forme plutôt qu'une autre. Enfin, l'état analysé doit respecter certaines caractéristiques de l'état vrai. Les variations des champs sont parfois limitées. Il existe des relations physiques entre les différentes variables. Ces contraintes physiques ne sont pas prises en compte par ce type de méthode et les corrections apportées par l'analyse peuvent parfois générer des structures non physiques.

Du fait de sa simplicité, ce type de méthode est vite limité mais reste très utile comme point de départ. Pour obtenir une analyse de bonne facture, les ingrédients sont aujourd'hui bien connus. Il faut d'abord une bonne première estimation de l'état du système. Une précédente analyse ou une prévision sont un bon choix en tant qu'ébauche. Ensuite, quand les observations sont nombreuses, leur moyenne est souvent proche de la valeur vraie. Il faut donc faire un bon compromis entre l'ébauche et les observations. Il faut être capable de donner un poids plus important aux observations de confiance et minimiser l'impact des observations suspicieuses. L'état analysé doit rester suffisamment lisse car l'état vrai l'est. Il faut donc que les observations aient une influence sur une région de la taille des phénomènes physiques mis en jeux et que cette influence diminue doucement pour revenir vers l'ébauche. L'analyse doit enfin être capable de tenir compte des structures physiques connues et aussi d'être capable de reconnaître des événements extrêmes pour ne pas les limiter car ils sont aussi très importants.

Les informations utilisées sont donc les observations, l'ébauche et propriétés physiques connues du système. Toutes ces sources d'informations sont importantes et doivent être prises en compte pour obtenir une bonne analyse. Par ailleurs, toutes ces sources d'information sont empreintes d'erreur et il n'est pas possible de leurs faire entièrement confiance. Il faut donc réussir un compromis, mais comme il y a des erreurs dans le modèle et dans les observations, il est difficile de savoir en quelles sources d'information avoir confiance. L'idée est donc de construire un système qui tente de minimiser en moyenne l'écart entre l'état analysé et vrai.

Pour construire ce type d'algorithme, il faut représenter mathématiquement l'incertitude sur les différentes sources d'information. Cette incertitude peut être définie en mesurant (ou en supposant) les statistiques d'erreur et modélisée avec des probabilités. L'algorithme d'analyse peut alors être écrit pour que, formellement, l'erreur d'analyse soit, en moyenne, minimale dans une norme définie par l'utilisateur. L'analyse devient ainsi un problème d'optimisation.

3.2 Description des vecteurs et espaces

3.2.1 Vecteur d'état

Pour formaliser mathématiquement le problème d'analyse, il faut définir un espace de travail. L'état du modèle est défini par une série de nombres ordonnés en une matrice colonne appelé vecteur d'état. Ce vecteur d'état $\mathbf{x}$ est le même que celui utilisé dans le modèle de prévision. Le choix de la discrétisation détermine comment les composantes vectorielles sont liées à l'état vrai du système. La discrétisation est ainsi équivalant au choix d'une base en mathématique.

Plusieurs vecteurs d'état peuvent être définis. Il n'est pas possible de décrire la réalité parfaitement car elle ne peut pas être représentée dans un vecteur d'état. Néanmoins, un vecteur $\mathbf{x}_c^t$ pourra permettre de décrire cet état vrai en continu. Plus classiquement, $\mathbf{x}^t$ représentera la meilleure représentation possible de la réalité $\mathbf{x}_c^t$ dans un vecteur d'état nommé *état vrai*. La première estimation de l'état analysé est appelé l'*ébauche* $\mathbf{x}^b$, tandis que l'état analysé lui-même est noté $\mathbf{x}^a$.

3.2.2 Variable de contrôle

En général, l'analyse n'est pas effectuée sur le vecteur d'état. En effet, il est souvent intéressant de ne pas résoudre le problème en prenant en compte toutes les variables du vecteur d'état. Ceci pour plusieurs raison. D'abord, certaines variables sont difficiles à prendre en compte. Ensuite, il est parfois souhaitable de réduire la taille du problème pour des questions de coûts informatiques. En effet, les besoins des modèles de prévision évoluent avec l'évolution des capacités informatiques en augmentant la résolution ou en améliorant la sophistication. Les méthodes d'assimilation sont alors d'autant plus coûteuses et il est primordiale réduire ces coûts. Au lieu de résoudre le problème dans l'espace du modèle, il est alors judicieux de le résoudre dans l'espace permettant de corriger l'ébauche. C'est l'espace des variables de contrôle. Le problème n'est plus de trouver l'état analysé $\mathbf{x}^a$, mais l'incrément vers l'état analysé $\delta\mathbf{x}$ tel que

$$\mathbf{x}^a = \mathbf{x}^b + \delta\mathbf{x} \tag{3.3}$$

le plus proche possible de $\mathbf{x}^t$. Une simple translation permet alors de transformer le problème. Au lieu de regarder $\mathbf{x}^a$, il faut regarder $(\mathbf{x}^a - \mathbf{x}^b)$ dans le sous-espace adéquat.

3.2.3 Observations

Chaque analyse utilise un nombre limité d'observations qui sont rangées dans un vecteur d'observation $\mathbf{y}^o$. Leur utilisation dans la méthode d'analyse dépend de la possibilité de les comparer avec le vecteur d'état. Dans l'idéal, il y aurait une observation pour chaque variable du vecteur d'état. En pratique, le nombre d'observations est très inférieur à celui des variables du vecteur d'état. De plus, les observations sont disposées très irrégulièrement. Il faut donc définir une fonction permettant de passer de l'espace du modèle à celui des observations. Cette fonction, nommée opérateur d'observation H, permet d'obtenir un équivalent du modèle $H\mathbf{x}$ dans l'espace des observations. Cette équivalent modèle est la valeur que devrait avoir l'observation si la mesure de l'observation et l'état du modèle étaient parfaits. En pratique, l'opérateur d'observation H est construit sur la base d'opérateurs d'interpolation et d'opérateurs permettant de transformer les variables du modèle en paramètres observés. Par exemple, en météorologie, il peut transformer les températures des différents niveaux de pression en une radiance mesurée par les satellites.

3.2.4 Représenteur

L'élément essentiel de l'assimilation de données est la comparaison entre les observations et le vecteur d'état. En utilisant les éléments introduits dans le paragraphe précédent, il est possible de définir un représenteur $\mathbf{y} - H\mathbf{x}$ aux points d'observations. Ce représenteur est appelé vecteur d'innovations quand le vecteur d'état est l'ébauche et vecteur des résidus d'analyse quand le vecteur d'état est l'analyse. L'étude des ces différents vecteurs permet de déterminer la qualité de la méthode d'analyse.

3.3 Modélisation des erreurs

Pour prendre en compte les incertitudes dans l'ébauche, les observations et l'analyse, il faut faire des hypothèses sur la modélisation des erreurs entre ces vecteurs et leurs équivalents « vrais ». L'utilisation des fonctions de densité de probabilité, nommées pdf, est une approche adaptée pour construire des modèles d'erreur. Les fonctions de densité de probabilité sont largement et rigoureusement décrites à travers des théories mathématiques. Une description simplifiée est donnée ci-dessous.

3.3.1 Représentation de l'incertitude par des pdfs

En prenant l'exemple du vecteur d'ébauche, il est possible de définir un seul et unique vecteur d'erreur séparant le vecteur d'ébauche de l'état vrai discrétisé :

$$\boldsymbol{\epsilon}^b \quad = \quad \mathbf{x}^b - \mathbf{x}^t. \tag{3.4}$$

Si la méthode d'analyse pouvait être répétée un très grand nombre de fois, dans des conditions strictement identiques mais avec des erreurs différentes dues à des causes aléatoires, le vecteur d'erreur $\boldsymbol{\epsilon}^b$ serait différent à chaque fois, mais il serait possible de calculer des grandeurs comme les moments (moyennes, variances...) ou de construire un histogramme des fréquences. Pour un très grand nombre d'expériences, ces estimateurs statistiques devraient converger vers des grandeurs ne dépendant que des processus physiques responsables des erreurs. A la limite, l'histogramme des fréquences peut être construit avec des classes infinitésimales et converger avec une fonction de densité de probabilité. Toutes les statistiques nécessaires sont alors dérivables de cette fonction de densité de probabilité et, entre autre, dans la plupart des problèmes physique, la moyenne et les variances. En effet, mathématiquement, une fonction de densité de probabilité n'a pas forcément ces grandeurs. Dans la suite, toutes les fonction de densité de probabilité seront supposées avoir ces grandeurs caractéristiques. Parmi toutes les fonction de densité de probabilité, la fonction Gaussienne est particulièrement intéressante.

3.3.2 Variables d'erreur

Il est possible de définir les erreurs d'ébauche, d'observation et d'analyse telle que présentées dans le Tab. 3.1.

L'erreur d'ébauche est la différence entre la première estimation de l'état du modèle et l'état vrai. Cette erreur n'inclue pas l'erreur de discrétisation.

Nom	Définition	Moyenne	Covariances
Erreur d'ébauche	$\boldsymbol{\epsilon}^b = \mathbf{x}^b - \mathbf{x}^t$	$\overline{\boldsymbol{\epsilon}}^b$	$\mathbf{B} = E[(\boldsymbol{\epsilon}^b - \overline{\boldsymbol{\epsilon}}^b)(\boldsymbol{\epsilon}^b - \overline{\boldsymbol{\epsilon}}^b)^T]$
Erreur d'observation	$\boldsymbol{\epsilon}^o = \mathbf{y}^o - H\mathbf{x}^t$	$\overline{\boldsymbol{\epsilon}}^o$	$\mathbf{R} = E[(\boldsymbol{\epsilon}^o - \overline{\boldsymbol{\epsilon}}^o)(\boldsymbol{\epsilon}^o - \overline{\boldsymbol{\epsilon}}^o)^T]$
Erreur d'analyse	$\boldsymbol{\epsilon}^a = \mathbf{x}^a - \mathbf{x}^t$	$\overline{\boldsymbol{\epsilon}}^a$	$\mathbf{A} = E[(\boldsymbol{\epsilon}^a - \overline{\boldsymbol{\epsilon}}^a)(\boldsymbol{\epsilon}^a - \overline{\boldsymbol{\epsilon}}^a)^T]$

TAB. 3.1 – Définition des variables d'erreur.

L'erreur d'observation est définie comme la différence entre les observations et l'équivalent modèle de l'état vrai. Cette erreur contient l'erreur de mesure ($\boldsymbol{\epsilon}^m$) due aux imprécisions de l'appareil de mesure vis-à-vis de la réalité, l'erreur due à l'opérateur d'observation H ($\boldsymbol{\epsilon}^i$) et l'erreur de représentativité ($\boldsymbol{\epsilon}^r$).

En effet, il est possible de définir le vecteur d'observation $\mathbf{y}^o$ comme la somme d'observations vraies $\mathbf{y}^t$ et de l'erreur de mesure $\boldsymbol{\epsilon}^m$.

$$\mathbf{y}^o = \mathbf{y}^t + \boldsymbol{\epsilon}^m \tag{3.5}$$

Il est ensuite possible d'écrire ces observations vraies $\mathbf{y}^t$ comme construites à partir d'un opérateur d'observation continu $\mathcal{H}$ et d'un état vrai et continu du modèle $\mathbf{x}_c^t$:

$$\mathbf{y}^o = \mathcal{H}\mathbf{x}_c^t + \boldsymbol{\epsilon}^m \tag{3.6}$$

L'état vrai continu du modèle peut être décomposé en portions discontinues résolues $\mathbf{x}^t$ et en portions discontinues non résolues. Les portions discontinues étant la projection de $\mathbf{x}_c^t$ dans l'espace de dimensions finies du modèle. L'équation 3.6 peut alors s'écrire :

$$\mathbf{y}^o = \mathcal{H}\mathbf{x}^t + \boldsymbol{\epsilon}^m + \boldsymbol{\epsilon}^r \tag{3.7}$$

où $\boldsymbol{\epsilon}^r$ représente les portions non résolues de l'état continu, c'est-à-dire l'erreur de représentativité.

L'équation 3.7 peut ensuite d'écrire :

$$\mathbf{y}^o = H\mathbf{x}^t + \boldsymbol{\epsilon}^m + \boldsymbol{\epsilon}^r + \boldsymbol{\epsilon}^i \tag{3.8}$$

où H est l'opérateur discret d'observation et $\boldsymbol{\epsilon}^i$ son erreur associée.

Enfin, l'erreur d'analyse est définie comme la différence entre l'état analysé et l'état vrai. La trace de cette matrice permet de définir une estimation de l'erreur de l'état analysé qui peut servir comme objet de minimisation. Par la suite, l'assimilation de données s'attachera à minimiser cette grandeur :

$$\mathrm{Tr}(\mathbf{A}) = \overline{||\boldsymbol{\epsilon}^a - \overline{\boldsymbol{\epsilon}}^a||^2}. \tag{3.9}$$

La moyenne de ces différentes erreurs sera appelé le biais et représente un problème systématique dans le système d'assimilation qui peut être une dérive du modèle, un biais dans les observations ou encore une erreur systématique dans la manière d'utiliser les observations. Il est important de comprendre que le biais est de même nature statistique que l'état du modèle ou le vecteur d'observation. Son interprétation est simple et les opérateurs linéaires utilisés pour le vecteur d'état du modèle ou celui d'observation peuvent lui être appliqué.

3.3.3 Covariances d'erreur

Les covariances d'erreur sont un petit plus compliquées que les erreurs de biais et seront illustrées par l'exemple des erreurs d'ébauche. Cependant, toutes les remarques peuvent aussi s'appliquer aux erreurs d'observation.

Dans un système scalaire, les erreurs de covariances se résument aux variances. Par contre, dans un système multidimensionnel, les covariances peuvent être décrites par une matrice carrée symétrique. Si le vecteur d'état du modèle est de dimension n, alors la matrice de covariance d'erreur est de dimension $n \times n$. La diagonale de cette matrice est alors constituée des variances de chaque variable du modèle et les termes non-diagonaux sont les covariances entre deux des variables du vecteur d'état du modèle. En définissant, pour trois variables, les erreurs d'ébauche non-biaisées ($\overline{\boldsymbol{\epsilon}}$) telles que $(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_k)$, alors la matrice de covariance d'erreur d'ébauche $\mathbf{B}$ s'écrit :

$$\mathbf{B} = \begin{pmatrix} \ddots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \mathrm{var}(\boldsymbol{\epsilon}_i) & \mathrm{cov}(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) & \mathrm{cov}(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_k) & \cdots \\ \cdots & \mathrm{cov}(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) & \mathrm{var}(\boldsymbol{\epsilon}_j) & \mathrm{cov}(\boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_k) & \cdots \\ \cdots & \mathrm{cov}(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_k) & \mathrm{cov}(\boldsymbol{\epsilon}_j, \boldsymbol{\epsilon}_k) & \mathrm{var}(\boldsymbol{\epsilon}_k) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \ddots \end{pmatrix} \quad (3.10)$$

En général, les variances sont non-nulles, car il est difficile de supposer que l'ébauche puisse représenter certains aspects de la réalité de manière parfaite. Dans ce cas, la matrice est alors définie positive et il est possible de réécrire les termes non-diagonaux de la matrice de covariance d'erreur sous forme de corrélation d'erreurs :

$$\rho(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j) = \frac{\mathrm{cov}(\boldsymbol{\epsilon}_i, \boldsymbol{\epsilon}_j)}{\sqrt{\mathrm{var}(\boldsymbol{\epsilon}_i)\mathrm{var}(\boldsymbol{\epsilon}_j)}}. \quad (3.11)$$

À la différence des erreurs de biais, il n'est pas possible d'appliquer les opérateurs linéaires utilisés sur le vecteur d'état du modèle ou sur le vecteur d'observation afin de transformer le champ de variances d'erreur (la diagonale de la matrice de covariance d'erreur). Il faut, en fait, définir des transformations linéaires avec des matrices pleines. Par exemple, si une transformation linéaire est définie par une matrice P telle que les nouvelles coordonnées de la transformation de $\mathbf{x}$ soient $P\mathbf{x}$, alors la matrice de covariance de cette nouvelle variable est $P\mathbf{x}P^T$.

Chapitre 4

Interpolation statistique

Le vrai génie réside dans l'aptitude à évaluer l'incertain, le hasardeux, les informations conflictuelles.

sir Winston Leonard Spencer Churchill

4.1 Notations et hypothèses

Soit l'espace du modèle de dimensions n et l'espace des observations de dimensions p. En reprenant les définitions et notations précédentes :

$\mathbf{x}^t$ est l'état du modèle vrai de dimension n ;

$\mathbf{x}^b$ est l'état de l'ébauche de dimension n ;

$\mathbf{x}^a$ est l'état analysé de dimension n ;

$\mathbf{y}^o$ est le vecteur d'observation de dimension p ;

H est l'opérateur d'observation passant de la dimension n à p ;

$\mathbf{B}$ est la matrice de covariance d'erreur d'ébauche $(\mathbf{x}^b - \mathbf{x}^t)$ de dimension $n \times n$;

$\mathbf{R}$ est la matrice de covariance d'erreur d'observation $(\mathbf{y} - H\mathbf{x}^t)$ de dimension $p \times p$;

$\mathbf{A}$ est la matrice de covariance d'erreur d'analyse $(\mathbf{x}^a - \mathbf{x}^t)$ de dimension $n \times n$.

De plus, certaines hypothèses sont émises telles que :

opérateur d'observation linéarisé : Les variations de l'opérateur d'observation au voisinage de l'ébauche est linéaire. Ainsi, pour tout $\mathbf{x}$ suffisamment proche de $\mathbf{x}^b$, $H\mathbf{x} - H\mathbf{x}^b = \mathbf{H}(\mathbf{x} - \mathbf{x}^b)$ où $\mathbf{H}$ est un opérateur linéaire ;

erreurs non triviales : les matrices $\mathbf{B}$ et $\mathbf{R}$ sont définies positives ;

erreurs non biaisées : les moyennes des erreurs d'ébauche et d'observation sont nulles ($E[\mathbf{x}^b - \mathbf{x}^t] = E[\mathbf{y} - H\mathbf{x}^t] = 0$) ;

erreurs non-corrélées : les erreurs d'ébauche et d'observation sont mutuellement décorrélées ($E[(\mathbf{x}^b - \mathbf{x}^t)(\mathbf{y} - H\mathbf{x}^t)^T] = 0$) ;

analyse linéaire : les corrections apportées à l'ébauche dépendent linéairement de l'innovation ;

analyse optimale : l'état analysé doit être aussi proche que possible de l'état vrai dans le sens du minimum de variance.

4.2 Le problème de l'estimation

Le système étudié est décrit par $\mathbf{x}^t$. La première estimation faite est $\mathbf{x}^b$ qui peut, par exemple, provenir d'une analyse antérieure. C'est la meilleure estimation du système en l'absence d'autres informations. Des observations $\mathbf{y}^o$ permettent d'obtenir des renseignements partiels au travers de l'opérateur d'observation non-linéaire H. Dans la suite de ce chapitre, l'opérateur d'observation sera considéré linéaire et noté $\mathbf{H}$. De plus, les erreurs d'ébauche $\boldsymbol{\epsilon}^b$ et d'observation $\boldsymbol{\epsilon}^o$ sont non-biaisées (ou débiaisées) et leurs statistiques sont connues.

L'objectif est alors, à l'aide des observations, d'améliorer l'estimation du système $\mathbf{x}^a$ vis-à-vis de la première estimation $\mathbf{x}^b$. De plus, l'erreur sur l'état du système analysé $\boldsymbol{\epsilon}^a$ est aussi recherchée.

Plusieurs possibilités existent pour obtenir l'analyse. Cependant, l'objectif est de réaliser une analyse aussi bonne qui possible (voire la meilleure). Il faut donc minimiser l'erreur commise *a posteriori* $\boldsymbol{\epsilon}^a$, soit par exemple en minimisant $\mathrm{Tr}(\mathbf{A})$.

4.3 Estimation non-optimisée

4.3.1 Définition du gain

Un ansatz pour le vecteur d'analyse $\mathbf{x}^a$ est de le décomposer de la manière suivante :

$$\mathbf{x}^a = \mathbf{L}\mathbf{x}^b + \mathbf{K}\mathbf{y}^o, \tag{4.1}$$

où $\mathbf{L}$ est une matrice de dimension $n \times n$ et $\mathbf{K}$ une matrice de dimension $n \times p$. L'état analysé est donc une combinaison linéaire des différentes informations disponibles.

Son erreur associée peut être obtenu facilement en soustrayant $\mathbf{x}^t$ à l'Eq. 4.1.

$$\begin{aligned} \mathbf{x}^a - \mathbf{x}^t &= \mathbf{L}(\mathbf{x}^b - \mathbf{x}^t + \mathbf{x}^t) + \mathbf{K}(\mathbf{H}\mathbf{x}^t + \boldsymbol{\epsilon}^o) - \mathbf{x}^t \\ \boldsymbol{\epsilon}^a &= \mathbf{L}\boldsymbol{\epsilon}^b + \mathbf{K}\boldsymbol{\epsilon}^o + (\mathbf{L} + \mathbf{K}\mathbf{H} - \mathbf{I})\mathbf{x}^t \end{aligned} \tag{4.2}$$

En recherchant une erreur d'analyse sans biais et comme les erreurs d'ébauche et d'observation sont, par hypothèse, non biaisées, l'Eq. 4.2 donne :

$$\begin{aligned} E(\boldsymbol{\epsilon}^a) &= \mathbf{L}E(\boldsymbol{\epsilon}^b) + \mathbf{K}E(\boldsymbol{\epsilon}^o) + (\mathbf{L} + \mathbf{K}\mathbf{H} - \mathbf{I})E(\mathbf{x}^t), \\ E(\boldsymbol{\epsilon}^a) &= (\mathbf{L} + \mathbf{K}\mathbf{H} - \mathbf{I})E(\mathbf{x}^t). \end{aligned} \tag{4.3}$$

Il faut donc que $\mathbf{L} + \mathbf{KH} - \mathbf{I} = 0$. L'ansatz prend donc la forme simplifiée suivantes :

$$\begin{aligned} \mathbf{x}^a &= (\mathbf{I} - \mathbf{KH})\mathbf{x}^b + \mathbf{K}\mathbf{y}^o \\ \mathbf{x}^a &= \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b) \end{aligned} \tag{4.4}$$

On retrouve donc l'innovation $\mathbf{y}^o - \mathbf{H}\mathbf{x}^b$ définit précédemment et un opérateur linéaire $\mathbf{K}$ allant de l'espace des observations à l'espace du modèle généralement appelé gain.

Comme $\mathbf{K}$ est linéaire, l'analyse est donc une régression linéaire. Néanmoins, on parle, en général, d'interpolation linéaire pour des raisons historiques. En effet, les premières méthodes d'analyse, comme l'analyse de Cressman présentée dans la section 3.1.3, étaient réellement des interpolations linéaires au sens mathématique du terme. Pour obtenir une bonne analyse, il suffit donc de déterminer un gain satisfaisant.

