

# Assimilation de données

François Bouttier (Météo-France), novembre 2004

Toulouse Cours de Master OASC à l'université Paul Sabatier,

## Préface

Ce cours d'assimilation de données décrit les principales techniques d'analyse de données et d'assimilation utiles en modélisation des fluides géophysiques. Par commodité on y fera souvent référence au fluide 3D atmosphérique, mais ces techniques s'appliquent aussi bien à l'océan, à la chimie atmosphérique, etc., ou à toute sorte de systèmes distribués dans l'espace (pour l'analyse de données) et présentant une évolution temporelle (pour l'assimilation) : hydrologie 2D, transfert radiatif 1D, modèles biologiques, machines, circulation sanguine, planètes, etc...

L'accent est mis sur les méthodes variationnelles, qui sont bien adaptées aux modèles dynamiques de grande dimension, mais d'autres techniques plus simples sont aussi mentionnées. L'objectif principal étant d'aider les jeunes chercheurs à appliquer ces méthodes à des problèmes concrets, les aspects théoriques sous-jacents, au demeurant fort intéressants (systèmes continus, probabilités et analyse numérique), ne sont que brièvement mentionnés pour laisser plus de place à la mise en oeuvre pratique des algorithmes. Le strict minimum nécessaire d'équations mathématiques est utilisé, ce qui permet avec un effort intellectuel minimal (et un bagage rarement au-delà d'un baccalauréat scientifique) d'introduire rigoureusement un large éventail de techniques d'assimilation, ainsi que d'autres techniques utiles qui reposent sur le même cadre : prévisibilité et études de sensibilité. Le vocabulaire spécialisé, indispensable pour comprendre les publications scientifiques, est introduit en français (et en anglais lorsque la traduction n'est pas triviale).

# Table des matières

<b>1</b>	<b>Concepts de base</b>	<b>4</b>
1.1	La définition du modèle . . . . .	7
1.2	Méthodes de base d'analyse de données . . . . .	7
1.3	Les avantages d'une approche statistique . . . . .	10
<b>2</b>	<b>Vecteur d'état, espaces de contrôle et d'observation</b>	<b>12</b>
2.1	Vecteur d'état . . . . .	12
2.2	Variable de contrôle . . . . .	12
2.3	Observations . . . . .	12
<b>3</b>	<b>La modélisation statistique des erreurs</b>	<b>14</b>
3.1	Modèles d'erreur . . . . .	14
3.2	Estimation pratique des covariances . . . . .	15
<b>4</b>	<b>Interpolation statistique aux moindres carrés, ou BLUE</b>	<b>17</b>
4.1	Notations et hypothèses . . . . .	17
4.2	Théorème : équations d'analyse aux moindres carrés . . . . .	18
4.3	Commentaires . . . . .	20
4.3.1	Cas de dégénérescence du BLUE . . . . .	20
4.3.2	Traitement des biais . . . . .	20
4.3.3	Autres hypothèses qui affectent l'analyse BLUE . . . . .	21
4.4	Sur l'hypothèses linéaire tangente . . . . .	21
4.5	Le point de vue des probabilités conditionnelles . . . . .	23
4.6	Coût numérique de l'analyse aux moindres carrés . . . . .	23
4.7	Conclusion . . . . .	24
<b>5</b>	<b>Exemple scalaire simple d'analyse BLUE</b>	<b>26</b>
<b>6</b>	<b>Modèles de covariances d'erreurs</b>	<b>30</b>
6.1	Technique d'estimation of error covariances . . . . .	31
<b>7</b>	<b>Optimal interpolation (OI) analysis</b>	<b>34</b>
<b>8</b>	<b>Three-dimensional variational analysis (3D-Var)