4.3.2 Erreur d'analyse commise

En supposant le gain $\mathbf{K}$ connu, il est possible de calculer l'erreur d'analyse commise en partant de l'Eq. 4.1 et en utilisant les erreurs introduites auparavant :

$$\boldsymbol{\epsilon}^a = \mathbf{L}\boldsymbol{\epsilon}^b + \mathbf{K}\boldsymbol{\epsilon}^o. \tag{4.5}$$

Comme les erreurs d'ébauche et d'observation sont décorrélées, alors la matrice de covariance d'erreur d'analyse est alors égale à :

$$\begin{aligned} \mathbf{A} &= E\left[(\boldsymbol{\epsilon}^a)(\boldsymbol{\epsilon}^a)^T\right] \\ &= E\left[(\mathbf{L}\boldsymbol{\epsilon}^b + \mathbf{K}\boldsymbol{\epsilon}^o)(\mathbf{L}\boldsymbol{\epsilon}^b + \mathbf{K}\boldsymbol{\epsilon}^o)^T\right] \\ &= E\left[\mathbf{L}(\boldsymbol{\epsilon}^b)(\boldsymbol{\epsilon}^b)^T\mathbf{L}^T\right] + E\left[\mathbf{K}(\boldsymbol{\epsilon}^o)(\boldsymbol{\epsilon}^o)^T\mathbf{K}^T\right] \\ &= \mathbf{LBL}^T + \mathbf{KRK}^T. \end{aligned} \tag{4.6}$$

En reprenant l'estimation de $\mathbf{L} = \mathbf{I} - \mathbf{KH}$ obtenue avec l'hypothèse d'une erreur d'analyse sans biais, la matrice de covariance d'erreur d'analyse s'écrit alors sous la forme :

$$\mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B}(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T. \tag{4.7}$$

4.4 Best Linear Unbiased Estimation

Best Linear Unbiased Estimation Connaissant la matrice de covariance d'erreur d'analyse, il est possible d'essayer de minimiser son erreur scalaire ($\mathrm{Tr}(\mathbf{A})$). Il doit donc exister un gain optimal $\mathbf{K}^*$ qui peut être obtenu en étudiant la variation de l'erreur scalaire d'analyse sous une variation du gain. Comme la trace est une fonction scalaire continue et différentiable des coefficients de $\mathbf{K}$, il est possible d'exprimer sa dérivé $d_{\mathbf{K}}$ au premier ordre :

$$\begin{aligned} d_{\mathbf{K}}\left(\mathrm{Tr}(\mathbf{A})\right) &= \mathrm{Tr}\left(-\mathbf{HBL}^T - \mathbf{LBH}^T + \mathbf{RK}^T + \mathbf{KR}\right) \\ &= \mathrm{Tr}\left(\mathbf{RK}^T - \mathbf{HBL}^T\right) + \mathrm{Tr}\left(\mathbf{KR} - \mathbf{LBH}^T\right) \\ &= 2\mathrm{Tr}\left(\mathbf{KR} - \mathbf{LBH}^T\right) \end{aligned} \tag{4.8}$$

L'équation 4.8 est obtenue en utilisant des propriétés de l'algèbre linéaire telles que la trace est linéaire ($\mathrm{Tr}(\mathbf{B}+\alpha\mathbf{R}) = \mathrm{Tr}(\mathbf{B})+\alpha\mathrm{Tr}(\mathbf{R})$), la trace de la transposée égale la trace ($\mathrm{Tr}(\mathbf{B}^T) = \mathrm{Tr}(\mathbf{B})$) et les matrices symétriques sont égales à leurs transposées ($\mathbf{B}^T = \mathbf{B}$).

Pour obtenir le gain optimal $\mathbf{K}^*$, il faut que $d_{\mathbf{K}}\left(\mathrm{Tr}(\mathbf{A})\right) = 0$. L'équation 4.8 donne alors le résultat suivant :

$$
\begin{aligned}
\mathbf{K}^*\mathbf{R} - \mathbf{L}\mathbf{B}\mathbf{H}^T &= 0, \\
\mathbf{K}^*\mathbf{R} - (\mathbf{I} - \mathbf{K}^*\mathbf{H})\mathbf{B}\mathbf{H}^T &= 0, \\
\mathbf{K}^*(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T) &= \mathbf{B}\mathbf{H}^t. \qquad (4.9)
\end{aligned}
$$

A l'optimalité, on a donc un gain égale

$$\mathbf{K}^* = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}. \qquad (4.10)$$

Avec ce gain optimal, il est alors possible d'estimer $\mathbf{x}^a$ et $\mathbf{A}$. C'est une estimation **BLUE** (Best Linear Unbiased Estimation) car elle est linéaire (Eq. 4.1), sans biais (Eq. 4.3) et optimale (Eq. 4.9).

4.5 Propriétés du BLUE

4.5.1 Formule de Sherman-Morrison-Woodbury

Le gain optimal est, en général, donné sous la forme de l'Eq. 4.10. Cependant, il peut être réécrit sous la forme :

$$
\begin{aligned}
\mathbf{K}^* &= \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}, \\
&= \left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}, \\
&= \left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\left(\mathbf{H}^T + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}\mathbf{H}^T\right)(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}, \\
&= \left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\left(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T\right)(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}, \\
&= \left(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}. \qquad (4.11)
\end{aligned}
$$

Cette autre formule est du gain optimal permet de changer l'espace dans lequel il faut faire une inversion matricielle. En effet, dans l'Eq. 4.10, il faut inverser $\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T$ dans l'espace des observations, tandis qu'avec l'Eq. 4.11, il faut maintenant inverser $\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$ dans l'espace du modèle. Comme, en général, l'espace des observations est plus petit que celui du modèle, l'Eq. 4.10 est beaucoup plus souvent utilisée car bien moins coûteuse.

4.5.2 Erreur d'analyse optimale

L'erreur d'analyse a été obtenue par l'Eq. 4.7. En réinjectant la valeur du gain optimal dedans, il est possible d'obtenir une erreur d'analyse optimale :

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}^*\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}^*\mathbf{H})^T + \mathbf{K}^*\mathbf{R}\mathbf{K}^{*T},$$

$$
\begin{aligned}
&= (\mathbf{I}-\mathbf{K}^*\mathbf{H})\mathbf{B}-(\mathbf{I}-\mathbf{K}^*\mathbf{H})\mathbf{B}\mathbf{H}^T\mathbf{K}^{*T}+\mathbf{K}^*\mathbf{R}\mathbf{K}^{*T},\\
&= (\mathbf{I}-\mathbf{K}^*\mathbf{H})\mathbf{B}+\left(\mathbf{K}^*\mathbf{R}+\mathbf{K}^*\mathbf{H}\mathbf{B}\mathbf{H}^T-\mathbf{B}\mathbf{H}^T\right)\mathbf{K}^{*T},\\
&= (\mathbf{I}-\mathbf{K}^*\mathbf{H})\mathbf{B}+\left(\mathbf{K}^*(\mathbf{R}+\mathbf{H}\mathbf{B}\mathbf{H}^T)-\mathbf{B}\mathbf{H}^T\right)\mathbf{K}^{*T},\\
&= (\mathbf{I}-\mathbf{K}^*\mathbf{H})\mathbf{B}+\left(\mathbf{B}\mathbf{H}^T(\mathbf{R}+\mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\mathbf{R}+\mathbf{H}\mathbf{B}\mathbf{H}^T)-\mathbf{B}\mathbf{H}^T\right)\mathbf{K}^{*T},\\
&= (\mathbf{I}-\mathbf{K}^*\mathbf{H})\mathbf{B}. \qquad (4.12)
\end{aligned}
$$

Il est aussi possible de calculer le résidu d'analyse en réutilisant l'Eq. 4.4 :

$$
\begin{aligned}
\mathbf{y}^o-\mathbf{H}\mathbf{x}^a &= \mathbf{y}^o-\mathbf{H}\left(\mathbf{x}^b+\mathbf{K}(\mathbf{y}^o-\mathbf{H}\mathbf{x}^b)\right)\\
&= (\mathbf{y}^o-\mathbf{H}\mathbf{x}^b)-\mathbf{H}\mathbf{K}(\mathbf{y}^o-\mathbf{H}\mathbf{x}^b)\\
&= (\mathbf{I}-\mathbf{H}\mathbf{K})(\mathbf{y}^o-\mathbf{H}\mathbf{x}^b) \qquad (4.13)
\end{aligned}
$$

En repartant de l'Eq. 4.11 et en réutilisant l'Eq. 4.12, il est possible de définir le gain optimal en fonction de l'erreur d'analyse. En effet :

$$
\begin{aligned}
\mathbf{K}^* &= \left(\mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1},\\
&= \left(\mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}\mathbf{H}\right)^{-1}\mathbf{B}^{-1}\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1},\\
&= \left(\mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}\mathbf{H}\right)^{-1}\left((\mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})-\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1},\\
&= \left(\mathbf{I}-(\mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1},\\
&= (\mathbf{I}-\mathbf{K}^*\mathbf{H})\,\mathbf{B}\mathbf{H}^T\mathbf{R}^{-1},\\
&= \mathbf{A}\mathbf{H}^T\mathbf{R}^{-1}. \qquad (4.14)
\end{aligned}
$$

Il est aussi possible de définir l'inverse de l'erreur d'analyse optimale en fonction des erreurs d'ébauche et d'observation. En effet, en reprenant l'Eq. 4.12 définissant l'erreur d'analyse optimale et en lui injectant le gain optimal obtenu avec l'Eq. 4.14, on obtient :

$$
\begin{aligned}
\mathbf{A} &= (\mathbf{I}-\mathbf{K}^*\mathbf{H})\mathbf{B},\\
\mathbf{A} &= (\mathbf{I}-\mathbf{A}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})\mathbf{B},\\
\mathbf{A} &= \mathbf{B}-\mathbf{A}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B},\\
\mathbf{A}(\mathbf{I}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B}) &= \mathbf{B},\\
(\mathbf{I}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B})^{-1}\mathbf{A}^{-1} &= \mathbf{B}^{-1},\\
\mathbf{A}^{-1} &= (\mathbf{I}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{B})\mathbf{B}^{-1},\\
\mathbf{A}^{-1} &= \mathbf{B}^{-1}+\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}. \qquad (4.15)
\end{aligned}
$$

Cette équation est intéressante car elle veut dire que les matrices de confiance (les inverses des matrices de covariance d'erreurs) sont additives. En d'autres termes, tout apport d'information, quelque soit sa qualité objective, augmente forcément la confiance dans l'état analysé. D'autres formulations du gain optimal ou des l'erreur d'analyse sont encore possibles.

4.5.3 Représenteur sans biais

Il est possible d'écrire l'innovation en fonction des erreurs d'ébauche et d'observation :

$$\begin{aligned}
\mathbf{y}^o - \mathbf{H}\mathbf{x}^b &= \mathbf{H}\mathbf{x}^t + \boldsymbol{\epsilon}^o - \mathbf{H}\mathbf{x}^b, \\
&= \boldsymbol{\epsilon}^o - \mathbf{H}\boldsymbol{\epsilon}^b. \qquad (4.16)
\end{aligned}$$

Comme les erreurs d'ébauche et d'observation sont supposées sans biais, alors il apparaît logiquement que l'innovation est aussi de biais nul ($E[\mathbf{y}^o - \mathbf{H}\mathbf{x}^b] = 0$). De la même manière, le biais du résidu d'analyse est, lui aussi, nul ($E[\mathbf{y}^o - \mathbf{H}\mathbf{x}^a] = 0$).

4.5.4 Corrélation de l'analyse et de son erreur

Une dernière autre caractéristique intéressante du BLUE peut être obtenue en calculant la covariance entres l'analyse et son erreur $E[\mathbf{x}^a(\boldsymbol{\epsilon}^a)^T]$. En supposant que l'ébauche est décorrélée de son erreur $E[\mathbf{x}^b(\boldsymbol{\epsilon}^b)^T] = 0$ ainsi que de l'erreur d'observation $E[\mathbf{x}^b(\boldsymbol{\epsilon}^o)^T] = 0$, que les erreurs d'ébauche et d'observation sont mutuellement décorrélées $E[\boldsymbol{\epsilon}^b(\boldsymbol{\epsilon}^o)^T] = 0$ et en utilisant les Eqs. 4.4, 4.5 et 4.16, alors :

$$\begin{aligned}
E[\mathbf{x}^a(\boldsymbol{\epsilon}^a)^T] &= E\left[\left(\mathbf{x}^b + \mathbf{K}(\boldsymbol{\epsilon}^o - \mathbf{H}\boldsymbol{\epsilon}^b)\right)\left((\mathbf{I}-\mathbf{K}\mathbf{H})\boldsymbol{\epsilon}^b + \mathbf{K}\boldsymbol{\epsilon}^o\right)^T\right] \\
&= E\left[\left(\mathbf{K}(\boldsymbol{\epsilon}^o - \mathbf{H}\boldsymbol{\epsilon}^b)\right)\left((\mathbf{I}-\mathbf{K}\mathbf{H})\boldsymbol{\epsilon}^b + \mathbf{K}\boldsymbol{\epsilon}^o\right)^T\right] \\
&= \mathbf{K}E[\boldsymbol{\epsilon}^o(\boldsymbol{\epsilon}^o)^T]\mathbf{K}^T - \mathbf{K}\mathbf{H}E[\boldsymbol{\epsilon}^b(\boldsymbol{\epsilon}^b)^T](\mathbf{I}-\mathbf{K}\mathbf{H})^T \\
&= \mathbf{K}\left(\mathbf{R}\mathbf{K}^T - \mathbf{H}\mathbf{B}(\mathbf{I}-\mathbf{K}\mathbf{H})^T\right) \qquad (4.17)
\end{aligned}$$

A l'optimalité et en reprenant l'Eq. 4.9, on trouve alors que $E[\mathbf{x}^a(\boldsymbol{\epsilon}^a)^T] = 0$. Ainsi, sous ces hypothèses classiques, l'analyse est décorrélées de son erreur. En mathématique, on parle alors d'orthogonalité. Le BLUE associe donc optimalité avec orthogonalité.

4.6 Approche variationnelle

4.6.1 Équivalence avec le BLUE

En reprenant exactement les mêmes hypothèses que pour le BLUE, il est possible de résoudre le problème par une approche variationnelle. Pour cela, il faut définir une fonctionnelle :

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}^b)\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) + \frac{1}{2}(\mathbf{y}^o - \mathbf{H}\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}), \qquad (4.18)$$

appelée fonction coût et qui a pour caractéristique d'être quadratique en $\mathbf{x}$. Comme, de plus, les matrices $\mathbf{B}$ et $\mathbf{R}$ sont définies positives, alors cette fonction coût est convexe et possède un seul minimum qui peut être estimé par son gradient :

$$\nabla J(\mathbf{x}) = \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) - \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}). \qquad (4.19)$$

Le point selle $\mathbf{x}^*$ est donc logiquement obtenu pour $\nabla J(\mathbf{x}) = 0$:

$$\begin{aligned}
\mathbf{B}^{-1}(\mathbf{x}^* - \mathbf{x}^b) - \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^*) &= 0,\\
\mathbf{B}^{-1}(\mathbf{x}^* - \mathbf{x}^b) + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{x}^* - \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}^o &= 0,\\
\mathbf{B}^{-1}(\mathbf{x}^* - \mathbf{x}^b) + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}(\mathbf{x}^* - \mathbf{x}^b) &= \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}^o - \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{x}^b,\\
(\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})(\mathbf{x}^* - \mathbf{x}^b) &= \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b),
\end{aligned}$$

et la forme obtenue (Eq. 4.20) en écrivant cette égalité par rapport à x^* est identique à l'Eq. 4.11 :

$$\mathbf{x}^* = \mathbf{x}^b + (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b). \tag{4.20}$$

La formulation variationnelle est donc strictement identique au BLUE à l'optimalité.

4.6.2 Hessien

A partir du gradient de la fonction coût (Eq. 4.19), il est facile d'en déduire le Hessien :

$$\nabla^2 J(\mathbf{x}) = \mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}. \tag{4.21}$$

En comparant avec l'Eq. 4.15, il est possible de réécrire l'Eq. 4.21 en fonction de la matrice de covariance d'erreur d'analyse :

$$\mathbf{A} = \left(\nabla^2 J(\mathbf{x})\right)^{-1}. \tag{4.22}$$

Cette nouvelle formulation de la matrice de covariance d'erreur d'analyse est particulièrement intéressante, car elle permet de comprendre que la qualité de l'analyse est proportionnelle à la convexité de la fonction coût. Moins la fonction coût sera convexe, moins bonne sera l'analyse. Il est donc très important de formuler le problème de manière à obtenir une fonction très convexe.

4.6.3 Extension des méthodes variationnelles

A partir de la formulation variationnelle (Eq. 4.18), il est très facile de remplacer l'opérateur d'observation linéaire $\mathbf{H}$ par un opérateur non linéaire H. Dans ce cas, il faut introduire l'opérateur tangent-linéaire de H en $\mathbf{x}$ noté $\mathbf{H}^T$. La fonction coût devient alors

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y}^o - H\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{y}^o - H\mathbf{x}), \tag{4.23}$$

et son gradient

$$\nabla J(\mathbf{x}) = \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) - \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}^o - H\mathbf{x}). \tag{4.24}$$

Cette extension de la méthode variationnelle à des cas non-linéaires est donc très simple, alors qu'elle n'est pas possible dans le cadre du BLUE.

Il est, par ailleurs, possible de modifier la fonction coût de telle sorte que la nouvelle fonctionnelle s'optimise dans l'espace des observations et non dans celui du modèle. Ce formalisme se nomme PSAS pour Physical Space Assimilation System et est particulièrement utile lorsque le nombre d'observations est plus faible que le nombre de variables du modèle.

Enfin, la formulation variationnelle a un avantage important sur le BLUE en cela qu'elle ne nécessite pas d'inverser la matrice $\mathbf{R} + \mathbf{HBH}^T$. En effet, pour minimiser la fonction coût, il suffit de calculer le produit d'un vecteur par les inverses de $\mathbf{B}$ et$\mathbf{R}$. Ce qui, au niveau algorithmique, est beaucoup plus rapide.

4.7 L'exemple du naufragé

Pour illustrer les différents concepts abordés, un exemple sera très utile. Supposons qu'à la suite d'une tempête, un marin naviguant en suivant la ligne de côtes, s'échoue sur des récifs. Le bateau étant bien équipé, il relève sa dernière position sur le GPS et monte dans le canot de sauvetage. Malheureusement, ce canot est dépourvu de rames. Les vagues et le vent l'emportent donc loin de son navire échoué.

Définissons un référentiel de tel sorte que l'axe x soit parallèle à la côte et l'axe y lui soit perpendiculaire. La position du navire échoué dans ce référentiel est défini comme le point de référence de coordonnées $(0, 0)$. La position du canot de sauvetage est donc connu à l'instant $t_0 = 0$. Un peu plus tard, à l'instant t, le naufragé estime au jugé la distance qui le sépare de la côte. Le naufragé sait que son estimation est empreinte d'une erreur et il estime la variance de cette erreur s^o. Il se rappelle, par ailleurs, la position de l'épave et sait que le canot de sauvetage a dérivé malgré l'absence de courants marins prédominants dans cette région. Il suppose donc que la probabilité qu'il se trouve maintenant à la position (u^b, v^b) suit une loi normale de variance s^b qui dépend linéairement du temps écoulé. Après réflexion, il estime aussi que le processus de mesure au jugé n'est pas corrélé à celui de la dérive du canot. Il résume donc sa situation en faisant un schéma (Fig. 4.1).

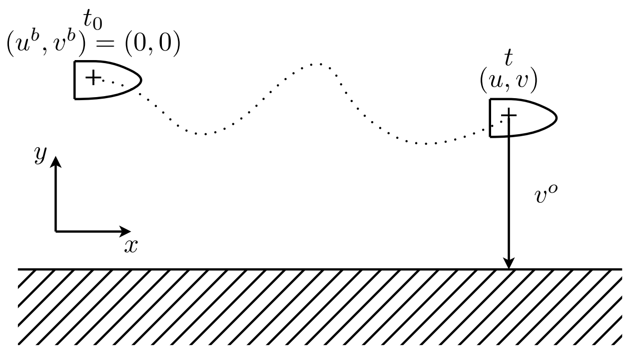


FIG. 4.1 – La géométrie de la situation du naufragé.

4.7.1 Définition des variables

Dans cette situation, le naufragé ne perdant pas son sang-froid, définit les variables lui permettant de résoudre son problème. Il commence par construire un vecteur d'état $\mathbf{x} = (u, v)^T$ et le vecteur d'observation $\mathbf{y}^o = v^o$. Il n'a, en effet, aucune possibilité de mesurer sa position le long de la côte. Dans ce cadre, il peut définir son opérateur d'observation simplement par $\mathbf{H} = (0,1)$. Comme il a estimé la variance de l'erreur de sa mesure, il construit la matrice de covariance d'erreur sur l'observation telle que $\mathbf{R} = (s^o)$. Finalement, il construit sa matrice de covariance d'erreur d'ébauche en utilisant son hypothèse sur une dérive suivant une loi normale et il obtient :

$$\mathbf{B} = \begin{pmatrix} s^b & 0 \\ 0 & s^b \end{pmatrix}.$$

4.7.2 Analyse optimale

Comme ce navigateur lisait régulièrement des ouvrages d'assimilation de données, il sait qu'il peut faire une estimation de sa nouvelle position en utilisant le BLUE. Il commence par écrire l'analyse de manière théorique $\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)$ puis l'applique à sa situation :

$$\begin{pmatrix} u^a \\ v^a \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \mathbf{K}\left(v^o - (0,1)\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right) = \begin{pmatrix} 0 \\ \mathbf{K}v^o \end{pmatrix}.$$

Connaissant l'Eq. 4.10, il sait que $\mathbf{K}^* = \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}$. Il calcule donc le gain optimal $\mathbf{K}^*$:

$$\begin{aligned} \mathbf{K}^* &= s^b \begin{pmatrix} 0 \\ 1 \end{pmatrix} \left(s^o + (0,1)s^b \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)^{-1}, \\ &= s^b \begin{pmatrix} 0 \\ 1 \end{pmatrix} \left(s^o + s^b\right), \\ &= \frac{s^b}{s^o + s^b} \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{aligned} \tag{4.25}$$

Le malheureux naufragé peut alors estimer plus précisément sa position et en déduit que :

$$\begin{pmatrix} u^a \\ v^a \end{pmatrix} = \frac{s^b}{s^o + s^b} \begin{pmatrix} 0 \\ v^b \end{pmatrix}.$$

Il constate donc logiquement que n'ayant pas apporté d'information sur sa position le long de la côte, la meilleure estimation possible est de conserver sa position au moment du naufrage $u^a = 0$. Il constate aussi que plus le temps s'écoule, plus l'erreur sur son ébauche s^b grandit de sorte que la meilleure estimation de la distance le séparant de la côte tend vers son observation au jugé ($\lim_{t\to\infty} v^a = v^o$).

4.7.3 Erreur commise

De plus, notre naufragé connaît aussi la formule liant la matrice de covariance d'erreur d'analyse à celle d'ébauche (Eq. 4.12) qui est $\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}$ et l'estime alors pour son cas :

$$\mathbf{A} = \begin{pmatrix} s^b & 0 \\ 0 & \frac{s^o s^b}{s^o + s^b} \end{pmatrix}.$$

Il lui apparaît alors encore plus clairement que son analyse ne permet pas d'augmenter sa connaissance sur sa position le long de la côte et que son incertitude est celle de sa dérive. Il apprend, par contre, que son analyse permet de réduire l'incertitude sur la distance le séparant de la côte d'un facteur $s^o/(s^o + s^b)$. Ainsi, plus il sera capable d'obtenir une observation précise, plus il réduira l'erreur. En effet, comme le montre l'Eq. 4.15, en notant s^a la variance de l'erreur sur la coordonnée v après analyse, on obtient :

$$\frac{1}{s^a} = \frac{1}{s^o} + \frac{1}{s^b}.$$

Il apparaît encore plus clairement que l'ajout d'une information augmente forcément la qualité de l'analyse. Et ceci, quelque soit la qualité de l'information.

Après ces différents résultats, le naufragé est bien avancé et décide d'utiliser son énergie à trouver une solution à son autre problème : regagner la terre ferme.

Chapitre 5

Méthodes d'assimilation

> Assimilation :
> Dans le raisonnement en général. Action de comparer en allant du différent au semblable pour rechercher ou établir l'identité.
>
> *Le Trésor de la Langue Française informatisé*

Dans la présentation du problème au chapitre 3, différentes familles de méthodes d'assimilation ont été introduites. Dans ce chapitre, les méthodes d'assimilation les plus connues et populaires seront décrites. L'interpolation optimale sera présentée en premier lieu. Ensuite, les différentes méthodes utilisant le filtre de Kalman seront décrites. Finalement, les méthodes variationnelles seront présentées. Ces différentes méthodes seront donc détaillées avec un effort particulier pour mettre en évidence les avantages et inconvénients de chacune. Les aspects pratiques seront particulièrement pris en compte, notamment les coûts informatiques et les difficultés d'implémentation.

À la différence de la présentation du BLUE, qui a permis de comprendre l'essentiel des problèmes de l'assimilation de données et pour laquelle tous les résultats ont été détaillés afin de prendre en main les concepts, les présentations de ces méthodes ne détailleront pas les calculs nécessaires permettant d'aboutir aux formules mise en œuvre en pratique.

5.1 Interpolation Optimale - OI

Interpolation Optimale L'interpolation optimale (Gandin, 1963 ; Lorenc, 1981 ou Daley, 1991)), notée OI, est une simplification algébrique du BLUE présenté en section 4.4. L'équation 4.4 est décomposée en un système d'équations résolvant cette équation pour chaque variable du modèle. L'hypothèse fondamentale de cette méthode est que pour chaque variable du modèle en chaque point de grille, un nombre réduit d'observations est pris prend en compte pour effectuer l'analyse. L'approximation vient donc de la technique de sélection d'une liste de données p_v utiles pour l'analyse de chaque variable $\mathbf{x}_v$ en chaque point de grille. Le calcul de $\mathbf{K}$ se fait ligne par ligne en n'utilisant qu'un nombre réduit d'observations voisines de chaque point de grille (Fig. 5.1). La matrice de gain $\mathbf{K}$ est construite par bloc et sa formulation

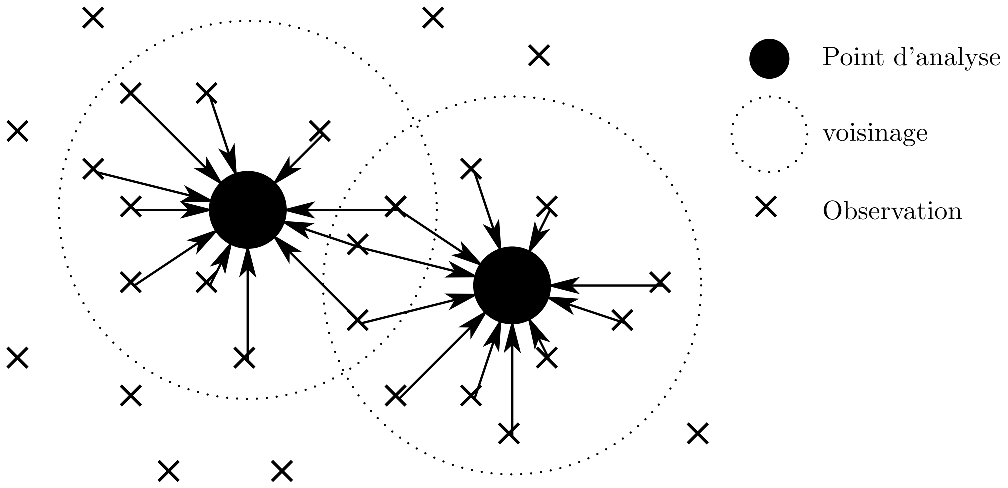


FIG. 5.1 – Représentation schématique de la sélection des observations au voisinage de deux points. Les analyses en ces deux points n'utilisent pas les mêmes observations bien qu'ils soient proches. Les champs analysé n'est donc généralement pas continu. De plus, le coût de l'analyse augmente avec la taille du voisinage utilisé.

se fait à un coût réduit :

$$\mathbf{K} = (\mathbf{K}_1, \cdots, \mathbf{K}_v, \cdots, \mathbf{K}_V), \tag{5.1}$$

où v représente les points au voisinage du point d'analyse variant de 1 à V et où :

$$\mathbf{K}_v = \mathbf{B}\mathbf{H}^T\mathbf{P}_v^T \left(\mathbf{P}_v \left(\mathbf{H}^T\mathbf{B}\mathbf{H} + \mathbf{R}\right) \mathbf{P}_v^T\right)^{-1}. \tag{5.2}$$

La matrice $\mathbf{P}_v$ de l'Eq. 5.2, de dimension, $(p_v \times p)$ définit quelles sont les observations au voisinage du point d'analyse qui seront utilisées pour le calcul de $\mathbf{K}_v$. Le calcul de $\mathbf{K}$ se résume donc à l'inversion de V systèmes réduits de $(p_v \times p_v)$ équations. La solution donnée par l'OI est optimale à l'instant de l'analyse sur chaque voisinage, mais est sous-optimale relativement à l'ensemble des observations (sur la période de mesure et sur l'ensemble du domaine de mesure).

Avantages : méthode simple et bon marché

L'OI présente l'avantage de s'implémenter aisément et de fournir une analyse à moindre coût par rapport à l'application exacte du BLUE, notamment pour le calcul de la matrice de gain $\mathbf{K}$. D'autre part, la matrice $\mathbf{B}$ ne doit pas nécessairement être explicitée dans sa globalité. Elle peut être définie pour chaque variable par des fonctions empiriques d'autocorrélations constantes au cours du temps, des contraintes physiques du système telles que l'équilibre hydrostatique ou géostrophique en océanographie ou par des EOF[1] Empirical Orthogonal Function univariées ou multivariées

[1]EOF : Empirical Orthogonal Function

Inconvénient : analyse bruitée

L'inconvénient majeur de cette méthode vient du choix des voisinages d'influence des observations en chaque point de grille qui produit une analyse bruitée (Cohn *et al.*, 1998). Ceci peut être aggravé par le choix de spécifications incompatibles des erreurs d'ébauche pour chaque variable du modèle.

5.2 Les filtres de Kalman de rang plein

5.2.1 Filtre de Kalman - KF

Filtre de Kalman En 1960-61, Kalman et Bucy ont décrit une solution récursive pour des problèmes de filtrage linéaire de données discrète. Cette solution est depuis nommée filtre de Kalman. Ce filtre peut être appréhendé comme une extension du BLUE pour laquelle l'état analysé pour une étape donnée définit l'ébauche à l'étape d'analyse suivante. Outre ceci, le filtre de Kalman incorpore un modèle d'évolution de l'état du système entre deux instants t_i et t_{i+1}.

Pour décrire cette méthode, les notations usuellement utilisées seront reprises. En particulier, la matrice des covariances d'erreur d'analyse, jusqu'ici notée $\mathbf{A}$, se nommera maintenant $\mathbf{P}^a$ et celle d'ébauche, jusqu'ici notés $\mathbf{B}$, s'appellera $\mathbf{P}^f$ de manière à mettre en évidence que l'ébauche (b comme *background*) est maintenant une prévision (f comme *forecast*). Le modèle d'évolution non-linéaire sera noté M, linéarisé il se nommera $\mathbf{M}$ et son adjoint $\mathbf{M}^T$. L'utilisation d'un modèle d'évolution entre deux instant t_i et t_{i+1} entraîne un nouveau type d'erreur nommée l'erreur modèle. Elle est supposée non-biaisée et est décrite par la matrice de covariance d'erreur du modèle à chaque instant t_i : $\mathbf{Q}_i$. De plus, les erreurs d'analyse et modèle sont supposées non-corrélées.

Le schéma d'assimilation peut être décrit de la manière suivante : à partir d'une prévision à l'instant t_i et de sa matrice de covariance d'erreur de prévision $\mathbf{P}_i^f$, une analyse est effectuée permettant d'obtenir un état analysé et une matrice de covariances d'erreur d'analyse $\mathbf{P}_i^a$ à l'instant t_i. Ensuite, une prévision du temps t_i à t_{i+1} est effectuée en partant de l'état analysé. De manière similaire, la matrice de covariance d'erreur d'analyse est propagée par le modèle d'évolution linéaire afin d'obtenir la matrice de covariance d'erreur de prévision $\mathbf{P}_{i+1}^f$ à l'instant t_{i+1}. Il suffit ensuite de répéter cette opération.

La deuxième étape, durant laquelle l'état analysé et la matrice de covariances d'erreur d'analyse est propagée jusqu'au temps d'observation suivant, est clairement la plus coûteuse.

De manière plus formelle, l'algorithme du filtre de Kalman entre les instants d'observation t_i et t_{i+1} peut être décrit les Eqs. 5.3, 5.4, 5.5, 5.6 et 5.7.

- Calcul de la matrice de gain $\mathbf{K}$ au temps t_i :

$$\mathbf{K}_i = \mathbf{P}_i^f \mathbf{H}_i^T \left(\mathbf{H}_i \mathbf{P}_i^f \mathbf{H}_i^T + \mathbf{R}_i \right)^{-1}. \tag{5.3}$$

- Analyse au temps t_i :

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}_i \left(\mathbf{y}_i^o - \mathbf{H}_i \mathbf{x}_i^f \right). \tag{5.4}$$

- Calcul de la matrice de covariance d'erreur d'analyse au temps t_i :

$$\mathbf{P}_i^a = \left(\mathbf{I} - \mathbf{K}_i \mathbf{H}_i \right) \mathbf{P}_i^f. \tag{5.5}$$

– Prévision au temps t_{i+1} par propagation de l'analyse de t_i à t_{i+1} par le modèle linéaire d'évolution :

$$\mathbf{x}_{i+1}^f = \mathbf{M}_{i\to i+1}(\mathbf{x}_i^a). \tag{5.6}$$

– Calcul de la matrice de covariance d'erreur de prévision au temps t_{i+1} par propagation de la matrice de covariance d'erreur d'analyse de t_i à t_{i+1} par le modèle linéaire d'évolution :

$$\mathbf{P}_{i+1}^f = \mathbf{M}_{i\to i+1}\mathbf{P}_i^a\mathbf{M}_{i\to i+1}^T + \mathbf{Q}_i. \tag{5.7}$$

5.2.2 Filtre de Kalman Étendu - EKF

Filtre de Kalman étendu Dans le filtre de Kalman classique, le modèle d'évolution et l'opérateur d'observation sont supposés linéaires. Cependant, il arrive souvent que l'hypothèse de linéarité ne soit pas valide. Dans ce cas, il est possible de généraliser le filtre de Kalman en utilisant des formes linéarisées de l'opérateur d'observation et du modèle d'évolution pour les Eqs. 5.3, 5.5 et 5.7 et la forme non-linéaire pour les Eqs. 5.4 et 5.6. Ce filtre est appelé filtre de Kalman étendu (EKF).

Les cinq étapes de l'analyse peuvent alors s'écrire ainsi :

– Calcul de la matrice de gain $\mathbf{K}$ au temps t_i :

$$\mathbf{K}_i = \mathbf{P}_i^f\mathbf{H}_i^T\left(\mathbf{H}_i\mathbf{P}_i^f\mathbf{H}_i^T + \mathbf{R}_i\right)^{-1}. \tag{5.8}$$

– Analyse au temps t_i à l'aide le l'opérateur d'observation non-linéaire :

$$\mathbf{x}_i^a = \mathbf{x}_i^f + \mathbf{K}_i\left(\mathbf{y}_i^o - H_i\mathbf{x}_i^f\right). \tag{5.9}$$

– Calcul de la matrice de covariance d'erreur d'analyse au temps t_i :

$$\mathbf{P}_i^a = (\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\,\mathbf{P}_i^f. \tag{5.10}$$

– Prévision au temps t_{i+1} par propagation de l'analyse de t_i à t_{i+1} par le modèle non-linéaire d'évolution :

$$\mathbf{x}_{i+1}^f = M_{i\to i+1}(\mathbf{x}_i^a). \tag{5.11}$$

– Calcul de la matrice de covariance d'erreur de prévision au temps t_{i+1} par propagation de la matrice de covariance d'erreur d'analyse de t_i à t_{i+1} par le modèle linéaire d'évolution :

$$\mathbf{P}_{i+1}^f = \mathbf{M}_{i\to i+1}\mathbf{P}_i^a\mathbf{M}_{i\to i+1}^T + \mathbf{Q}_i. \tag{5.12}$$

A noter que, bien que le filtre de Kalman soit une analyse optimale, le filtre de Kalman étendu perd cette qualité (il ne fournit pas la solution de variance minimale). Néanmoins, l'utilisation du filtre de Kalman étendu dans un cadre faiblement non-linéaire permet d'obtenir de bonnes analyses. De plus, la linéarisation du modèle d'évolution M peut interagir avec les erreurs modèle de manière assez compliquée.

5.2.3 Coût de calcul

L'algorithme du filtre de Kalman complète le système d'équations lié à la détermination de l'état analysé et à sa propagation dans le temps avec deux équations de calcul et de propagation de la matrice de covariance d'erreur d'analyse. Le coût numérique du filtre de Kalman est donc la somme du coût du traitement du vecteur d'état et des covariances d'erreur. Pour les systèmes de grande taille tels que l'océan ou l'atmosphère, le coût de calcul principal provient du traitement des covariances d'erreur d'analyse. La première étape coûteuse est l'inversion de la matrice $\left(\mathbf{H}_i\mathbf{P}_i^f\mathbf{H}_i^T + \mathbf{R}_i\right)$. La propagation par les équations de la dynamique du modèle linéaire-tangent de $\mathbf{P}^a$ requiert ensuite la multiplication par la matrice $\mathbf{M}$ par chaque colonne (chaque ligne pour $\mathbf{M}^T$) de $\mathbf{P}^a$ (autour de $10^7 \times 10^7$ opérations). Au delà du coup de calcul exorbitant de ces opérations, il est impossible de stocker entre chaque étape d'analyse de telles matrices malgré les capacités déjà importantes disponibles. Pour ces raisons, l'algorithme du filtre de Kalman ne peut être appliqué qu'à des systèmes de taille réduite.

Il doit donc être simplifié pour permettre son application aux systèmes océaniques et atmosphériques. Plusieurs études visent notamment à réduire le nombre d'intégration du modèle linéaire-tangent en ne propageant pas la matrice de covariance d'erreur que suivant certaines directions (Fukomori *et al.*, 1995 ; Evensen, 1994 ; Fisher, 1998 et Evensen, 2003). Il faut tout d'abord identifier un sous-espace de dimension réduite. Ensuite, seule la projection de la matrice de covariance dans ce sous-espace, et non la matrice complète, est propagée. Divers de ces filtres seront présentés dans la section 5.3.

5.2.4 L'exemple du naufragé

Revenons aux mésaventures de notre naufragé introduites à la section 4.7. Finalement, ne sachant comment atteindre le rivage, il se résout à évaluer la distance le séparant du rivage toutes les heures. Il dispose ainsi de i mesures de la distance du canot au rivage (v_i^o) entre l'instant de son naufrage t_0 et la dernière mesure au temps t_i. Cette évaluation est supposée sans biais et sa variance, notée comme précédemment s^o, est supposée stationnaire. Les coordonnées réelles du canot sont (u_i, v_i), tandis que celle issues de l'analyse (u_i^a, v_i^a) et celles de la prévision (u_i^f, v_i^f). A l'instant du naufrage (t_0), la position du canot est $(u_0^a, v_0^a) = (0, 0)$. Entre deux mesures aux instants t_i et t_{i+1}, le canot dérive mais sa direction n'est pas connue. Le naufragé imagine donc un modèle d'évolution comme un modèle de diffusion autour de son point d'origine. Il peut donc écrire le modèle (linéaire en l'occurrence) tel que $\mathbf{M}_{i\to i+1} = \mathbf{I}$) et l'erreur modèle, qu'il suppose importante, telle que

$$\mathbf{Q}_i = \begin{pmatrix} s^m & 0 \\ 0 & s^m \end{pmatrix},$$

où s^m est proportionnel au temps écoulé entre t_{k+1} et t_0.

Le naufragé fait ensuite la supposition qu'il n'y a aucune corrélation entre les deux coordonnées de la position du canot de sauvetage. Il écrit alors les matrices de covariances d'erreur d'analyse et de prévision sous la forme :

$$\mathbf{P}_i^a = \begin{pmatrix} \lambda_i & 0 \\ 0 & \mu_i \end{pmatrix}, \quad \text{et} \quad \mathbf{P}_i^f = \begin{pmatrix} \nu_i & 0 \\ 0 & \rho_i \end{pmatrix}.$$

Analyse

En raisonnant à partir de l'instant t_i, le naufragé dispose du vecteur résultant de sa dernière analyse $(u_i^f, v_i^f)^T$. Il effectue alors une nouvelle mesure v_i^o et effectue une analyse optimale en utilisant le gain de Kalman (Eq. 5.3) :

$$\begin{aligned}\mathbf{K}_i^* &= \begin{pmatrix} \nu_i & 0 \\ 0 & \rho_i \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix}\left(s^o + (0,1)\begin{pmatrix} \nu_i & 0 \\ 0 & \rho_i \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right)^{-1} \\ &= \frac{\rho_i}{s^o + \rho_i}\begin{pmatrix} 0 \\ 1 \end{pmatrix}.\end{aligned}$$

Il peut alors en déduire les coordonnées de la position du canot (Eq. 5.4)

$$\begin{aligned}\begin{pmatrix} u_i^a \\ v_i^a \end{pmatrix} &= \begin{pmatrix} u_i^f \\ v_i^f \end{pmatrix} + \mathbf{K}_i^*\left(v_i^o - (0,1)\begin{pmatrix} u_i^f \\ v_i^f \end{pmatrix}\right) \\ &= \begin{pmatrix} u_i^f \\ v_i^f \end{pmatrix} + \mathbf{K}_i^*\left(v_i^o - v_i^f\right) \\ &= \begin{pmatrix} u_i^f \\ v_i^f + \frac{\rho_i}{s^o+\rho_i}\left(v_i^o - v_i^f\right) \end{pmatrix},\end{aligned}$$

et la matrice de covariance d'erreur d'analyse (Eq. 5.5)

$$\begin{aligned}\begin{pmatrix} \lambda_i & 0 \\ 0 & \mu_i \end{pmatrix} &= \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{\rho_i}{s^o + \rho_i}\begin{pmatrix} 0 \\ 1 \end{pmatrix}(0,1)\right)\begin{pmatrix} \nu_i & 0 \\ 0 & \rho_i \end{pmatrix} \\ &= \begin{pmatrix} \nu_i & 0 \\ 0 & \frac{s^o\rho_i}{s^o+\rho_i} \end{pmatrix}.\end{aligned} \tag{5.13}$$

Il en déduit alors que

$$\lambda_i = \nu_i \quad \text{et} \quad \frac{1}{\mu_i} = \frac{1}{s^o} + \frac{1}{\rho_i}.$$

Prévision

Le naufragé passe maintenant à l'étape de prévision du filtre de Kalman et applique son modèle d'évolution $\mathbf{I}$ (Eq. 5.6) pour estimer sa prochaine position

$$\begin{pmatrix} u_{i+1}^f \\ v_{i+1}^f \end{pmatrix} = \begin{pmatrix} u_i^a \\ v_i^a \end{pmatrix}.$$

Ensuite, il calcule sa prochaine matrice de covariance d'erreur de prévision (Eq. 5.7) facilement et trouve que $\mathbf{P}_{i+1}^f = \mathbf{P}_i^a + s^m\mathbf{I}$, ce qu'il écrit aussi sous la forme

$$\nu_{i+1} = \lambda_i + s^m \quad \text{et} \quad \rho_{i+1} = \mu_i + s^m.$$

Il résulte alors de la succession de l'analyse et de la prévision que

$$\begin{cases} u_{i+1}^f = u_i^f, \\ v_{i+1}^f = v_i^f + \frac{\rho_i}{s^o+\rho_i}\left(v_i^o - v_i^f\right), \end{cases}$$

et

$$\left\{\begin{array}{l}\nu_{i+1} = \nu_i + s^m,\\ \frac{1}{\rho_{i+1}-s^m} = \frac{1}{s^o} + \frac{1}{\rho_i}.\end{array}\right.$$

Comme le naufragé sait que $u_0^f = 0$, il en déduit que $u_k^f = 0$. Sans mesure, il n'apprend rien sur sa position le long de la côte. Cependant, l'incertitude croît linéairement avec le temps puisque $\nu_k = ks^m$.

De l'incertitude sur la coordonnée v, le naufragé en déduit que, dans un premier temps, la confiance sur l'analyse est la somme de la confiance sur l'observation et sur la prévision, et, dans un second temps, que l'erreur de la prévision est la somme de l'erreur modèle et d'analyse. C'est donc le résultat d'un compromis entre la réduction de l'incertitude par l'assimilation et son accroissement due à la dérive non contrôlée de l'embarcation.

5.3 Les filtres de Kalman réduits

Depuis R. E. Kalman, les filtres ont été utilisés dans de nombreuses applications. Mais très vite, les aspects limitants de l'implémentation du filtre de Kalman sont apparus. Ainsi, l'assimilation de données n'était pas possible dans des domaines comme la météorologie, ou plus tard, l'océanographie car les dimensions du problème rendaient excessif le coût numérique et, de plus, les statistiques nécessaires au filtre de Kalman ne sont que rarement connues.

Pour résoudre ce problème, une hypothèse peut permettre de le contourner. L'idée est, qu'à un instant donné, la physique du modèle est contrôlée par un nombre ou une combinaison limitée de variables. L'hypothèse est alors que les statistiques d'erreurs significatives sont données par celles portant sur ces variables contrôlant la physique du modèle (les modes réduits). Il est alors nécessaire des les identifier. De plus, il faut aussi être capable d'enrichir stochastiquement le système afin que la base de modes réduits puisse évoluer sans contraintes trop fortes. En effet, le risque est que ces modes, s'ils dégénèrent, ne sous-tendent plus la fraction de l'espace des états dans lequel évolue le système.

5.3.1 Filtre RRSQRT

Filtre de Kalman de rang réduit Le filtre RRSQRT est une réponde à ce problème. Il permet d'éviter les différentes difficultés d'implémentation mise en évidence auparavant en représentant les directions principales des matrices d'erreur par des modes réduits. Ainsi, il possible d'utiliser exclusivement les modes au détriment des matrices.

En reprenant l'espace du modèle de dimension n, avec un état du système initial $\mathbf{x}_0^f$ auquel est associé la matrice de covariance d'erreur $\mathbf{P}_0^f$, il faut réaliser une décomposition en mode principaux de cette matrice telle que

$$\mathbf{P}_0^f \quad \simeq \quad \mathbf{S}_0^f(\mathbf{S}_0^f)^T, \tag{5.14}$$

où $\mathbf{S}_0^f$ est une matrice de dimension $(m \times n)$ avec m représentant les mes m premiers modes principaux de $\mathbf{P}_0^f$. L'erreur sur l'ébauche a donc été réduite. Il est alors possible de définir un opérateur d'observation

$\mathbf{\Psi} = (H_i\mathbf{S}_i^f)^T$. Le gain de Kalman, calculé dans l'espace d'analyse, peut alors être décrit en fonction de $\mathbf{\Psi}^T$ de taille $(p \times m)$

$$\mathbf{K}_i^* \quad = \quad \mathbf{S}_i\mathbf{\Psi}\left(\mathbf{\Psi}^T\mathbf{\Psi} + \mathbf{R}_i\right)^{-1}. \tag{5.15}$$

Et, il est aussi possible d'obtenir la matrice racine de covariance d'analyse sans faire de calcul directement avec les matrices de covariance d'erreur :

$$\mathbf{S}_i^a \quad = \quad \mathbf{S}_i^f\left(\mathbf{I} - \mathbf{\Psi}\left(\mathbf{\Psi}^T\mathbf{\Psi} + \mathbf{R}_i\right)^{-1}\mathbf{\Psi}^T\right)^{1/2}. \tag{5.16}$$

Le calcul de la racine de $\mathbf{S}_i^a$ pourrait être coûteux, mais il n'en est rien puisque $\mathbf{I} - \mathbf{\Psi}\left(\mathbf{\Psi}^T\mathbf{\Psi} + \mathbf{R}_i\right)^{-1}\mathbf{\Psi}^T$ est de taille $(m \times m)$. De plus, la matrice racine est mieux conditionnée, ce qui assure une meilleure précision numérique.

Après l'analyse, la dimension du système est réduit en passant de m modes à $m - q$ modes. Pour cela, il suffit de diagonaliser $(\mathbf{S}_i^a)^T\mathbf{S}_i^a$ et de ne retenir que les $m - q$ plus grandes valeurs propres de la matrice de passage, puis de réduire $\mathbf{S}_i^a$.

À l'étape de prévision, l'état analysé est propagé par le modèle d'évolution et la racine de la matrice réduite $\tilde{\mathbf{S}}_i^a$ par le modèle linéaire-tangent. Elle est ensuite élargie en ajoutant q modes imputés à l'erreur modèle :

$$\mathbf{S}_{i+1}^f \quad = \quad [\mathbf{M}_{i\to i+1}\tilde{\mathbf{S}}_i^a(\mathbf{M}_{i\to i+1})^T, \mathbf{T}_i]. \tag{5.17}$$

Cette matrice comporte alors m modes.

5.3.2 Filtre SEEK

Singular Evolutive Extended Kalman filter Le filtre SEEK[2] a été introduit par Pham *et al.* en 1998. Il s'agit d'un filtre réduit déduit du filtre de Kalman étendu. Il repose sur la stagnation ou la décroissance du rang des matrices de covariances d'erreur, une propriété avérée ou forcée selon les cas.

Dans le cas d'un filtre sans erreur modèle, il résulte des Eqs. 5.10 et 5.14 que le rang de $\mathbf{P}_i^f$, noté $r = \text{rang}(\mathbf{P}_i^f)$, est une fonction décroissante de t_i, puisque, *in fine*, la récurrence est de la forme $\mathbf{P}_{i+1}^f = \mathbf{A}_i\mathbf{P}_i^f\mathbf{B}_i$. En conséquence, si le rang de la matrice de covariance d'erreur initiale est faible comparée à la dimension n de l'espace du modèle, il le restera. Il est alors possible de décomposer la matrice $\mathbf{P}_i^a$ avec une matrice diagonale à coefficient positifs ou nuls et une matrice orthogonale décrivant les r directions principales d'erreur. À partir de cette décomposition, il est possible de poser le problème dans l'espace des directions principales d'erreur de taille très inférieure à la dimension du système original. L'analyse est alors effectuée dans cet espace réduit et a pour caractéristique de ne pas modifier l'espace engendré par les directions principales des erreurs. Ce qui n'empêche pas, en général, à ces directions de changer.

Dans un cadre plus général, le modèle n'est pas parfait. Il n'est pas possible de négliger $\mathbf{Q}_i$. Ainsi, il apparaît que le rang de $\mathbf{P}_{i+1}^f$ peut être supérieur à celui de $\mathbf{P}_i^a$. De plus, il ne peut plus y avoir de réduction du filtre sans approximation. La solution la plus naturelle et la plus simple pour réduire le rang

[2]Singular Evolutive Extended Kalman filter

consiste à projeter $\mathbf{Q}_i$ avec une projection orthogonale. C'est-à-dire que seule la composante de l'erreur modèle agissant dans le sous-espace sur lequel agit $\mathbf{P}_i^a$ est retenue.

L'idée de filtre SEEK peut donc être résumé à ceci : le système d'évolution M amplifie les erreurs associées à un sous-espace $\mathcal{A}_i$ de l'espace tangent à l'espace modèle, tandis que les erreurs associées au complémentaire de cet espace sont atténuées. Si le rang de la matrice de covariance d'erreur est supérieur à la dimension de $\mathcal{A}_i$, alors il est possible d'espérer que le système d'évolution se chargera d'atténuer toutes les erreurs commises et non corrigées dans l'espace complémentaire de $\mathcal{A}_i$.

5.3.3 Filtre SEIK

Singular Evolutive Interpolation Kalman filter Le filtre de Kalman étendu peut présenter des instabilités en présence de fortes non-linéarités jusqu'à, parfois, diverger complètement (Evensen, 1992 ; Gauthier *et al.*, 1994 et Kushner, 1967). Une possibilité pour tenter de résoudre cette difficulté est de remplacer la linéarisation dans le filtre de Kalman étendu par un développement de Taylor d'ordre supérieur. Malheureusement, cette approche n'est pas envisageable sur des systèmes de grandes dimensions comme l'océanographie. Une autre approche est possible en utilisant des méthodes stochastiques de type Monte Carlo pour estimer l'évolution de la matrice de covariance d'erreur par un nuage d'états centrés autour de l'état courant et donc la matrice de covariance empirique est celle de la matrice considérée. Cette approche, introduite par Evensen en 1994 avec son filtre de Kalman d'ensemble, est un très bon moyen pour traiter les modèles d'évolution fortement non-linéaires. Cette méthode sera présentée dans la section 5.3. Néanmoins, cette méthode est limitée par la taille de l'échantillon à considérer.

En 2001, Pham *et al.* ont proposé une variante du filtre de SEEK, appelé filtre de Kalman Singulier Évolutif Interpolé (SEIK), dans lequel la taille de l'échantillon est, en un certain sens, minimale. En effet, il substitue à la linéarisation opérée dans le filtre de Kalman étendu et dans le SEEK une interpolation sur un échantillon d'états bien choisis propagés dans l'étape de prévision. L'idée du SEIK est donc de faire évoluer la matrice de covariance d'erreur à l'aide d'un nuage de points de taille raisonnable. Dans ce but, Pham a émit l'hypothèse de rang faible r de la matrice de covariance d'erreur pour réduire la taille du nuage de points à $r+1$ points exactement. L'autre originalité de ce filtre réside dans le choix des états d'interpolation qui sont tirés « au hasard » à chaque pas de filtrage afin de ne pas privilégier une direction particulière de l'espace d'état. La Fig. 5.2 permet de mettre en évidence les différentes étapes nécessaire au filtre SEIK.

5.3.4 Coût de Calcul et filtres dégradés

Les différents filtres de Kalman à rang réduit représentent des approches réalistes permettant l'implémentation du filtre de Kalman à des problèmes complexes et de grandes tailles.

En effet, alors que l'utilisation des filtres de Kalman ou de Kalman étendu nécessitent des ressources informatiques hors de portée pour des problèmes de grandes tailles comme l'océanographie ou la météorologie, le passage à un sous-espace représentatif de taille beaucoup plus petite permet la mise en œuvre réaliste des méthodes utilisant cette technique.

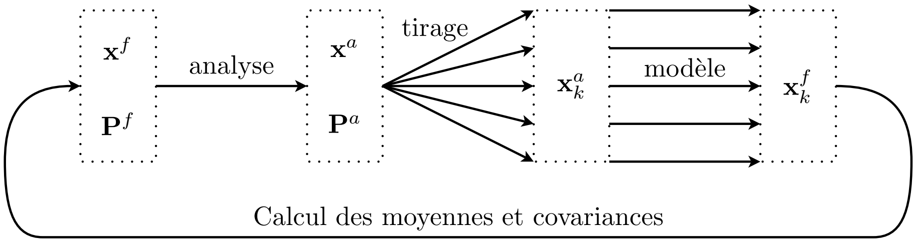


FIG. 5.2 – Représentation schématique des différentes étapes du filtre SEIK lors d'un cycle d'assimilation du temps t_i au temps t_{i+1}. L'indice k variant de 1 à $r+1$ représente les différents membres du nuage de points.

L'intérêt pour ces différentes formes d'assimilation de données séquentielles varie en fonction des besoins et capacités. Autant l'interpolation optimale est simple à mettre en œuvre et d'un coût extrêmement réduit, autant les filtres SEEK et SEIK représentent déjà un investissement important dans le cadre de problèmes de grandes tailles. Le filtre SEIK peut être considéré comme une variante du SEEK pour lequel la linéarisation est remplacée par une interpolation linéaire. Il résout partiellement le problème de la non-linéarité du filtre SEEK. Outre sa plus grande stabilité, il est plus simple à implémenter car il ne nécessite par le calcul du gradient du modèle d'évolution. Néanmoins, son coût est quasiment équivalent à celui du filtre SEEK car l'évolution de leurs bases de correction respectives requiert l'intégration du modèle $r+1$ fois afin de faire évoluer les états interpolés du premier et les vecteurs de la base du deuxième. Ces deux filtres restent donc chers pour des problèmes de grandes tailles comme l'océanographie. En effet, le coût de ces deux filtres est supérieur à $r+1$ fois le coût du modèle d'évolution.

Il est néanmoins possible de réduire ces coûts en simplifiant l'évolution de leur base de correction. Ce qui est évidement le seul moyen permettant de réduire les coûts.

Brasseur *et al.* (1999) ont été les premiers à suivre cette idée en proposant de ne garder que la base de correction initiale du filtre SEEK, calculée par une analyse EOFs, fixe dans le temps. Néanmoins, l'expérience montre que le filtre ainsi obtenu présente des instabilités lors du passage du modèle dans une période instable. En effet, les EOFs formant la base de correction ne captent, le plus souvent, que la variabilité globale de l'état du système. Ainsi, comme ce filtre ne fait pas évoluer sa base de correction pour suivre la dynamique du modèle, sa correction ne sera plus efficace lors des instabilités du modèle. Ce filtre se nomme Filtre de Kalman Singulier à base Fixe (SFEK) Filtre de Kalman Singulier à base fixe et assume donc l'hypothèse que l'opérateur du modèle d'évolution peut être considérer comme l'identité ($M = \mathbf{I}$).

Un autre moyen efficace de réduire les coûts des filtres SEEK et SEIK est d'utiliser la propriété de convergence du filtre de Kalman vers un régime permanent. Ainsi, après un certains temps, la base de correction des filtres SEEK et SEIK reste constante. Il est alors possible de construire un nouveau filtre de Kalman étendu (ou interpolé) singulier à base asymptotique noté SAEK[3] Filtre de Kalman étendu singulier à base asymptotique (ou SAIK[4] Filtre de Kalman interpolé singulier à base asymptotique) qui

[3]Filtre de Kalman étendu singulier à base asymptotique
[4]Filtre de Kalman interpolé singulier à base asymptotique

opère exactement comme le SEEK (ou le SEIK) durant les premiers temps du filtrage et comme le SFEK ensuite. Néanmoins, la propriété de convergence n'est pas valable dans le cas de modèles non-linéaires comme ceux en océanographie. Ce n'est pas non plus une condition suffisante et seule l'expérience et la comparaison des filtres SEEK et SAEK (ou SEIK et SAIK) peuvent permettre de vérifier la valeur de ces filtres. Enfin, en ce qui concerne le coût, ces filtres sont nettement moins coûteux que des filtres SEEK ou SEIK.

Dans le cas de problèmes faiblement non-linéaires, il est imaginable que la matrice de covariance d'erreur d'analyse tende rapidement vers un régime "semi-fixe" dans lequel elle évolue lentement. Dans ce cas, la base de correction va aussi évoluer lentement. Il alors possible de créer une variante du filtre SEEK (ou SEIK) nommé SIEEK[5] Filtre de Kalman étendu singulier à base évolutive par intermittence (ou SIEIK[6]) Filtre de Kalman interpolé singulier à base évolutive par intermittence qui est capable de faire évoluer cette base selon deux régimes après une période d'initialisation avec le filtre SEEK (ou SEIK). Un régime fixe pour lequel la méthode est la même que pour le SFEK et un régime de rattrapage pour lequel la méthode redevient un SEEK (ou SEIK). Ces nouveaux filtres sont, eux aussi, beaucoup moins coûteux.

Il est aussi possible de faire évoluer les vecteurs de la base de correction sur une grille plus grossière que celle du modèle. Cette méthode réduit les coûts de calcul mais aussi les besoins de stockage.

D'une autre manière, il est aussi possible de ne faire évoluer qu'une partie de la base de correction et de garder l'autre partie fixe dans le temps. Ce filtre se nomme SSEEK[7] Filtre de Kalman étendu singulier à base semi-évolutive (ou SSEIK[8]) Filtre de Kalman interpolé singulier à base semi-évolutive dans le cas d'une évolution du SEEK (ou du SEIK). Son coût est, bien sûr, dépendant du nombre de vecteurs de base qui évoluent.

Dans tous ces exemples, lorsque le modèle est stable, les formes dégradées du SEEK et SEIK se comportent aussi bien que le SEEK et SEIK mais peuvent être de deux à dix fois plus rapides. Lorsque le modèle est en période instable, les performances de ces filtres se dégradent notablement.

Enfin, il est aussi possible travailler sur les EOFs en utilisant une base d'EOFs locales ou mixte. La thèse de Ibrahim Hoteit (2001) permet d'avoir une vision précise d'un grand nombre de ces variantes.

5.4 Filtre de Kalman d'ensemble - EnKF

Filtre de Kalman d'ensemble

Le filtre de Kalman d'ensemble a été proposé par Evensen en 1994, puis corrigé en 1998. Pour une description détaillée, il est possible de se référer à Evensen (2003). Cette méthode a d'abord été présentée comme une alternative stochastique au filtre de Kalman étendu qui est déterministe. L'utilisation d'une méthode de Monte Carlo a été imaginée pour résoudre les deux principaux problèmes du filtre de Kalman étendu dans le cadre de système de grande taille non linéaire qui sont son coût très important et sa mauvaise réponse en cas de forte non-linéarité.

[5]Filtre de Kalman étendu singulier à base évolutive par intermittence
[6]Filtre de Kalman interpolé singulier à base évolutive par intermittence
[7]Filtre de Kalman étendu singulier à base semi-évolutive
[8]Filtre de Kalman interpolé singulier à base semi-évolutive

Le filtre de Kalman d'ensemble est très populaire car il est conceptuellement très simple et sa mise en œuvre est aisée. En effet, il ne nécessite ni dérivation des opérateurs tangent-linéaires et des équations adjointes, ni intégration rétrograde du modèle d'évolution.

Le filtre de Kalman d'ensemble reste un filtre Gaussien et n'est pas un filtre particulaire malgré l'emprunt de la nation de particule (comme le filtre SEIK) car il ne gère les statistiques d'erreur que jusqu'à l'ordre deux. Au lieu de propager une matrice de covariance, les erreurs sont représentées statistiquement par un nuage de points propagés par le modèle d'évolution, sans aucune linéarisation. L'étape d'analyse est ensuite celle d'un filtre de Kalman standard.

Comme la montré Burgers *et al.* (1998), il est essentiel de perturber les observations pour chacun des membres de l'ensemble avec l'estimation de la matrice de covariance d'erreur d'observation $\mathbf{R}$. En effet, comme un échantillon statistique a tendance à s'appauvrir par coalescence des points, l'ajout de bruit dans les observations peut être interprété comme l'adjonction d'une partie stochastique permettant d'enrichir l'échantillon.

L'algorithme du filtre de Kalman d'ensemble peut être décrit de la manière suivante (cf. Fig. 5.3). À partir d'un ensemble de conditions initiales, un ensemble d'états d'ébauche à l'instant t_i est construit par de courtes intégrations du modèle d'évolution. La matrice de covariance d'erreur de prévision $\mathbf{P}_i^f$ est calculée à partir de cet échantillon de telle manière que $\mathbf{P}^f = E\left[(\mathbf{x}^f - \overline{\mathbf{x}}^f)(\mathbf{x}^f - \overline{\mathbf{x}}^f)^T\right]$. La matrice de gain $\mathbf{K}$ peut alors être calculée. Chaque ébauche est utilisée pour effectuer une analyse à l'instant t_i comme décrit par l'Eq. 5.9. Les analyses sont obtenues avec des données bruitées. L'ensemble de ces états analysés est ensuite propagé jusqu'à l'instant t_{i+1} et permet alors d'estimer la matrice $\mathbf{P}_{i+1}^f$. Le rang des matrices ainsi estimées est inférieur ou égal à la taille de l'échantillon stochastique, c'est-à-dire très largement inférieur à la taille du vecteur d'état. Cette déficience de rang signifie que l'utilisation directe de la matrice $\mathbf{P}^f$ dans l'algorithme d'assimilation contraint les corrections identifiées par l'analyse à être définies dans l'espace des membres de l'échantillon. Afin de palier à ce problème, le vecteur d'état $\mathbf{x}$ peut être séparé en un vecteur $\mathbf{x}^p$ projeté sur le sous-espace constitué par les échantillons et un vecteur orthogonal. La matrice estimée $\mathbf{P}^f$ est alors liée dans l'algorithme à $\mathbf{x}^p$ et une matrice de covariance d'erreur statique est liée au vecteur orthogonal à $\mathbf{x}^p$. Outre la conséquente économie de calcul et de stockage par rapport au filtre de Kalman étendu, l'algorithme du filtre de Kalman d'ensemble présente l'avantage d'être particulièrement adapté aux machines de calcul parallèle puisque chaque membre de l'échantillon d'analyse peut être calculé indépendamment des autres, et donc simultanément.

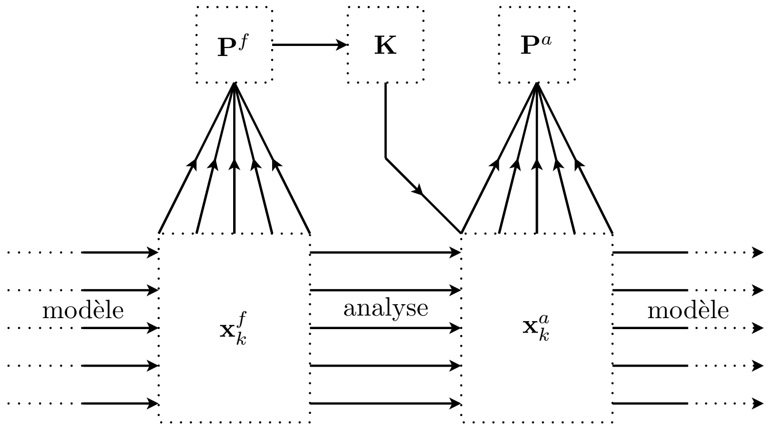


FIG. 5.3 – Représentation schématique des différentes étapes du filtre filtre de Kalman d'ensemble lors d'un cycle d'assimilation du temps t_i au temps t_{i+1}. L'indice k variant de 1 à N représente les différents membres de l'ensemble.

5.5 Méthodes variationnelles

Introduites au début des années cinquante par Sasaki (Sasaki, 1955 et Sasaki, 1958), les méthodes variationnelles sont devenues pendant les années 1990 très populaires. De grands centres de prévisions météorologiques, tels que le NMC[9] U.S. National Meteorological Center maintenant appelé NCEP National Centers for Environmental Prediction en 1991 (Parrish et Derber, 1992), le CEPMMT[10] en 1996 (Courtier *et al.*, 1998 et Anderson *et al.*, 1998) ou Météo-France en 1997, ont adopté ce type de méthode.

Cette approche de l'assimilation de données n'est plus basée sur des théories statistiques, mais sur la théorie de l'optimisation. En opposition aux méthodes séquentielles qui ne traitent les observations qu'au fur et à mesure de leur disponibilité sans jamais utiliser des observations futures, l'approche variationnelle traite le problème globalement sous la forme de la minimisation d'une fonctionnelle (fonction objective) mesurant les caractéristiques indésirables de la solution du modèle. Ces caractéristiques peuvent être l'écart aux observations, la présence d'onde de gravité, le non respect de certains équilibres, ou d'autres. Dans la suite, seul l'écart aux observations et l'éloignement à l'ébauche de la condition initiale seront pris en compte.

Si les statistiques ne sont plus les bases de ces méthodes, elles restent indispensables pour les calibrer et définir la fonction à minimiser.

L'approche variationnelle a déjà été abordée dans la section 4.6 en mettant en évidence, entre autre, l'équivalence à l'optimalité avec le BLUE.

5.5.1 3D-Var

Méthode d'assimilation variationnelle tri-dimensionnelle La méthode d'assimilation variationnelle tri-

[9]U.S. National Meteorological Center, maintenant appelé National Centers for Environmental Prediction
[10]Centre Européen pour les Prévisions Météorologiques à Moyen Terme, aussi appelé ECMWF

dimensionnelle, notée 3D-Var pour « 3Dimensional VARiational assimilation », consiste à chercher l'état le plus vraisemblable à partir des connaissances disponibles sur les lois de probabilités des erreurs d'observation et d'ébauche.

Comme sont nom l'indique clairement, le 3D-Var traite de problèmes tri-dimensionels. Par abus de langage, cette appellation est aussi utilisée pour des problèmes à une ou deux dimensions afin d'éviter les risques de confusions avec l'extension temporelle de cette méthode. En effet, sur un problème bi-dimensionnel, le 3D-Var s'appellerait 2D-Var, tandis que le 4D-Var se nommerait 3D-Var. Ce qui serait particulièrement ambigu. De ce fait, tous les problèmes ne prenant pas en compte l'aspect temporel sont appelés 3D-Var.

Comme pour le filtre de Kalman, le 3D-Var consiste à minimiser la distance au sens des moindres carrés entre l'état estimé et les différentes sources d'informations telles que la prévision précédente et les observations. Le nouvel état analysé est, en général, utilisé comme point de départ de la prévision suivante.

En reprenant les Eqs. 4.18 à 4.24, il est possible d'écrire la fonction coût

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) + \frac{1}{2}(\mathbf{y}^o - H\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{y}^o - H\mathbf{x}), \\ J(\mathbf{x}) &= J^b(\mathbf{x}) + J^o(\mathbf{x}). \end{aligned} \tag{5.18}$$

En général, $\mathbf{x}^b$ est issu de l'intégration par le modèle d'évolution de l'état analysé à l'étape précédente. L'équilibre entre le terme d'écart aux observations J^o et celui de l'ébauche J^b est effectué grâce aux inverses des matrices de covariances d'erreur d'observation et d'ébauche. C'est-à-dire grâce à la confiance portée dans les observations et l'ébauche.

La minimisation se fait à l'aide du gradient égale à

$$\nabla J(\mathbf{x}) = \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b) - \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}^o - H\mathbf{x}). \tag{5.19}$$

Comme montré dans la section 4.6, si l'opérateur d'observation est linéaire ($H = \mathbf{H}$), le 3D-Var est alors équivalent au BLUE à l'optimalité, et donc aussi à l'interpolation optimale (Lorenc, 1986).

3D-Var classique

Dans un cadre théorique, il est possible d'imaginer des observations réparties régulièrement dans le temps. il est alors possible d'effectuer une analyse à chaque pas de temps où celles-ci sont présentes. En pratique, les observations sont diverses et très inégalement réparties. Afin de ne ma multiplier les phases d'analyse, une fenêtre temporelle de taille arbitraire est définie sur laquelle une analyse est effectuée. Dans le 3D-Var classique, toutes les observations sont regroupées (moyennées) à l'instant où est effectuée l'assimilation (Fig. 5.4). Dans ce cas, la méthode ne tient absolument pas compte de l'origine temporelle des observations hormis qu'elles doivent être incluses dans cette fenêtre temporelle. En faisant cette approximation, l'Eq. 5.18 s'écrit alors

$$\begin{aligned} J^b(\mathbf{x}) &= \frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b), \\ J^o(\mathbf{x}) &= \frac{1}{2}(\overline{\mathbf{y}}^o - \overline{H}\mathbf{x})^T\overline{\mathbf{R}}^{-1}(\overline{\mathbf{y}}^o - \overline{H}\mathbf{x}). \end{aligned} \tag{5.20}$$

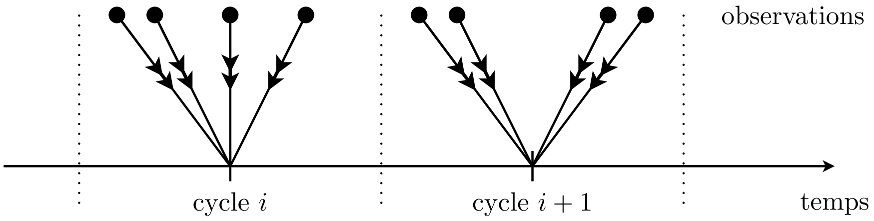


FIG. 5.4 – Répartition des observations dans un 3D-Var classique.

3D-Var FGAT

Le 3D-Var FGAT[11] First Guess at Appropriate Time est un 3D-Var pour lequel les observations sont utilisées en tenant compte de l'instant de leurs mesures (Fig. 5.5). C'est-à-dire que chaque observation est comparée à l'état du modèle au pas de temps le plus proche. Néanmoins, la correction se fait toujours à un instant donné de la fenêtre (typiquement au début ou au milieu). L'équation 5.18 peut alors s'écrire

$$J^b(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}^b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}^b), \tag{5.21}$$

$$J^o(\mathbf{x}) = \frac{1}{2}\sum_{i=0}^{N}(\mathbf{y}_i^o - H\mathbf{x}(t_i))^T R^{-1}(\mathbf{y}_i^o - H\mathbf{x}(t_i)), \tag{5.22}$$

où $\mathbf{y}_i$ représente les N observations de la fenêtre d'assimilation aux instants t_i et $\mathbf{x}(t_i)$ représente l'état du modèle aux instants t_i. L'avantage du 3D-Var FGAT sur le 3D-Var classique est qu'il prend réellement

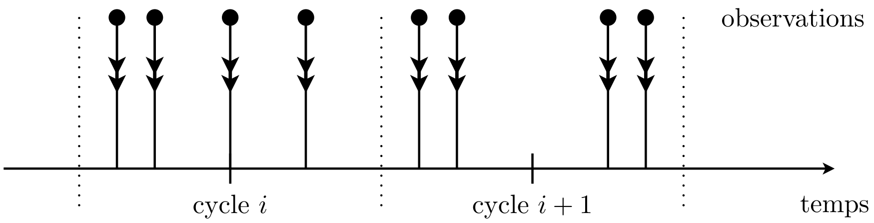


FIG. 5.5 – Répartition des observations dans un 3D-Var FGAT.

en compte l'instant de mesure des observations. De plus, il est très similaire au 4D-Var présenté dans la suite.

3D-Var incrémental

Il est possible d'écrire le 3D-Var de manière incrémentale. L'intérêt de cette formulation sera discutée dans la section 5.8 où il apparaîtra clairement. Cette formulation consiste à considérer comme contrôle l'écart entre l'état du système et l'ébauche, et non plus l'état du système. L'introduction de l'approche incrémentale en météorologie a été motivée par la réduction de coût qu'elle propose en 4D-Var. Soit

$$\delta\mathbf{x} = \mathbf{x}-\mathbf{x}^b. \tag{5.23}$$

[11] First Guess at Appropriate Time

La fonction coût devient alors

$$J(\delta\mathbf{x}) \quad = \quad \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}(\mathbf{d} - \mathbf{H}\delta\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\delta\mathbf{x}), \tag{5.24}$$

où $\mathbf{d}$ est le vecteur d'innovation $\mathbf{y} - H\mathbf{x}^b$.

La formulation 3D-Var FGAT (Eqs. 5.21 et 5.22) prend alors la forme suivante :

$$\begin{aligned} J^b(\delta\mathbf{x}) &= \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x}, \\ J^o(\delta\mathbf{x}) &= \frac{1}{2}\sum_{i=0}^{N}(\mathbf{d}_i - \mathbf{H}\delta\mathbf{x})^T R^{-1}(\mathbf{d}_i - \mathbf{H}\delta\mathbf{x}), \end{aligned} \tag{5.25}$$

avec $\mathbf{d}_i = \mathbf{y}_i^o - H\mathbf{x}(t_i)$.

Une illustration schématique (Fig. 5.6) du 3D-Var FGAT incrémental permet de comprendre simplement son utilisation. Il apparaît ainsi clairement que chaque analyse introduit un « choc » dans le modèle au moment de son introduction. Pour éviter cet inconvénient, il est possible d'utiliser une technique nommée IAU[12] Incremental Analysis Updating et introduite par Bloom *et al.* (1996) qui consiste à répartir l'incrément d'analyse sur toute le cycle d'assimilation comme s'il s'agissait d'un forçage (Fig. 5.7).

5.5.2 4D-Var

Méthode d'assimilation variationnelle quadri-dimensionnelle Le 4D-Var est l'extension temporelle du 3D-Var. Cette méthode ne vise pas à obtenir l'état optimal à un instant donné, mais la trajectoire optimale sur une fenêtre de temps donné. Les observations sont donc prises en compte aussi bien dans leur distribution spatiale que temporelle. Cet aspect est déjà pris en compte par le 3D-Var FGAT présenté en section 5.5.1. Néanmoins, le 4D-Var apporte un aspect temporel en plus car il propage l'information apportée par les observations à l'instant initial de la fenêtre d'assimilation. De ce fait, l'analyse obtenue doit permettre au modèle d'évolution d'avoir la trajectoire la plus proche possible de l'ensemble des observations utilisées.

Cette amélioration du 3D-Var permet d'ajouter la connaissance de l'évolution du système comme information pour l'analyse.

De nombreuses applications à des modèles réalistes météorologiques (Thépaut et Courtier, 1991 et Zupanski, 1993) et océanographiques (Moore, 1986 ; Shröter *et al.*, 1993 ; Luong *et al.*, 1998 et Greiner *et al.*, 1998) ont depuis longtemps été effectuées.

L'amélioration ainsi apportée, conjuguée au fort développement des moyens de calculs, font que le 4D-Var est venu remplacer le 3D-Var dans les systèmes de prévision opérationnels atmosphériques du CEPMMT en 1997 et de Météo-France en 2000.

4D-Var classique

Soit $M_{0\rightarrow i}$ l'opérateur *a priori* non-linéaire qui permet de propager l'état du système $\mathbf{x}$ de t_0 à t_i :

$$\forall i, \qquad \mathbf{x}(t_i) \quad = \quad M_{0\rightarrow i}(\mathbf{x}). \tag{5.26}$$

[12]Incremental Analysis Updating

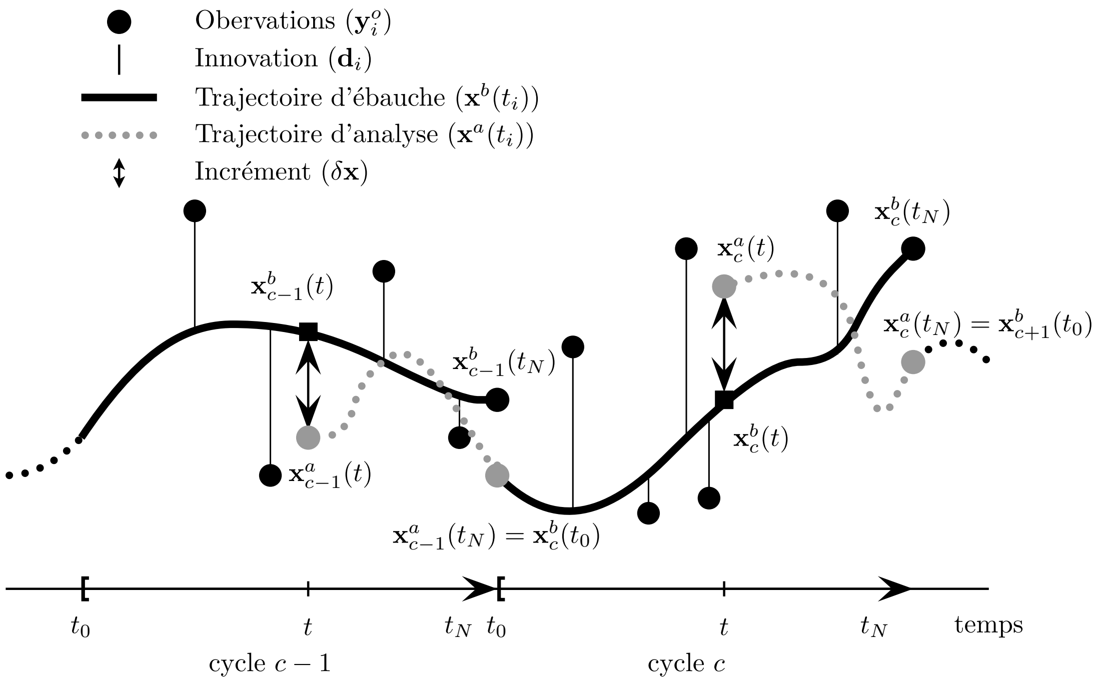


FIG. 5.6 – Illustration de la procédure pour cycler le 3D-Var FGAT incrémental. Pour chaque cycle c, le modèle d'évolution est intégré de t_0 à t_N à partir de l'état initial d'ébauche $\mathbf{x}^b_c(t_0)$ (courbe noir pleine) et le vecteur d'innovation $\mathbf{d}_i$ est calculé pour les différentes observations $\mathbf{y}^o_i$ avec $i = 1, \cdots, N$ (ligne fine verticale). L'analyse est effectuée à l'instant t compris entre t_0 et t_N. Après l'analyse, un incrément est obtenu et est rajouté à l'état d'ébauche $\mathbf{x}^b_c(t)$ permettant d'obtenir l'état analysé $\mathbf{x}^a_c(t)$. L'état analysé est ensuite propagé jusqu'à le fin du cycle c (courbe grise pointillée). Cet état analysé $\mathbf{x}^a_c(t_N)$ est ensuite utilisé comme état initial d'ébauche pour le cycle suivant.

En supposant, dans un premier temps, le modèle parfait, la fonction coût J du 4D-Var se décompose, comme pour le 3D-Var, en un terme J^b lié à l'ébauche et un autre J^o lié aux observations.

En tenant compte des instants de mesures t_i, la matrice de covariance d'erreur des observation est notée $\mathbf{R}_i$ et l'opérateur d'observation non-linéaire H_i. Les observations $\mathbf{y}^o_i$ sont donc comparées à leur équivalent modèle $H_i\mathbf{x}(t_i)$ à chaque instant d'observation. Le calcul du terme J^o nécessite l'intégration du modèle d'évolution de t_0 à t_N. Le vecteur d'état $\mathbf{x}$ est ainsi propagé par le modèle numérique $M_{0\to N}$ de t_0 à t_N où N représentent le nombre de pas de temps de l'intégration du modèle à l'intérieur d'un cycle d'assimilation.

L'algorithme d'assimilation identifie un état $\mathbf{x}^a$ de la variable $\mathbf{x}$ à l'instant t_0 (une condition initiale), qui, intégré par le modèle d'évolution fournit une trajectoire optimale au sens des moindres carrés (la trajectoire analysée) sur l'ensemble de la fenêtre d'assimilation[13] :

$$J^b(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}-\mathbf{x}^b\right)^T \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right), \tag{5.27}$$

[13]Les vecteurs $\mathbf{x}^a(t_0)$ et $\mathbf{x}^b(t_0)$ sont notés $\mathbf{x}^a$ et $\mathbf{x}^b$. À tout autre moment que l'instant initial t_0, les notations $\mathbf{x}^a(t_i)$ et $\mathbf{x}^b(t_i)$ seront utilisées.

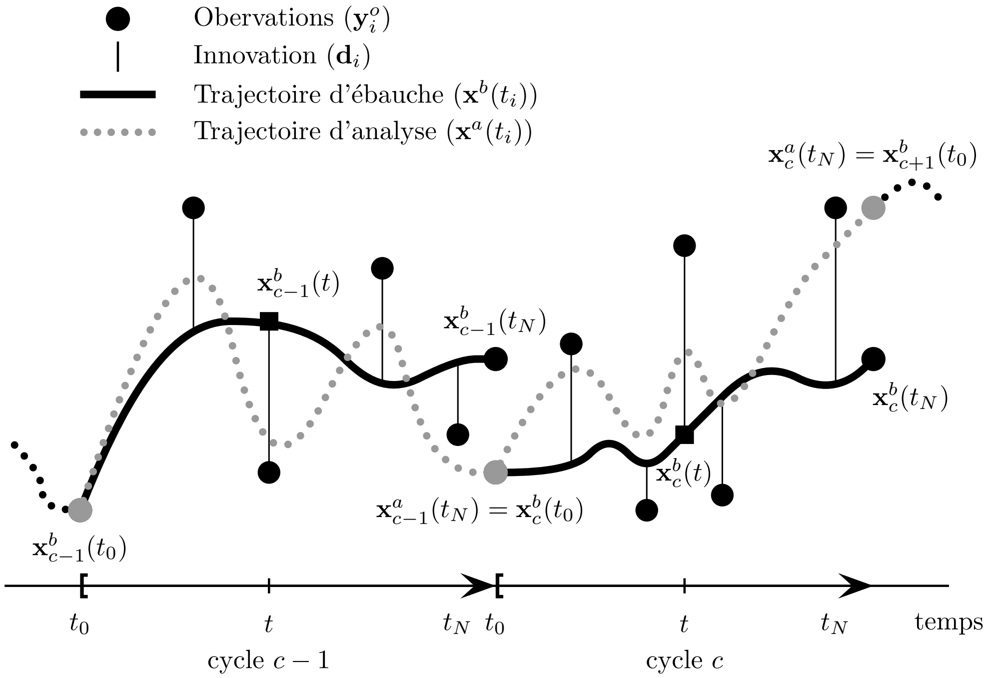


FIG. 5.7 – Illustration de la procédure pour cycler le 3D-Var FGAT incrémental avec IAU. Pour chaque cycle c, le modèle d'évolution est intégré de t_0 à t_N à partir de l'état initial d'ébauche $\mathbf{x}_c^b(t_0)$ (courbe noir pleine) et le vecteur d'innovation $\mathbf{d}_i$ est calculé pour les différentes observations $\mathbf{y}_i^o$ avec $i = 1, \cdots, N$ (ligne fine verticale). L'analyse est effectuée à l'instant t compris entre t_0 et t_N. Après l'analyse, un incrément est obtenu qui est rajouté comme un forçage lors de l'intégration du modèle d'évolution de t_0 à t_N à partir de l'état initial d'ébauche $\mathbf{x}_c^b(t_0)$ (courbe grise pointillée) Cet état analysé $\mathbf{x}_c^a(t_N)$ est ensuite utilisé comme état initial d'ébauche pour le cycle suivant.

$$
\begin{aligned}
J^o(\mathbf{x}) &= \frac{1}{2}\sum_{i=0}^{N}\left(\mathbf{y}_i^o - H_i\mathbf{x}(t_i)\right)^T \mathbf{R}_i^{-1}\left(\mathbf{y}_i^o - H_i\mathbf{x}(t_i)\right), \\
&= \frac{1}{2}\sum_{i=0}^{N}\left(\mathbf{y}_i^o - H_iM_{0\to i}(\mathbf{x})\right)^T \mathbf{R}_i^{-1}\left(\mathbf{y}_i^o - H_iM_{0\to i}(\mathbf{x})\right), \\
&= \frac{1}{2}\sum_{i=0}^{N}\left(\mathbf{y}_i^o - G_i\mathbf{x}\right)^T \mathbf{R}_i^{-1}\left(\mathbf{y}_i^o - G_i\mathbf{x}\right), \qquad (5.28)
\end{aligned}
$$

où $G_i\mathbf{x} = H_i\mathbf{x}(t_i) = H_iM_{0\to i}(\mathbf{x})$. L'opérateur G_i est appelé l'opérateur d'observation généralisé pour l'état $\mathbf{x}$ propagé par le modèle de t_0 à t_i.

L'état optimal $\mathbf{x}^a$, qui minimise la fonction coût J, est obtenu quand le gradient de cette fonctionnelle est nul. Comme pour le 3D-Var, ce gradient s'obtient simplement :

$$
\nabla J(\mathbf{x}) = \mathbf{B}^{-1}\left(\mathbf{x} - \mathbf{x}^b\right) - \sum_{i=0}^{N}\mathbf{M}_{0\to i}^T\mathbf{H}_i^T\mathbf{R}_i^{-1}\left(\mathbf{y}_i^o - G_i\mathbf{x}\right). \qquad (5.29)
$$

Les opérateurs $\mathbf{H}_i$, $\mathbf{M}_{0\to i}$ et $\mathbf{G}_i = \mathbf{H}_i\mathbf{M}_{0\to i}$ sont les opérateurs linéarisés de H_i, $M_{0\to i}$ et G_i au voisinage de l'ébauche. L'opérateur $\mathbf{M}_{0\rightarrow i}^T$ est l'adjoint de l'opérateur linéarisé $\mathbf{M}_{0\rightarrow i}$.

Équivalence avec le filtre de Kalman

Si les opérateurs H et M sont linéaires, alors la fonction coût J est quadratique. Si de plus le modèle est parfait (hypothèse émise dans la section précédente), alors la solution du 4D-Var à la fin de la fenêtre d'assimilation est identique à celle du filtre de Kalman (Jazwinski, 1970 ; Ghil *et al.*, 1981 et Lorenc, 1986). En météorologie comme en océanographie, H et M sont souvent faiblement non-linéaires. Dans ce cas, la minimisation peut être effectuée avec un algorithme adapté aux fonctions coûts non-quadratiques. Généralement, l'opérateur généralisé d'observation linéarisé $\mathbf{H}_i$ et le modèle linéaire-tangent $\mathbf{M}_{0\rightarrow i}$ sont supposés de bonnes approximations de H_i et $M_{0\rightarrow i}$ sur la fenêtre temporelle d'assimilation. La validité du linéaire-tangent dépend d'une part de la formulation du modèle numérique et de l'opérateur d'observation considérés mais aussi du contexte de l'assimilation, notamment la durée de la fenêtre d'assimilation, de la physique des phénomènes représentés et de la région d'étude.

Calcul de la fonction coût et de son gradient

Le terme d'ébauche de la fonction coût J^b est identique à celui décrit pour le 3D-Var. Son évaluation est directe pour tout état de la variable d'état $\mathbf{x}$. L'évaluation du terme lié aux observations J^o est, par contre, plus ardue. Dans le cadre du 3D-Var, J^o est une combinaison linéaire des écarts entre les observations et l'état du modèle à un instant donné. À présent, chaque évaluation de J^o requiert l'intégration du modèle sur la fenêtre d'assimilation.

L'équation $\nabla J = 0$ ne peut être résolue directement. Une solution minimisant la fonction coût J par une méthode de descente itérative utilisant la valeur de ∇J est envisageable. Généralement, l'état d'ébauche $\mathbf{x}^b$ est utilisé comme une première estimation de l'état analysé. Le gradient de la fonction coût ∇J peut être estimé de manière très efficace par la méthode adjointe[14]. En fait, l'évaluation du terme ∇J^o peut se résumer à une intégration du modèle direct et une intégration du modèle adjoint (Le Dimet et Talagrand, 1986). Le gradient de la fonction coût est évalué par rapport à $\mathbf{x}$ décrivant le vecteur d'état à l'instant initial de la fenêtre d'assimilation

$$\nabla J^o(\mathbf{x}) \quad = \quad \mathbf{x}^*, \tag{5.30}$$

où $\mathbf{x}^*$ représente l'adjoint de la variable $\mathbf{x}$.

La méthode adjointe permet une évaluation efficace du gradient de la fonction coût dans l'algorithme du 4D-Var. Néanmoins, cette opération implique qu'une version adjointe du modèle d'évolution linéarisé soit disponible. Cette contrainte est souvent l'étape cruciale à surmonter lors de l'implémentation de l'algorithme d'assimilation 4D-Var. En effet, dans de nombreux domaines, comme la météorologie et l'océanographie, les modèles d'évolution sont souvent complexes et non-linéaires. Deux étapes sont généralement nécessaire : écrire le modèle linéaire-tangent du modèle d'évolution ; puis écrire son adjoint. Ces opérations peuvent être effectuées soit manuellement, soit par des méthodes de linéarisation et d'adjoint automatiques (Giering et Kaminski, 1998). Malheureusement, ces méthodes automatiques ne sont pas

[14]Une description complète des méthodes adjointes est présentée dans Thacket et Long (1988

capables de présumer les hypothèses de linéarisation qui sont faites manuellement et les modèles linéaire-tangent et adjoint obtenus sont souvent très coûteux. Néanmoins, l'utilisation conjointe d'une méthode automatique et d'une écriture manuelle permet d'obtenir assez rapidement des modèles tangent-linéaire et adjoint efficaces.

Coût de calcul du 4D-Var

L'intégration du modèle direct pour le calcul des innovations, puis pour la propagation de l'analyse, ainsi que l'intégration du modèle adjoint font du 4D-Var une méthode très coûteuse. Elle est néanmoins applicable en météorologie et, dans une moindre mesure, en océanographie. En effet, la plupart des grands centres opérationnels de météorologie utilisent des méthodes 4D-Var. Néanmoins, à la différence des centres océanographiques, leur capacité de calcul est très importante. Ainsi, le 4D-Var a été utilisé dans de nombreuses études en océanographie, notamment pour assimiler des données altimétriques dans des modèles régionaux quasi-géostrophiques (Moore, 1991), dans des modèles à gravité réduite (Weaver et Anderson, 1997), dans des modèles « shallow-water » (Greiner et Périgaud, 1994), dans des modèles aux équations primitives (Stammer *et al.*, 1997 ; Greiner *et al.*, 1998a, b) et dans des modèles couplés (Lee *et al.*, 2000).

Le 4D-Var à contrainte faible

Dans la présentation du 4D-Var, le modèle a été considéré comme parfait. C'est-à-dire que le modèle décrivait exactement le comportement du système. Dans ce cas, le 4D-Var est décrit comme à contrainte forte (Sasaki, 1970). Cependant, malgré l'utilisation de modèles extrêmement sophistiqués, ceux-ci comportent des erreurs qui ne peuvent être négligées et qui ne pourront jamais l'être pour des systèmes aussi complexes que l'atmosphère ou l'océanographie.

Comme pour le filtre de Kalman présenté dans la section 5.2, l'ajout d'un terme correctif directement dans la fonction coût est possible. En définissant l'erreur du modèle telle que

$$\mathbf{q}_i = \mathbf{x}(t_i) - M_{i\to i+1}\mathbf{x}(t_{i-1}), \tag{5.31}$$

la fonction coût J peut alors s'écrire

$$\begin{aligned} J(\mathbf{x}, \mathbf{q}_1, \cdots, \mathbf{q}_N) &= \frac{1}{2}\left(\mathbf{x}-\mathbf{x}^b\right)^T \mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}^b\right) \\ &+ \frac{1}{2}\sum_{i=0}^{N}\left(\mathbf{y}_i^o - G_i\mathbf{x}\right)^T \mathbf{R}_i^{-1}\left(\mathbf{y}_i^o - G_i\mathbf{x}\right) \\ &+ \frac{1}{2}\sum_{i=0}^{N}\mathbf{q}_i^T\mathbf{Q}_i^{-1}\mathbf{q}_i, \end{aligned} \tag{5.32}$$

où $\mathbf{Q}_i$ est la matrice de covariance d'erreur modèle à l'instant t_i.

Cette formulation du 4D-Var est dite à contrainte faible et il est nécessaire de proposer une modélisation de la matrice de covariance d'erreur modèle $\mathbf{Q}$. De nombreux travaux par Derber (1989), Stammer *et al.* (1997), Bennet *et al.* (1998), Lee et Marotzke (1998) ou Vidard (2001) utilisent cette algorithme du 4D-Var à contrainte faible.

Évolution temporelle de la matrice de covariance d'erreur d'ébauche

L'impact de la prise en compte du caractère temporel est clairement mis en évidence dans la fonction coût liée aux observations J^o. Cependant, l'aspect temporel de la matrice de covariance d'erreur d'ébauche $\mathbf{B}$ est moins évident dans un premier temps. En effet, la dynamique du modèle d'évolution est prise en compte implicitement par la matrice de covariance d'erreur d'ébauche $\mathbf{B}$ sur chaque fenêtre d'assimilation, et donc dans la fonction coût liée à l'ébauche J^b.

L'erreur d'ébauche à l'instant t_0 est l'écart entre l'ébauche et l'état vrai au début de la fenêtre d'assimilation $\boldsymbol{\epsilon}^b = \mathbf{x}^b - \mathbf{x}^t$. À tout instant t_i du cycle d'assimilation, cette erreur notée $\boldsymbol{\epsilon}^b(t_i)$, est l'écart entre l'état vrai $\mathbf{x}^t(t_i)$ et l'ébauche $\mathbf{x}^b(t_i)$ sous l'hypothèse que le modèle est linéaire :

$$\begin{aligned} \boldsymbol{\epsilon}^{t_i} &= \mathbf{x}^b(t_i) - \mathbf{x}^t(t_i) \\ &= M_{0\to i}\mathbf{x}^b - M_{0\to i}\mathbf{x}^t \\ &\approx \mathbf{M}_{0\to i}\left(\mathbf{x}^b - \mathbf{x}^t\right) \\ &\approx \mathbf{M}_{0\to i}\boldsymbol{\epsilon}^b. \end{aligned} \tag{5.33}$$

La matrice de covariance d'erreur à l'instant t_i est alors donnée par

$$\begin{aligned} E\left[\boldsymbol{\epsilon}^b(t_i)(\boldsymbol{\epsilon}^b(t_i))^T\right] &= E\left[\mathbf{M}_{0\to i}\boldsymbol{\epsilon}^b(\boldsymbol{\epsilon}^b)^T\mathbf{M}_{0\to i}^T\right] \\ &= \mathbf{M}_{0\to i}\mathbf{B}\mathbf{M}_{0\to i}^T. \end{aligned} \tag{5.34}$$

Cette matrice décrit les erreurs liées à l'ébauche aux instants d'observations. La matrice de covariance d'erreur d'ébauche $\mathbf{B}$ est donc implicitement propagée en temps par le 4D-Var à travers la dynamique du modèle linéaire-tangent $\mathbf{M}_{0\to i}$ et son adjoint $\mathbf{M}_{0\to i}^T$. Le 4D-Var présente donc une analyse cohérente avec la dynamique du système (Thépaut *et al.*, 1993).

Équivalence avec le 3D-Var FGAT

Il existe une similarité frappante entre le 3D-Var FGAT (Eqs. 5.21 et 5.22) et le 4D-Var à contrainte forte (Eqs. 5.27 et 5.28). En effet, il suffit de définir le modèle d'évolution dans le 4D-Var comme l'identité ($M = \mathbf{I}$), pour que le l'opérateur généralisé d'observation G se réduise à l'opérateur d'observation non-linéaire H et que le 4D-Var à contrainte forte devienne un 3D-Var FGAT.

C'est une caractéristique très intéressante car elle permet très facilement, du point de vue de l'implémentation informatique, de passer d'une méthode à l'autre. Étant donné le coût informatique du 4D-Var, cette approche permet une évolution simple et naturelle du 3D-Var FGAT vers le 4D-Var en fonction de l'évolution des capacités informatiques disponibles.

Formulation incrémentale du 4D-Var

La formulation incrémentale a déjà été abordée rapidement à propos du 3D-Var dans la section 5.5.1, mais elle prend avec le 4D-Var tout son sens.

En effet, l'introduction de l'approche incrémentale en météorologie a été motivée par la réduction de coût qu'elle propose. Dans le cadre du 4D-Var classique décrit précédemment, à chaque itération de la minimisation de la fonction coût (Eqs. 5.27 et 5.28), l'intégration du modèle direct non-linéaire et de l'adjoint du modèle linéarisé est très coûteuse. Les non-linéarités des modèles numériques atmosphériques peuvent conduire à des fonctions coûts complexes. Les minimiseurs utilisés sur ces fonctions n'aboutissent pas forcément à une minimisation fiable et ce, de surcroît, à un coût élevé. Ces non-linéarités compliquent de plus, lourdement la tâche de l'écriture de l'adjoint (Xu, 1996). Dans l'approche incrémentale, la fonction coût est rendue quadratique, ce qui garantit l'identification d'un minimum unique par une méthode de descente pour un coût de calcul notablement inférieur à celui du problème non-linéaire. Le modèle linéaire-tangent est écrit avec une physique simplifiée, ce qui facilite grandement l'écriture de l'adjoint du modèle. De plus, une approche communément choisie en météorologie est d'utiliser un modèle linéaire-tangent à une résolution plus basse que celle du modèle non-linéaire. Le coût de la minimisation du 4D-Var en est significativement réduit. C'est formulation incrémentale de l'approche variationnelle du 4D-Var qui a permis de le rendre applicable de façon opérationnelle pour la prévision météorologique (Courtier *et al.*, 1994 et Rabier *et al.*, 2000).

Dans la formulation incrémentale de l'assimilation variationnelle, l'objectif est de minimiser la fonction coût, non plus par rapport à la variable d'état $\mathbf{x}$, mais par rapport à un incrément $\delta\mathbf{x}$ tel que $\mathbf{x} = \mathbf{x}^b + \delta\mathbf{x}$. L'hypothèse principale de la formulation incrémentale est d'utiliser un modèle d'évolution et un opérateur d'observation linéarisé pour propager l'incrément mais de conserver le modèle non-linéaire pour la propagation de l'ébauche $\mathbf{x}^b$. La solution de la minimisation est l'incrément d'analyse $\delta\mathbf{x}^a$ à t_0 tel que le vecteur d'analyse $\mathbf{x}^a$ soit

$$\mathbf{x}^a = \mathbf{x}^b + \delta\mathbf{x}^a. \tag{5.35}$$

En supposant que l'ébauche $\mathbf{x}^b$ est une « bonne » approximation *a priori* de l'état optimal du système au sens des moindres carrés, l'incrément $\delta\mathbf{x}$ devrait être petit. Les opérateurs non-linéaires d'observation H et du modèle M sont linéarisés au voisinage de l'ébauche de sorte que pour tout état du modèle $\mathbf{x}$, à chaque instant t_i de la fenêtre temporelle d'assimilation,

$$\begin{aligned} \mathbf{x}(t_{i+1}) &= M_{0\to i}\mathbf{x} \\ &= M_{0\to i}(\mathbf{x}^b + \delta\mathbf{x}) \\ &\approx M_{0\to i}\mathbf{x}^b + \mathbf{M}_{0\to i}\delta\mathbf{x}. \end{aligned} \tag{5.36}$$

Par conséquent

$$H_i\mathbf{x}(t_i) = H_i\mathbf{x}^b(t_i) + \mathbf{H}_i\delta\mathbf{x} + \parallel \delta\mathbf{x} \parallel^2 + \cdots, \tag{5.37}$$

$$G_i\mathbf{x}(t_i) = G_i\mathbf{x}^b(t_i) + \mathbf{G}_i\delta\mathbf{x} + \parallel \delta\mathbf{x} \parallel^2 + \cdots, \tag{5.38}$$

où $\mathbf{M}_{0\to i}$, $\mathbf{H}_i$ et $\mathbf{G}_i$ sont les opérateurs linéarisés à t_0 autour de l'ébauche x^b de $M_{0\to i}$, H_i et G_i.

$$\begin{aligned} \mathbf{M}_{0\to i} &= \left.\frac{\partial M_{0\to i}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^b}, \\ \mathbf{H}_i &= \left.\frac{\partial H_i}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^b}, \\ \mathbf{G}_i &= \left.\frac{\partial G_i}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^b}. \end{aligned} \tag{5.39}$$

En insérant les Eqs. 5.36, 5.37 et 5.38 dans la formulation de la fonction coût donnée par les Eqs. 5.27 et 5.28, elle peut être reformulée de manière incrémentale

$$J^b(\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^T\mathbf{B}^{-1}\delta\mathbf{x}, \tag{5.40}$$

$$J^o(\mathbf{x}) = \frac{1}{2}\sum_{i=0}^{N}\left(\mathbf{d}_i - \mathbf{G}_i\delta\mathbf{x}\right)^T\mathbf{R}_i^{-1}\left(\mathbf{d}_i - \mathbf{G}_i\delta\mathbf{x}\right), \tag{5.41}$$

où $\mathbf{d}_i = \mathbf{y}_i^o - G_i\mathbf{x}^b = \mathbf{y}_i^o - H_i\mathbf{x}^b(t_i)$ représente l'innovation au temps t_i, c'est-à-dire l'écart entre les observations et l'équivalent de l'ébauche donné par l'opérateur d'observation généralisé G_i dans l'espace des observations à chaque temps t_i.

La fonction coût $J = J^b + J^o$ du 4D-Var incrémental à contrainte forte est quadratique et la minimisation possède une solution unique. Si la linéarisation des opérateurs H_i, $M_{0\to i}$ et G_i est exacte, alors la solution est identique à celle obtenue par le filtre de Kalman étendu.

L'incrément d'analyse qui minimise la fonction coût donné par les Eqs. 5.40 et 5.41 est

$$\delta\mathbf{x}^a = \mathbf{B}\mathbf{G}^T\left(\mathbf{G}\mathbf{B}\mathbf{G}^T + \mathbf{R}\right)^{-1}\mathbf{d}. \tag{5.42}$$

La fonction coût incrémentale (Eqs. 5.40 et 5.41) est minimisée par une méthode itérative de descente. Cette minimisation nécessite le calcul de la fonction coût et de son gradient à chaque itération de la minimisation comme pour l'algorithme classique du 4D-Var. Avant le début de la minimisation, l'état d'ébauche à t_0 noté $\mathbf{x}^b$ est propagé par le modèle non-linéaire permettant le calcul des innovations $\mathbf{d}_i$ à chaque instant d'observation t_i.

La figure 5.8 permet de représenter simplement l'utilisation du 4D-Var incrémental. Elle est à comparer aux Figs. 5.6 et 5.7 qui représentent, selon les mêmes codes, les 3D-Var FGAT incrémental avec ou sans IAU.

Prise en compte des non-linéarités

À chaque itération de la minimisation, le terme de la fonction coût lié aux observations $J^o(\delta\mathbf{x})$ est calculé en propageant l'incrément $\delta\mathbf{x}$ dans le temps avec le modèle linéaire-tangent $\mathbf{M}$. Le calcul du gradient de la fonction coût, notamment de la partie relative aux observations $\nabla J^o(\delta\mathbf{x})$, nécessite l'intégration de l'adjoint du modèle linéaire-tangent $\mathbf{M}^T$ sur la fenêtre d'assimilation. À la fin de la minimisation, l'incrément d'analyse est ajouté à l'ébauche $\mathbf{x}^b$ (Eq. 5.35). L'état analysé à l'instant initial de la fenêtre d'assimilation $\mathbf{x}^a$ est ensuite propagé par le modèle non-linéaire M jusqu'à la fin de la fenêtre permettant d'obtenir une trajectoire analysée $\mathbf{x}^a(t_i)$. En pratique, il est possible de prendre en compte les faibles non-linéarités des opérateurs H et M en mettant à jour la trajectoire de référence au cours de la minimisation. Ces mises à jours sont aussi appelée boucles externes. Le modèle linéaire est relinéarisé au voisinage du nouvel état de référence à chaque boucle externe et la fonction coût est ensuite minimisée par une série de boucles internes. Cette méthode permet de conserver la fonction coût quadratique tout en tenant compte, jusqu'à un certain point, des non-linéarités du système.

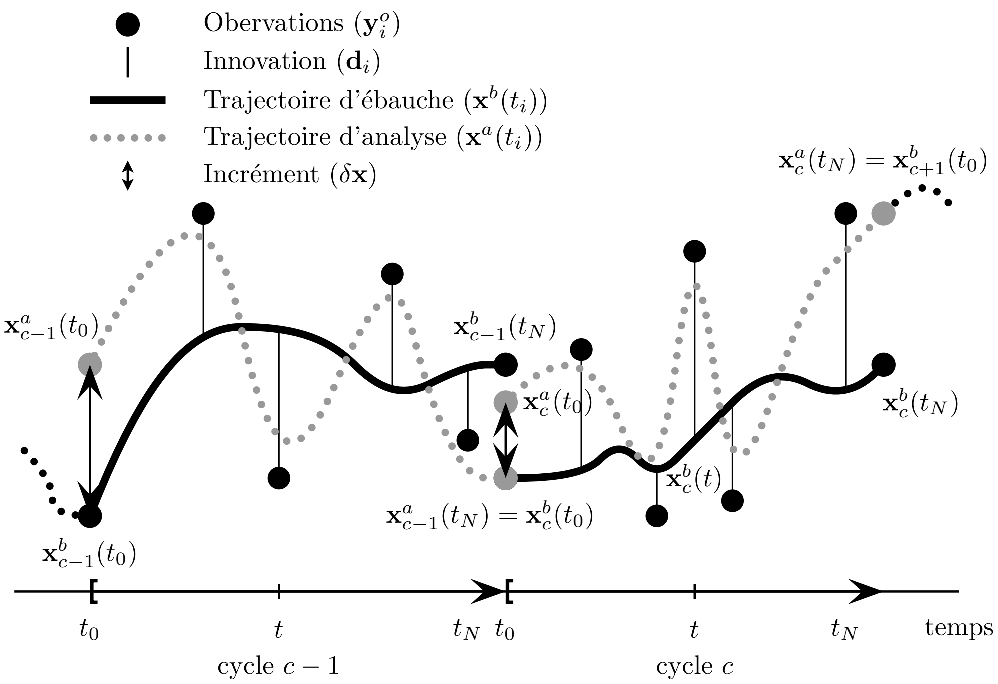


FIG. 5.8 – Illustration de la procédure pour cycler le 4D-Var incrémental. Pour chaque cycle c, le modèle d'évolution est intégré de t_0 à t_N à partir de l'état initial d'ébauche $\mathbf{x}_c^b(t_0)$ (courbe noir pleine) et le vecteur d'innovation $\mathbf{d}_i$ est calculé pour les différentes observations $\mathbf{y}_i^o$ avec $i = 1, \cdots, N$ (ligne fine verticale). L'analyse est effectuée à l'instant t_0 en ramenant les innovations à l'instant t_0 à l'aide du modèle adjoint. Après la minimisation, un incrément est obtenu qui est rajouté à l'état de l'ébauche initial pour obtenir l'état analysé $\mathbf{x}_c^a(t_0) = \mathbf{x}_c^b(t_0) + \delta\mathbf{x}^a$. Cet état analysé tient compte de la dynamique du modèle de sorte que la trajectoire analysée (courbe grise pointillée) minimise au mieux l'écart aux observations tout au long du cycle d'assimilation (de t_0 à t_N). L'état analysé $\mathbf{x}_c^a(t_N)$ est ensuite utilisé comme état initial d'ébauche pour le cycle suivant.

Chapitre 6

Modèle de covariance d'erreur

la connaissance est une navigation dans un océan d'incertitudes à travers des archipels de certitudes.

Edgar Morin in Les sept savoirs nécessaires à l'éducation du futur

Une spécification correcte et adaptée des covariances d'erreur d'observation et d'ébauche est primordiale pour la qualité de l'analyse. En effet, ces covariances déterminent comment et à quel point les observations corrigeront l'état d'ébauche. Les variances d'erreur sont les paramètres essentiels. Pour autant, les corrélations sont aussi très importantes car elles déterminent comment les informations apportées par les observations seront lissées dans l'espace du modèle s'il y a un décalage entre la résolution du modèle et la densité d'observations. Dans le cadre des filtres de Kalman ou du 4D-Var à contrainte faible, les covariances d'erreur modèle sont aussi à spécifier.

6.1 Description des erreurs

6.1.1 Variances d'erreur d'observation

Les variances d'erreur d'observation sont le plus souvent estimées à l'aide des connaissances sur les caractéristiques techniques des instruments de mesures. Ces caractéristiques peuvent être, par exemple, déterminées par des observations positionnées au même endroit. Ces variances doivent aussi inclure les variances d'erreur de représentativité qui sont loin d'être négligeables tant qu'il existe des phénomènes physiques qui ne sont pas bien représentés dans l'espace du modèle.

D'autre part, il ne faut absolument pas considérer qu'un biais puisse être considéré comme une contribution aux variances d'erreur d'observation. En effet, il occasionnerait un biais dans l'incrément d'analyse. Ainsi, à chaque fois qu'un biais est mis en évidence, il doit être retiré des observations ou de l'état d'ébauche en fonction de son origine supposée. Il est cependant souvent difficile de déterminer son origine.

6.1.2 Covariances d'erreur d'observation

Les covariances d'erreur d'observation sont le plus souvent considérées comme nulles. En d'autres termes, des mesures distinctes sont affectées par des erreurs physiques indépendantes. Cette hypothèse paraît raisonnable pour des observations mesurées par des instruments différents. Elle paraît moins évidente quand un jeu d'observations est obtenu par le même instrument de mesure (mesures satellite, bouées dérivant es...) ou quand une série temporelle de mesures d'une même station, par exemple par une bouée fixe, est utilisée dans un 4D-Var. Dans de tels cas, il apparaît intuitivement que des corrélations d'erreur doivent exister entre ces mesures proches les une des autres.

La présence d'un biais, par exemple, se traduit par une corrélation d'erreur permanente. De plus, le prétraitement des observations peut produire artificiellement des corrélations d'erreur entre les observations transformées. Ces transformations sont assez courantes et permettent, par exemple, de transformer des températures en températures potentielles, la distance altimétrique d'un satellite en anomalie de hauteur de mer ou des radiances obtenues par des satellites en température de surface de la mer.

Quand l'état d'ébauche est utilisé dans le prétraitement des observations, cela peut créer artificiellement des corrélations entre l'erreur d'observation et l'erreur d'ébauche qui sont très difficile à prendre en compte. Par exemple, rapprocher une observation de l'état d'ébauche donne l'impression de réduire les erreurs d'observation et d'ébauche, mais cela réduit de manière irréaliste le poids de l'information apportée par l'observation originale.

Enfin, les erreurs de représentativité sont corrélées par nature. Les d'erreurs d'interpolation sont toujours corrélées quelque soit la densité des observations vis-à-vis de la résolution du modèle. D'autre part, les erreurs dans la définition de l'opérateur d'observation, comme le modèle d'évolution pour le 4D-Var, sont corrélées aux mêmes échelles que le problème modélisé.

La présence de corrélations d'erreur d'observation positives réduit le poids moyen des observations. Ainsi, elle augmente l'importance relative des différences entre les observations, comme les gradients ou les tendances. Cependant, les corrélations d'erreur d'observation sont très difficiles à estimer et peuvent causer des problèmes numériques lors de l'analyse ou dans les algorithmes de contrôle qualité. En pratique, il est souvent plus facile de minimiser leurs impacts en utilisant des méthodes de corrections de biais, en évitant trop de prétraitement des observations, en refusant des observations dans les régions très denses et en améliorant le modèle d'évolution et les opérateurs d'observation. Ainsi, les modèles de covariances d'erreur d'observation sont, en général, diagonaux ou presque.

6.1.3 Variances d'erreur d'ébauche

En général, l'ébauche provient d'une prévision obtenue avec le modèle d'évolution. Les variances d'erreur d'ébauche sont donc les variances d'erreur de la prévision utilisée pour obtenir l'état d'ébauche initial $\mathbf{x}^b$. Avec les filtres de Kalman, ces variances d'erreur sont estimées automatiquement à travers le modèle tangent-linéaire. Il n'est donc pas nécessaire de les spécifier. Cependant, le problème n'est que déplacé puisqu'il faut alors spécifier l'erreur modèle $\mathbf{Q}$ et, pour les filtres réduits, mettre au point les approximations nécessaires à ces algorithmes.

Une première estimation des variances d'erreur d'ébauche peut être obtenue en prenant une fraction des variances climatologiques des champs des variables du vecteur d'état.

Une autre possibilité est d'utiliser des quantités dont les statistiques d'erreur sont équivalentes à celle de l'erreur d'ébauche. Parmi ces méthodes, les plus connues sont la méthode NMC et la méthode d'ensemble décrites par la suite. Une des hypothèses de ces méthodes est que l'analyse soit de bonne qualité. C'est-à-dire, en d'autres termes, que les observations soient nombreuses.

Enfin, la méthode la meilleure est, sans conteste, celle qui utilise l'innovation pour estimer les variances d'erreurs. Cette méthode sera aussi décrite par la suite mais elle repose sur l'hypothèse que les erreurs d'observation ne sont pas corrélées.

Néanmoins, dans la plupart des problèmes, l'erreur d'ébauche est supposée dépendre largement de l'état lui-même de l'ébauche. Il est alors très bénéfique que l'erreur d'ébauche dépende de l'écoulement et prenne en compte les variations temporelles. Cette caractéristique est obtenue avec les filtres de Kalman, dans les fenêtres temporelles des 4D-Var, par l'utilisation de lois empiriques basées les connaissances physiques du système ou par des méthodes d'ensemble (ou équivalentes).

Si les variances d'erreurs d'ébauches sont mal spécifiées, l'incrément d'analyse sera ou trop grand ou trop petit. Avec les algorithmes variationnelles basées sur la méthode des moindres carrés, seul le rapport des variances d'erreur d'ébauche et d'observation est important pour l'analyse. Néanmoins, la valeur absolue peut aussi avoir son importance lors du contrôle de la qualité des observations. En effet, les observations éloignées de l'ébauche peuvent être tout à fait acceptées si l'erreur d'ébauche est importante.

6.1.4 Covariances d'erreur d'ébauche

Les covariances d'erreur d'ébauche sont essentielles pour faire une bonne analyse.

Propagation des informations

Dans les régions pauvres en observations, la forme de l'incrément d'analyse est complètement déterminée par les structures de covariances d'erreur d'ébauche. Ainsi, la forme de l'incrément d'analyse d'une observation esseulée est donnée directement par $\mathbf{BH}^T$. C'est donc les corrélations de $\mathbf{B}$ qui propagent l'information spatialement autour du point d'observation.

Lissage des informations

Dans les régions riches en observations, le lissage des informations est gouverné par les corrélations de la matrice de covariances d'erreur d'ébauche $\mathbf{B}$. Ceci est clairement mis en évidence au regard du gain d'analyse optimal $\mathbf{K}$ (Eq. 4.10) dont le dernier terme à être utilisé, celui le plus à gauche, est $\mathbf{B}$. Le lissage de l'incrément d'analyse est très important en ceci qu'il doit permettre à l'analyse d'avoir des échelles statistiquement compatibles avec les propriétés des champs physiques. La spécification des corrélations d'erreur d'ébauche est donc à la fois importante et délicate, car les échelles spatiales des champs physiques sont diverses et variables.

Propriétés d'équilibre

Le nombre de degrés de liberté d'un modèle est souvent supérieur à celui de la réalité. Par exemple, les courants marins sont supposés géostrophique hormis à l'équateur. Cette hypothèse d'équilibre entre gradient de pression et force de Coriolis dans les équations de Navier-Stokes permet d'obtenir une relation directe entre courant géostrophique et dérivée au premier ordre de la hauteur de mer. Ces propriétés d'équilibre peuvent être considérées comme des contraintes gênantes au problème d'analyse et être *a posteriori* appliquées brutalement. Un autre point de vue est de considérer qu'il existe des propriétés statistiques qui lient les différentes variables du modèle. S'il existe des relations d'équilibre entre les différentes variables du modèle, il doit donc y avoir aussi des relations d'équilibre linéarisés dans la matrice de covariances d'erreur d'ébauche. Ces équilibres sont très intéressants car ils permettent d'apporter des informations sur toutes les variables en équilibre avec celle observée. Ainsi, par exemple, les mesures de température en océanographie permettent de corriger la salinité. Combiné avec le lissage spatial des informations, les propriétés d'équilibre peuvent avoir un impact considérable sur la qualité de l'analyse. Une mesure de température propagée autour du point d'observation peut modifier, en plus de la température, la salinité, la hauteur de mer et les courants. C'est-à-dire toutes les variables utilisées dans un modèle océanique. L'amplitude de toutes ces modifications dépendra de l'estimation de la corrélation entre deux variables différentes et des variances d'erreur de ces variables.

Paramètres additionnels

Avec les méthodes variationnelles, il est possible d'inclure dans le vecteur de contrôle d'autres paramètres additionnels, tels que des paramètres de réglage du modèle ou des estimations du biais. Cette technique indirecte d'estimation des paramètres peut être très efficace à condition qu'il y ait une véritable relation entre ces paramètres et les observations. Cette relation passe, en général, au travers de l'opérateur d'observation ou du modèle dans le cas du 4D-Var. Il n'est généralement ni possible ni prudent de spécifier explicitement les corrélations avec les autres variables de l'état du modèle dans $\mathbf{B}$. De plus, les erreurs d'ébauche de tous les paramètres du vecteur de contrôle doivent être spécifiées prudemment, à moins d'être sûr que le problème est surdéterminé par les observations. Une variance d'erreur trop faible évitera, logiquement, de corriger les paramètres additionnels. Une variance d'erreur trop forte pourra, par contre, transformer les paramètres additionnels en source de bruit et créer des variations sans justification physique. D'importants problèmes peuvent alors apparaître car des couplages implicites dans l'analyse sont souvent créés par des dépendances dans l'opérateur d'observation ou dans le modèle en 4D-Var. Ainsi, la spécification des erreurs d'ébauche des paramètres additionnels peut avoir un impact sur les variables d'état du modèle analysées.

Dépendance à l'écoulement

La matrice de covariances d'erreur d'ébauche $\mathbf{B}$ peut dépendre de l'incertitude d'une précédente analyse ou prévision à condition que la dynamique du problème soit suffisamment connue. Non seulement les variances d'erreur d'ébauche peuvent évoluer au cours du temps, mais aussi les corrélations. En atmosphère comme en océanographie, certaines ondes suivent des motifs spécifiques qui peuvent apparaître dans les erreurs d'ébauche. Par exemple, dans un secteur enclin au développement cyclonique (région de basses pressions), les erreurs d'ébauche le plus probables devraient avoir la forme des structures les plus instables, avec peut-être une inclinaison des ondes baroclines et des anti-corrélations entre les masses d'air chaud et froid. C'est dépendance à l'écoulement est assez équivalent aux propriétés d'équilibre. Ainsi, si de tels informations peuvent être incorporées dans la matrice de covariances d'erreur $\mathbf{B}$, alors les observations pourront être mieux propagées spatialement et mieux distribuées entre les variables du modèle. Ce type d'information peut être utilisé dans le cadre des filtres de Kalman, en 4D-Var mais aussi en 3D-Var comme le mettra en évidence ce manuscrit.

6.2 Estimation des erreurs

Il est difficile d'estimer les erreurs car elles ne sont jamais observées directement. En effet, l'état vrai n'étant pas accessible, il est impossible d'obtenir des échantillons des erreurs d'ébauche et d'observation. Les données statistiques sont donc difficilement disponibles et très largement insuffisantes pour déterminer tous les éléments. Par ailleurs, les matrices de covariances d'erreur sont très grandes. Pour ces deux raisons, elles doivent être simplifiées et modélisées. De tailles réduites, ces matrices sont manipulables informatiquement parlant et nécessitent moins de d'informations statistiques pour les décrire.

La modélisation des covariances d'erreur est donc un problème difficile et il est nécessaire de faire des hypothèses d'homogénéités. La meilleure source d'information est clairement l'étude de l'innovation ($\mathbf{d} = \mathbf{y} - H\mathbf{x}^b$) et peut être utilisée de plusieurs manières différentes. D'autres informations peuvent être obtenues à partir du vecteur d'erreur d'analyse ($\mathbf{y} - H\mathbf{x}^a$) ou à partir de la valeur de la fonction coût pour les méthodes variationnelles. D'autres méthodes permettent d'estimer les covariances d'erreur d'ébauche avec des quantités dont les statistiques d'erreur sont équivalentes à celle de l'erreur d'ébauche. Parmi ce type de techniques, la méthode NMC est très connue mais ces bases théoriques sont friables. Une autre possibilité est d'utiliser une méthode d'ensemble de la même manière que pour le filtre de Kalman éponyme. Cette méthode est néanmoins applicable quelque soit la méthode d'assimilation utilisée.

6.2.1 Méthode basée sur l'innovation

Cette méthode a été introduite en météorologie à la fin de années 80 (Hollingworth et Lönnberg, 1986). Elle est basée sur l'utilisation de l'innovation ($\mathbf{d} = \mathbf{y}^o - H\mathbf{x}^b$) d'un réseau d'observations suffisamment grand et dense tel qu'il puisse fournir des informations sur les différentes échelles présentes dans la physique du système. Cette méthode permet d'obtenir des statistiques moyennes permettant de construire des matrices de covariances d'erreur statiques. Deux hypothèses importantes sont émises : les erreurs d'ébauches sont indépendantes des erreurs d'observation et les erreurs d'observation ne sont pas corrélées spatialement. Le principe est ensuite assez simple (Fig. 6.1). Il suffit de construire un histogramme représentant les covariances de du vecteur d'innovation en fonction de la distance de séparation. Pour une séparation nulle, l'histogramme fournit une information moyenne sur les variances d'erreurs d'ébauche et d'observation. Pour une distance non-nulle, l'histogramme ne fournit plus qu'une information moyenne sur les corrélations d'erreur d'ébauche.

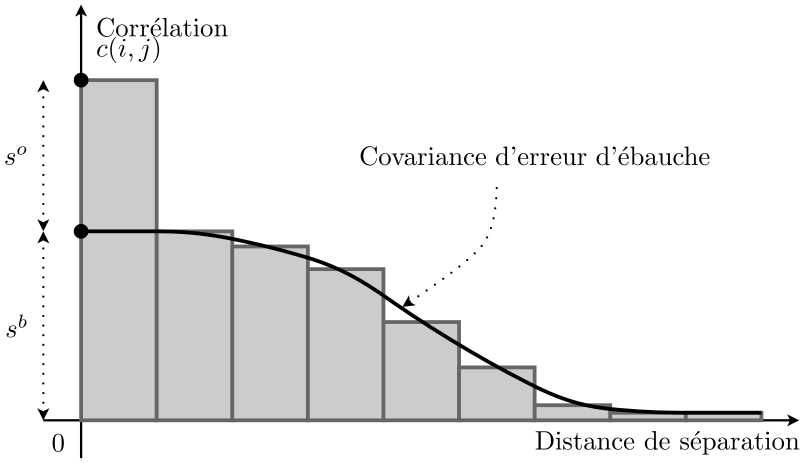


FIG. 6.1 – Représentation de la méthode basée sur l'innovation. Les statistiques de covariances de l'innovation $(\mathbf{y}^o - H\mathbf{x}^b)$ d'un système d'assimilation sont rangées dans un histogramme en fonction de la distance séparant les deux points. L'histogramme à l'origine permet d'estimer les variances d'ébauche et d'observation moyennes.

Soit deux points d'observation i et j, la covariance d'innovation $c(i,j)$ s'écrit

$$\begin{aligned}
c(i,j) &= E\left[(\mathbf{y}_i^o - \mathbf{H}_i\mathbf{x}^b)(\mathbf{y}_j^o - \mathbf{H}_j\mathbf{x}^b)^T\right] \\
&= E\left[\left((\mathbf{y}_i^o - \mathbf{H}_i\mathbf{x}^t) + (\mathbf{H}_i\mathbf{x}^t - \mathbf{H}_i\mathbf{x}^b)\right)\left((\mathbf{y}_j^o - \mathbf{H}_j\mathbf{x}^t) + (\mathbf{H}_j\mathbf{x}^t - \mathbf{H}_j\mathbf{x}^b)\right)^R\right] \\
&= E\left[(\mathbf{y}_i^o - \mathbf{H}_i\mathbf{x}^t)(\mathbf{y}_j^o - \mathbf{H}_j\mathbf{x}^t)^T\right] + \mathbf{H}_i E\left[(\mathbf{x}^t - \mathbf{x}^b)(\mathbf{x}^t - \mathbf{x}^b)^T\right]\mathbf{H}_j^T \\
&+ E\left[(\mathbf{y}_i^o - \mathbf{H}_i\mathbf{x}^t)(\mathbf{x}^t - \mathbf{x}^b)^T\right]\mathbf{H}_j^T + \mathbf{H}_i E\left[\mathbf{x}^t - \mathbf{x}^b)(\mathbf{y}_j^o - \mathbf{H}_j\mathbf{x}^t)^T\right] \\
&= \mathbf{R}_{ij} + \mathbf{H}_i\mathbf{B}\mathbf{H}_j^T
\end{aligned} \tag{6.1}$$

En utilisant l'hypothèse que l'erreur d'ébauche n'est pas corrélée à l'erreur d'observation, l'Eq. 6.1 ne conserve que deux termes : le premier est la covariance d'erreur d'observation entre les points i et j ; le second est la matrice de covariances d'erreur d'ébauche interpolée (si l'opérateur d'observation n'agit qu'en tant qu'opérateur d'interpolation) en ces points. Ceci en supposant que ces deux termes sont homogènes sur l'ensemble des observations.

Si les points i et j sont identiques ($i = j$), alors la corrélation du vecteur d'innovation au point i est la somme des variances d'erreurs d'ébauche et d'observation ($c(i,j) = s_i^o + s_i^b$). Si les points i et j sont différents ($i \neq j$) et que l'erreur d'observation n'est pas corrélée spatialement, alors la corrélation du vecteur d'innovation entre les points i et j est la covariance d'erreur d'ébauche entre ces points ($c(i,j) = \mathbf{H}_i\mathbf{B}\mathbf{H}_j^T$). À noter que la décorrélation spatiale d'erreur d'observation est fondamentale, car seule cette hypothèse permet de séparer l'information provenant de la matrice de covariances d'erreur d'observation $\mathbf{R}$ et d'ébauche $\mathbf{B}$.

A partir de ces hypothèses, si les points i et j sont très proches l'un de l'autre sans jamais être égaux, alors la corrélation du vecteur d'innovation entre les point i et j tend vers la variance d'erreur d'ébauche au point i ($\lim_{i\to j} c(i,j) = s_i^b$). En prolongeant la courbe formée par la corrélation du vecteur d'innovation vers une séparation nulle, il est donc possible d'obtenir la variance d'erreur d'ébauche. La variance d'erreur d'observation est alors la différence entre la corrélation du vecteur d'innovation pour une séparation nulle et la variance d'erreur d'ébauche obtenue ($s_i^o = c(i,j) - s_i^b$). Il est aussi possible d'obtenir les corrélations d'erreur d'ébauche en fonction de la distance de séparation en prenant le rapport de la corrélation du vecteur d'innovation sur la variance d'erreur d'ébauche ($c(i,j)/s_i^b$). Ce résultat n'est possible que si les variances d'erreur d'ébauche sont homogènes sur tout le jeu d'observations.

Si les covariances d'erreur d'ébauche ne tendent pas vers zéro pour une grande distance de séparation, c'est le signe de la présence d'un biais dans l'ébauche et/ou les observations. Dans ce cas, cette méthode ne fonctionnera pas correctement.

La méthode basée sur l'innovation est la seule méthode directe permettant de diagnostiquer les statistiques d'erreur. Cependant, elle ne fournit des informations que dans l'espace des observations et donc que dans les régions observées. Pour obtenir de bons résultats, il faut un réseau d'observations uniforme et pas trop dense pour ne pas biaiser les statistiques. Cette méthode n'est donc pas toujours très pratique pour spécifier les statistiques des erreurs. De plus, elle ne fournit que des valeurs moyennes ne permettant que de construire des matrices de covariances d'erreur statiques.

6.2.2 Méthode NMC

La méthode NMC (Parrish et Derber, 1992) est ou a été utilisée dans de nombreux centres de prévision météorologique. Elle permet de construire de construire une matrice de covariances d'erreur d'ébauche statique. L'idée est de calculer des différences entre des prévisions valides au même instant mais de durées différentes. À partir de ces différences, il est possible de d'obtenir des statistiques qui peuvent être liées à la matrice de covariances d'erreur d'ébauche $\mathbf{B}$. À partir d'un système d'assimilation séquentielle, il est très facile de mettre cette méthode en œuvre. À la fenêtre d'assimilation démarrant à l'instant t_{i-1}, une prévision d'une durée de deux cycles d'assimilation est effectuée (de t_{i-1} à t_{i+1}). À partir du même état à l'instant t_{i-1}, un cycle d'assimilation est effectué et permettent d'obtenir un état à l'instant t_i à partir duquel une prévision est effectuée de d'une durée d'un cycle (de t_i à t_{i+1}). Ces deux prévisions sont donc valides au même instant t_{i+1}. Il est possible d'envisager ces deux prévisions comme des prévisions d'une durée d'un seul cycle d'assimilation mais dont les conditions initiales à t_i varient. Ces différences de conditions initiales reflètent l'impact de l'assimilation. En calculant les différences entre ces deux prévisions et en reproduisant ces expériences suffisamment de fois, il est alors possible de calculer des statistiques sur ces différences. Le principe de la méthode NMC est illustré par la Fig. 6.2.

Comme le met en évidence Berre *et al.*(2006), l'erreur d'analyse s'écrit

$$\begin{aligned} \boldsymbol{\epsilon}_i^a &= \boldsymbol{\epsilon}_i^b + \mathbf{K}(\boldsymbol{\epsilon}_i^o + H\boldsymbol{\epsilon}_i^b), \\ &= (\mathbf{I} - \mathbf{K}H)\boldsymbol{\epsilon}_i^b + \mathbf{K}\boldsymbol{\epsilon}_i^o, \end{aligned} \tag{6.2}$$

et la différence entre les conditions initiales des deux prévisions à l'instant t_i s'écrit

$$\begin{aligned} \delta\mathbf{x}_i^a &= \mathbf{x}_i^a - \mathbf{x}_i^b, \\ &= \mathbf{K}(\boldsymbol{\epsilon}_i^o - H\boldsymbol{\epsilon}_i^b), \\ &= -\mathbf{K}H\boldsymbol{\epsilon}_i^b + \mathbf{K}\boldsymbol{\epsilon}_i^o. \end{aligned} \tag{6.3}$$

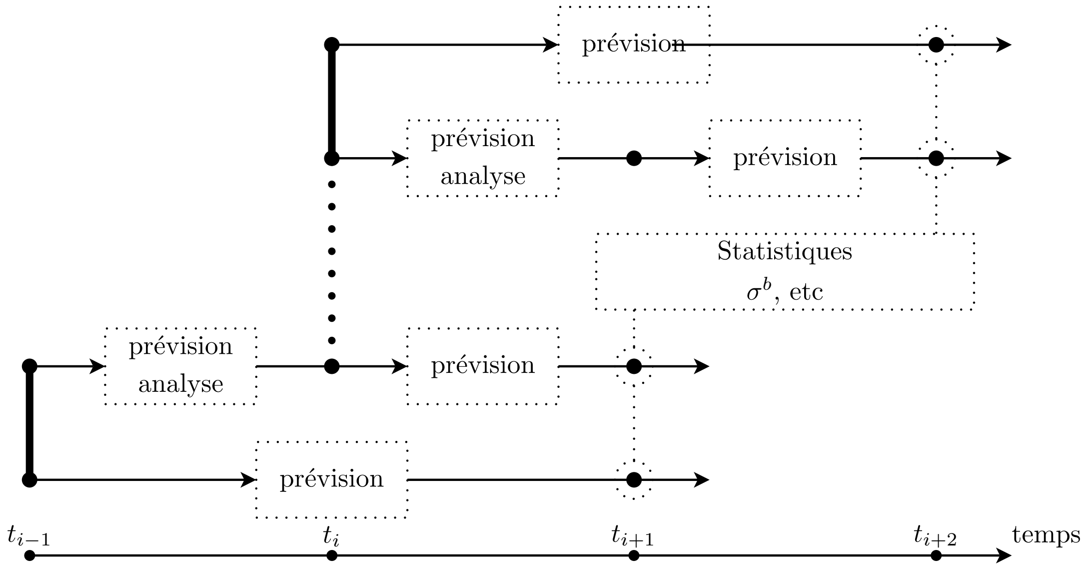


FIG. 6.2 – Méthode NMC. Une prévision démarre à t_{i-1} et dure jusqu'à t_{i+1}. Au même instant, une analyse est effectuée entre t_{i-1} et t_i. Suite à l'analyse, une prévision est effectuée jusqu'à l'instant t_{i+1}. Les différences entre les deux prévisions à l'instant t_{i+1} sont calculées. Ce processus est répété à partir de l'instant t_i et ainsi de suite. Toutes les différences permettent alors d'estimer la matrice de covariances d'erreur d'ébauche $\mathbf{B}$.

Les matrices $\mathbf{I} - \mathbf{K}H$ et $\mathbf{K}$ représentent les poids des erreurs d'ébauche et d'observation dans l'équation d'analyse. Dans la méthode NMC, le poids de l'erreur d'ébauche est approximé par $-\mathbf{K}H$. Cette approximation est raisonnable si $\mathbf{K} \sim \mathbf{I}/2$ (Boutier, 1994). Ce cas de figure est décrit par un réseau d'observations très denses ($H \sim \mathbf{I}$) et des matrices de covariances d'erreur presque identiques ($\mathbf{R} \sim H\mathbf{B}H^T \sim \mathbf{B}$). La deuxième condition signifie que l'erreur d'observation possède la même intensité et les mêmes structures spatiales que l'erreur d'ébauche.

Cependant, dans les régions pauvres en observations ou avec des observations de piètre qualité, l'incrément d'analyse risque d'être faible tandis que l'erreur d'analyse sera grande.

De plus, l'erreur d'observation est généralement moins corrélée spatialement que l'erreur d'ébauche. Comme le montre Daley (1991, section 4.5), l'opérateur $\mathbf{K}H$ agit comme un filtre passe-bas. Par conséquence, l'opérateur $\mathbf{I} - \mathbf{K}H$ devrait agir comme un filtre passe-haut. Ce qui signifie que l'incrément d'analyse doit avoir un spectre plus large que l'erreur d'analyse. Les corrélations d'erreur d'analyse risquent donc d'être surestimées avec la méthode NMC.

La perturbation d'analyse à l'instant t_i est ensuite propagée par le modèle d'évolution jusqu'à t_{i+1}. Les différences entre les prévisions permettent donc d'estimer l'erreur d'ébauche à la condition que $\mathbf{K} \sim \mathbf{I}/2$.

La méthode NMC a de nombreux avantages. Elle permet d'obtenir des statistiques dans l'espace du modèle et donc pour toutes les variables du modèle. De plus, elle est très bon marché. Cependant, elle ne représente pas parfaitement l'erreur d'ébauche car les hypothèses faites ne sont pas respectées. Ainsi, l'estimation de l'erreur d'ébauche est trop faible dans les régions peu ou mal observée.

6.2.3 Méthode d'ensemble

La méthode d'ensemble a d'abord été proposé par Evensen (Evensen, 1994) dans le cadre du filtre de Kalman d'ensemble présenté dans le chapitre 5.3. Néanmoins, cette méthode peut s'appliquer aux autres méthodes d'assimilation.

L'idée de cette méthode est de construire un ensemble composé d'une série de membres perturbés. Chacun des membres est analysé puis propagé de fenêtre d'assimilation en fenêtre d'assimilation. Ainsi, chaque membre est traité individuellement. Il est alors possible de calculer des différences entre ces membres à n'importe quel instant, puis d'obtenir des statistiques sur ces différences. La figure 6.3 permet d'illustrer l'algorithme. Il existe un lien entre les statistiques obtenues avec les différences entre les membre

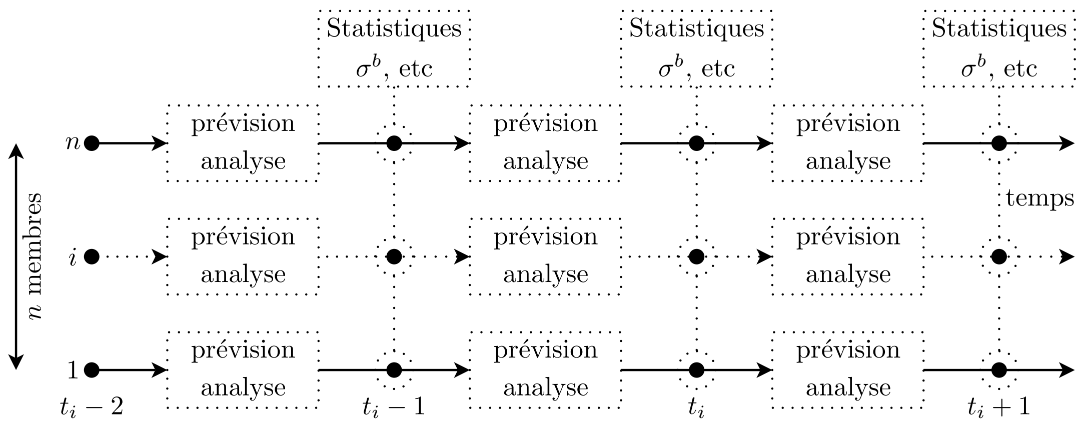


FIG. 6.3 – Méthode d'ensemble. Un ensemble est constitué de n membres perturbés qui analysés et propagés indépendamment. Après chaque cycle d'assimilation, les différences entre ces membres permettend d'obtenir des statistiques estimant la matrice de covariances d'erreur d'ébauche **B**.

et l'erreur d'ébauche. En effet, les perturbations ajoutées aux membres de l'ensemble évoluent de manière similaire à l'erreur du système d'assimilation. Ainsi, à condition de bien spécifier les perturbations, les statistiques obtenues avec les différences entre les membres sont une très bonne estimation de l'erreur d'ébauche.

Cependant, il est difficile de bien perturber les membres de l'ensemble, car les perturbations appliquées aux divers champs doivent être similaires aux covariances d'erreur de ces champs. Le problème de la connaissance de la matrice de covariances d'erreur d'ébauche est ainsi déplacé vers la connaissance des matrices de covariances d'erreur des champs perturbés. Néanmoins, ces champs à perturber peuvent être mieux connus ou leurs covariances d'erreur plus accessible.

La méthode d'ensemble est donc une méthode complexe et coûteuse. Elle a cependant des attraits non-négligeables. Elle permet d'estimer réellement les erreurs d'ébauche de toutes les variables du modèle au cours du temps. Il est ainsi possible d'obtenir une matrice de covariance d'erreur d'ébauche **B** dynamique. Cette méthode a cependant un défaut important. Si le système d'analyse est bruité, les statistiques le seront aussi et amplifieront le bruit du système d'analyse. Il est donc nécessaire d'être attentif aux risques de rétroaction.

6.3 Modélisation des erreurs

Comme il a été expliqué précédemment, la matrice de covariances d'erreur d'ébauche **B** complète est trop grande pour être spécifiée explicitement. En général, les variances qui représentent les n termes de la diagonale de **B** sont tous définis. Les termes non-diagonaux sont plus difficiles à définir. En effet, la matrice **B** doit être définie positive. Les modélisations des termes non-diagonaux doivent donc conserver cette qualité. De plus, ces modélisations doivent, en général, imposer des propriétés physiques qui seront réfléchies dans l'analyse :

- Les corrélations doivent être lissées à l'échelle des processus physiques ;
- Les corrélations doivent tendre vers zéro pour des grandes distances de séparation car les observations ne doivent avoir qu'un impact local ;
- Les corrélation ne doivent pas avoir des variations en fonction des directions ou de la position qui ne puissent être expliquées physiquement ;
- Les principales propriétés d'équilibre du système doivent être renforcées ;
- Les corrélations ne doivent pas amener des variances d'erreur d'ébauche sur les paramètres observés qui ne soient pas raisonnable.

Ces différentes exigences conduisent à une spécification des covariances d'erreur d'ébauche très complexe et qui peut être comparée à un réglage de paramètres physiques. Les hypothèses reposant sur la physique doivent effectuées et testées avec beaucoup de précaution. La liste ci-dessous cite un certains nombre de techniques populaires :

- Les modèles de corrélation peuvent être définis indépendamment des champs de variances à la condition que les échelles de variation des variances soit plus grande que celles des corrélations (sinon la forme des covariances sera très différente de celle des corrélations et les conséquences sur les propriétés d'équilibre seront imprévisibles) ;
- Les matrices d'autocorrélations verticales de chaque paramètre sont suffisamment petites pour être spécifiées explicitement ;
- Les autocorrélations horizontales ne peuvent pas être spécifiées explicitement, mais elles peuvent être réduites à des matrices creuses sous l'hypothèse d'homogénéité et d'isotropie. Ces matrices creuses sont alors diagonales dans l'espace spectral et comparables à des filtres digitales passe-bas dans l'espace physique ;
- Des modèles de corrélations tridimensionnelles et multivariées peuvent être construits en utilisant les hypothèses de séparabilité, d'homogénéité et d'indépendance ;
- Les contraintes d'équilibres peuvent être utilisées en transformant les variables du modèle en composantes équilibrées et non-équilibrées. La partie non-équilibrée est supposée avoir une variance d'erreur d'ébauche plus faible que la partie équilibrée, ce qui signifie quelle contribue moins à la construction de l'incrément d'analyse ;

Il existe, bien sûr, beaucoup d'autres techniques permettant de construire des opérateurs modélisant les covariances d'erreur d'ébauche.

Chapitre 7

Conclusion

> Si furent leurs propos à la premiere rencontre les plus gracieux qu'il est possible : mais à la fin la conclusion n'en fut ny belle ny bonne.
>
> *Jacques Amyot* in *Pompée, 47*

L'objectif de cette conclusion est de mettre en évidence les avantages des différentes méthodes dans le cadre d'une utilisation réaliste. Comme il a déjà été montré dans la section 4.6, le BLUE peut-être écrit de manière variationnelle. De ces deux écritures découlent, globalement, deux façons d'aborder le problème de l'assimilation de données. La première est construite autour des filtres de Kalman, tandis que la seconde utilise l'approche variationnelle. Dans un cadre linéaire, le filtre de Kalman étendu et le 4D-Var sont identiques à l'optimalité. De ce point de vue, il est difficile de s'orienter vers une méthode plutôt qu'une autre.

Dans la pratique, les systèmes étudiés sont souvent de grandes tailles et non-linéaires. C'est le cas en météorologie ou en océanographie. Le coût de calcul devient alors une contrainte très forte sur le choix de la méthode d'assimilation. Le filtre de Kalman et sa version étendue ainsi que le 4D-Var sont, en général, inaccessibles. Il existe néanmoins des approximations qui permettent de réduire notablement les coûts de calculs. Pour le filtre de Kalman, il existe tout une série de filtres réduits (SEEK, SEIK, SFEK, EnKF...). L'approche variationnelle peut aussi se décliner dans sa version incrémentale et 3D-Var.

Dans le filtre de Kalman, les matrices de covariances d'erreur sont propagées explicitement. Au contraire, dans les méthodes variationnelles, elles doivent être construites. Cette différence est un des aspects les plus positifs du filtre de Kalman comparée à l'approche variationnelle. Cependant, d'une certaine manière, la complexité du filtre de Kalman est reportée sur la matrice de covariance d'erreur modèle. Ainsi, l'avantage du filtre de Kalman disparaît quasiment pour les filtres réduits où de nombreuses hypothèses et contraintes sont rajoutées en plus pour diminuer les coûts de calcul. L'approche variationnelle, qui nécessite la définition d'opérateurs pour modéliser les matrices de covariances d'erreur, a ainsi l'intérêt de travailler sur l'espace complet du vecteur d'état et à haute résolution. D'un côté, il y a donc des matrices de rang réduit propagées explicitement, de l'autre il y a des matrices de rang complet modélisées par des opérateurs. Ces deux approches ont donc, toutes les deux, des défauts.

Les problèmes de non-linéarités sont très importants et le deviennent actuellement de plus en plus. En effet, la résolution des modèles d'évolution ne cesse d'augmenter et les phénomènes non-linéaires sont de plus en plus présents. Le filtre de Kalman étendu et le 4D-Var sont bien adaptés pour répondre à ces problèmes de non-linéarités. Malheureusement, les versions réduites du filtre de Kalman étendu et l'approche incrémentale du 4D-Var perdent une partie de ces capacités. Aucune des deux approches n'a, aujourd'hui, fait la preuve d'une plus grande efficacité.

La prise en compte de l'erreur modèle dans le filtre de Kalman est un aspect très intéressant. Cependant, en conséquence, le coût de la méthode est important. Le 4D-Var à contrainte faible permet, lui aussi, de tenir compte de l'erreur modèle. De la même manière, les coûts sont importants. Néanmoins, les deux approches permettent d'en tenir compte et ce critère n'est pas pertinent pour faire un choix.

Une qualité intrinsèque de la méthode variationnelle est la possibilité de rajouter d'autres termes à la fonction coût. Il est ainsi possible de rajouter un terme pour l'erreur modèle, mais aussi de rajouter un autre terme pour corriger le biais des observations par exemple.

Un autre point positif de l'approche variationnelle est la facilité de passage entre une 3D-Var FGAT et un 4D-Var. Cet aspect permet d'obtenir un système économique permettant de faire beaucoup de tests avec le 3D-Var FGAT et un système très évolué mais beaucoup plus onéreux avec le 4D-Var.

Ceci met en évidence un des défauts majeur du 4D-Var qui requiert l'écriture du modèle tangent-linéaire et adjoint de modèle non-linéaire. Le codage de ces modèles est souvent long et difficile du fait de la complexité des modèles non-linéaires.

Une autre critique à l'approche variationnelle est qu'elle ne fournit pas *a priori* d'information sur la matrice de covariance d'erreur d'analyse **A**. L'ajout de son calcul est possible (Fisher, 1998), mais coûteux.

Concernant les non-linéarités, le filtre de Kalman d'ensemble permet théoriquement une prise en compte efficace par l'utilisation du modèle non-linéaire pour propager l'état des différents membres. Son point faible est la taille de l'ensemble qui doit être conséquente pour que les statistiques calculées soient pertinentes. Néanmoins, cette approche profite de la présence de plus en plus courante d'ensemble nécessaire aux prévisions probabilistes. Si un ensemble est requit pour des raisons extérieures à l'assimilation de données, son utilisation par celle-ci peut-être considérée à coût nul. Cependant, les méthodes variationnelles peuvent aussi utiliser des ensembles à leur profit de la même manière que le filtre de Kalman d'ensemble.

Cette énumération non exhaustive des différents points positifs et négatifs ne permettent pas de faire ressortir clairement une méthode par rapport aux autres. La plupart des centres opérationnels en météorologie ont, pour leur part, opté pour des 4D-Var incrémentaux. Les filtres de Kalman d'ensemble sont aussi actuellement très en vogue. Au CERFACS, c'est l'approche variationnelle qui a été choisie sous sa forme incrémentale et avec la possibilité d'utiliser l'algorithme en tant que 3D-Var FGAT ou 4D-Var.

Table des figures

Liste des tableaux

Index

Bibliographie

[Alves et al., 2004] Alves, O., Balmaseda, M. A., Anderson, D. L. T., and Stockdale, T. N. (2004). Sensitivity of dynamical seasonal forecasts to ocean initial condition. *Quart. J. Roy. Meteor. Soc.*, 130 :647–668.

[Andersson et al., 1998] Andersson, E., Haseler, J., Undén, P., Courtier, P., G. Kelly, D. V., Brankovic, C., Cardinali, C., Gaffard, C., Hollingsworth, A., Jakob, C., Janssen, P., Klinker, E., A. Lanzinger, M. M., Rabier, F., Simmons, A., Strauss, B., ThÃ©paut, J.-N., and Viterbo, P. (1998). The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part III : Experimental results. *Quart. J. Roy. Meteor. Soc.*, 124 :1831–1860.

[Behringer et al., 1998] Behringer, D. W., Ji, M., and Leetmaa, A. (1998). An improved coupled model for enso prediction and implications for ocean initialization. part i : The ocean data assimilation system. *Mon. Wea. Rev.*, 126 :1013–1021.

[Bennett et al., 1998] Bennett, A. F., Chua, B. S., Harrison, D. E., and McPhaden, M. J. (1998). Generalized inversion of tropical atmosphere-ocean data and a coupled model of the tropical pacific. *J. Clim.*, 11 :1768–1792.

[Bennett and McIntosh, 1982] Bennett, A. F. and McIntosh, P. C. (1982). Open ocean modeling as an inverse problem : tidal theory. *J. Phys. Oceanogr.*, 12 :1004–1018.

[Bloom et al., 1996] Bloom, S., Takacs, L., da Silva, A., and Ledvina, D. (1996). Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, 124 :1256–1271.

[Brasseur et al., 1999] Brasseur, P., Ballabrera-Poy, J., and Verron, J. (1999). Assimilation of altimetric observations in a primitive equation model of the Gulf Stream using singular evolutive extented Kalman filter. *J. Mar. Sys.*, 22 :269–294.

[Burgers et al., 1998] Burgers, G., Leeuwen, P. J. V., and Evensen, G. (1998). Analysis scheme in the ensemble kalman filter. *Mon. Wea. Rev.*, 126 :1719–1724.

[Carton et al., 1996] Carton, J. A., Giese, B. S., Cao, X., and Miller, L. (1996). Impact of altimeter, thermistor, and expendable bathythermograph data on retrospective analyses of the tropical pacific ocean. *J. Geophys. Res.*, 101 :14147–14159.

[Cohn et al., 1998] Cohn, S. E., Silva, A. D., Guo, J., Sienkiewicz, M., and Lamich, D. (1998). Assessing the effects of data selection with DAO i Physical-Space statistical analysis system. *Mon. Wea. Rev.*, 126 :2913–2926.

[Courtier et al., 1998] Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., and Fisher, M. (1998). The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part I : Formulation. *Quart. J. Roy. Meteor. Soc.*, 124 :1783–1808.

[Daley, 1991] Daley, R. (1991). *Atmospheric Data Analysis*. Cambridge University Press.

[Deltel, 2002] Deltel, C. (2002). *Estimation du cycle saisonnier de l'Atlantique Sud par assimilation variationnelle de données in situ. Impact du contrôle optimal de l'hydrologie aux frontières ouvertes et des champs de forçage air-mer*. PhD thesis, Université de Bretagne Sud, Brest, France.

[Derber, 1989] Derber, J. C. (1989). A variational continuous assimilation technique. *Mon. Wea. Rev.*, 117 :2437–2446.

[Dimet and Talagrand, 1986] Dimet, F. X. L. and Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations : Theoretical aspects. *Tellus*, 38A :97–110.

[Evensen, 1992] Evensen, G. (1992). Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.*, 97 :831–835.

[Evensen, 1994] Evensen, G. (1994). Sequential data assimilation with non-linear quasi-geostrophic model using monte carlo methods to forecast error statistics. *J. Geophys. Res.*, 99 :10143–10162.

[Evensen, 2003] Evensen, G. (2003). The ensemble kalman filter : theoretical formulation and practical implementation. *Ocean Dynamics*, 53 :343–367.

[Fisher, 1998] Fisher, M. (1998). Minimization algorithms for variational data assimilation. In *Seminar on Recent Developments in Numerical Methods for Atmospheric Modelling*, pages 364–385. ECMWF.

[Fisher, 1912] Fisher, R. A. (1912). On an absolute criterion for fitting frequency curves. *Mess. Math.*, 41 :155–160.

[Fisher, 1925] Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

[Fu and Smith, 1996] Fu, L.-L. and Smith, R. D. (1996). Global ocean circulation from satellite altimetry and high-resolution computer simulation. *Bull. Am. Met. Soc.*, 77 :2625–26r326.

[Fukumori and Malanotte-Rizzoli, 1995] Fukumori, I. and Malanotte-Rizzoli, P. (1995). An approximate Kalman filter for ocean data assimilation : An example with idealized Gulf Stream model. *J. Geophys. Res.*, 100 :6777–6793.

[Gandin, 1963] Gandin, L. S. (1963). *Objective Analysis of Meteorological Fields*. Gidrometeorologich eskoe Izdatel 'sivo.

[Gauthier et al., 1994] Gauthier, P., Courtier, P., and Moll, P. (1994). Assimilation of simulated wind lidar with a Kalman filter. *Mon. Wea. Rev.*, 121 :1803–1820.

[Gelb, 1974] Gelb, A. (1974). *Applied Optimal Estimation*. Academic Press, New York.

[Gent and McWilliams, 1990] Gent, P. R. and McWilliams, J. C. (1990). Isopycnal mixing in ocean circulation models. *J. Phys. Oceanogr.*, 20 :150–155.

[Ghil et al., 1981] Ghil, M., Cohn, S., Tavanzis, J., Bube, K., and Isaacson, E. (1981). *Application of estimation theory to numerical weather prediction. Dynamic Meteorology : Data assimilation method.* Eds. L. Bengtsson, M. Ghil and E. Källen, Springer-Verlag.

[Giering and Kaminski, 1998] Giering, R. and Kaminski, T. (1998). Recipes for adjoint code construction. *ACM Trans. Mathematical Software*, 4 :437–474.

[Greiner et al., 1998a] Greiner, E., Arnault, S., and Moliere, A. (1998a). Twelve monthly experiments of 4D-Variational assimilation in the tropical Atlantic during 1987. Part I : Method and Statistical results. *Progress in Oceanography*, 41 :141–202.

[Greiner et al., 1998b] Greiner, E., Arnault, S., and Moliere, A. (1998b). Twelve monthly experiments of 4D-Variational assimilation in the tropical Atlantic during 1987. Part II : Oceanographic interpretation. *Progress in Oceanography*, 41 :203–247.

[Greiner and Périgaud, 1998] Greiner, E. and Périgaud, C. (1998). Assimilation of Geostat altimetric data in a non-linear reduced-gravity model of the Indian Ocean. Part I : Adjoint approach and model-data consistency. *J. Phys. Oceanogr.*, 24 :1783–1804.

[Griffies et al., 2000] Griffies, S. M., Böning, C., Bryan, F. O., Chassignet, E. P., Gerdes, R., Hasumi, H., Hirst, A., Treguier, A.-M., and Webb, D. (2000). Developments in ocean climate modelling. *Ocean Modelling*, 2 :123–192.

[Holland, 1989] Holland, W. R. (1989). Altimeter data assimilation into ocean ciculation models – some preliminary results. In Anderson, D. L. T. and Willebrand, J., editors, *Oceanic Circulation Models : Combining Data and Dynamics*, pages 203–232. Kluwer Academic Publishers, Dordrecht, The Netherlands.

[Hollingsworth and Lönnberg, 1986] Hollingsworth, A. and Lönnberg, P. (1986). The statistical structure of short-range forecast errors as determined from radiosonde data. part i : The wind field. *Tellus*, 38A :111–136.

[Holloway, 1992] Holloway, G. (1992). Representing topographic stress for large-scale ocean models. *J. Phys. Oceanogr.*, 22 :1033–1046.

[Hoteit, 2001] Hoteit, I. (2001). *Efficient reduced Kalman filters for data assimilation in oceanography.* PhD thesis, Université Joseph-Fourier - Grenoble I - France.

[Houtekamer and Mitchell, 1998] Houtekamer, P. L. and Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, 126 :796–811.

[Jazwinski, 1970] Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*, volume 64 of *Mathematics in science and engineering.* Academic Press (New-York).

[Ji et al., 1998] Ji, M., Behringer, D. W., and Leetmaa, A. (1998). An improved coupled model for enso prediction and implications for ocean initialization. part ii : The coupled model. *Mon. Wea. Rev.*, 126 :1022–1034.

[Ji and Leetmaa, 1997] Ji, M. and Leetmaa, A. (1997). Impact of data assimilation on ocean initialization and el niño prediction. *Mon. Wea. Rev.*, 125 :742–753.

[Kalman, 1960] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *J. Bas. Eng.*, 82 :35–45.

[Kalman and Bucy, 1961] Kalman, R. E. and Bucy, R. S. (1961). New results in linear filtering and prediction theory. *J. Bas. Eng.*, 83 :95–107.

[Koblinsky and Smith, 2001] Koblinsky, C. J. and Smith, N. R. (2001). *Observing the oceans in the* 21^{th} *century*. GODAE Project office bureau of Meteorology.

[Kushner, 2003] Kushner, H. (2003). Approximation to optimal nonlinear filters. *IEEE trans. Autom. Control*, AC-12 :546–556.

[Large et al., 1994] Large, W. G., McWilliams, J. C., and Doney, S. C. (1994). Oceanic vertical mixing : a review and a model with a nonlocal boundary layer parameterization. *Rev. Geophys.*, 32 :363–403.

[Lee et al., 2000] Lee, T., Boulanger, J. P., Fu, L. L., Foo, A., and Giering, R. (2000). Data assimilation by a simple coupled ocean-atmosphere model : application to the 1997-98's El Niño. *J. Geophys. Res.*, 105(C11) :26063–26088.

[Lee and Marotzke, 1998] Lee, T. and Marotzke, J. (1998). Seasonal cycle of meridional overturning and heat transport of the Indian Ocean. *J. Phys. Oceanogr.*, 28 :923–943.

[Leetmaa and Ji, 1989] Leetmaa, A. and Ji, M. (1989). Operational hindcasting of the tropical pacific. *Dyn. Atmos. Oceans*, 13 :465–490.

[Lions, 1968] Lions, J.-L. (1968). *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*, volume 1. DUNOD.

[Lorenc, 1981] Lorenc, A. C. (1981). A global three-dimensional multivariate statistical interpolation scheme. *Quart. J. Roy. Meteor. Soc.*, 109 :701–721.

[Lorenc, 1986] Lorenc, A. C. (1986). Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, 112 :1177–1194.

[Luong et al., 1998] Luong, B., Blum, J., and Verron, J. (1998). A variational method for the resolution of data assimilation problem in oceanography. *Inverse problems*, 14 :979–997.

[Miller, 1990] Miller, R. N. (1990). Tropical data assimilation experiments with simulated data : The impact of the tropical ocean and global atmosphere thermal array for the ocean. *J. Geophys. Res.*, 95 :11461–11482.

[Moore, 1991] Moore, A. (1991). Data assimilation in a quasi-geostrophic operational model of the Gulf Stream region using the adjoint method. *J. Phys. Oceanogr.*, 21 :398–427.

[Moore, 1986] Moore, A. M. (1986). *Data assimilation in ocean models*. PhD thesis, University of Oxford.

[Parrish and C., 1992] Parrish, D. F. and C., J. C. D. J. (1992). The national meteorological center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, 120 :1747–1763.

[Pham, 2001] Pham, D. T. (2001). Stochastic methods for sequential data assimilation in strongly non-linear systems. *Mon. Wea. Rev.*, 129 :1194–1207.

[Pham et al., 1998] Pham, D. T., Verron, J., and Roubaud, M. C. (1998). A singular evolutive extented kalman filter for data assimilation in oceanography. *J. Mar. Sys.*, 16 :323–340.

[Rabier et al., 2000] Rabier, F., Järvinen, H., Kilnder, E., Mahfouf, J.-F., and Simmons, A. (2000). The ECMWF operational implementation of four-dimensional variational assimilation. part i : Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, 126 :1143–1170.

[Sasaki, 1955] Sasaki, Y. K. (1955). A fundamental study of the numerical prediction based on the variational principle. *J. Met. Soc. Japan*, 33 :262–275.

[Sasaki, 1958] Sasaki, Y. K. (1958). An objective analysis based on the variational method. *J. Met. Soc. Japan*, II(36) :77–88.

[Sasaki, 1970] Sasaki, Y. K. (1970). Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev.*, 98 :875–883.

[Segschneider et al., 2000] Segschneider, J., Anderson, D. L. T., and Stockdale, T. N. (2000). Towards the use of altimetry for operational seasonal forecasting. *J. Clim.*, 13 :3116–3138.

[Shröter et al., 1993] Shröter, J., Seiler, U., and Wensel, M. (1993). Variational Assimilation of Geosat Data into a Eddy-resolving Model of the Gulf Stream Extension Area. *J. Phys. Oceanogr.*, 23 :925–953.

[Smedstad and O'Brien, 1991] Smedstad, O. M. and O'Brien, J. J. (1991). Variational data assimilation and parameter estimation in an equatorial pacific ocean model. *J. Phys. Oceanogr.*, 26 :179–241.

[Stammer et al., 2002] Stammer, D., Wunsch, C., Giering, R., Eckert, C., Heimbach, P., Marotze, J., Adcroft, A., Hill, N. C., and Marshall, J. (2002). Global ocean circulation during 1992-1997, estimated from ocean observations and a general circulation model. *J. Geophys. Res.*, 107(C9) :3118.

[Stammer et al., 1997] Stammer, D., Wunsch, C., Giering, R., Zhang, Q., Marotze, J., and Hill, C. (1997). The global ocean circulation estimation with TOPEX/POSEIDON altimetry and the MIT general circulation model. Technical Report 49, Center for Global Change Science, Massachusett Institute of Technology, Cambridge, MA.

[Thacker and Long, 1988] Thacker, W. C. and Long, R. B. (1988). Fitting dynamics to data. *J. Geophys. Res.*, 93 :1227–1240.

[Thépaut and Courtier, 1991] Thépaut, J. N. and Courtier, P. (1991). Four dimensional variational data assimilation using the adjoint of a multilevel primitive equation model. *Quart. J. Roy. Meteor. Soc.*, 117 :1225–1254.

[Thépaut et al., 1993] Thépaut, J. N., Hoffman, R., and Courtier, P. (1993). Interaction of dynamic and observations in a four-dimensional variational assimilation. *Mon. Wea. Rev.*, 121 :3393–3414.

[Vidard, 2001] Vidard, P. A. (2001). *Vers une prise en compte des erreurs modèle en assimilation de données 4D-variationnelle. Application à un modèle réaliste d'océan.* PhD thesis, Université Joseph Fourier, Grenoble, France.

[Vossepoel and Behringer, 2000] Vossepoel, F. C. and Behringer, D. W. (2000). Impact of sea level assimilation on salinity variability in the western equatorial pacific. *J. Phys. Oceanogr.*, 30 :1706–1721.

[Weaver and Anderson, 1997] Weaver, A. T. and Anderson, D. L. T. (1997). Variational assimilation of altimeter data in a multi-layer model of the tropical Pacific Ocean. *J. Phys. Oceanogr.*, 27 :664–682.

[Wiener, 1949] Wiener, N. (1949). *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications.* MIT Press, Cambridge.

[Zupanski, 1993] Zupanski, M. (1993). Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Wea. Rev.*, 121 :2396–2408.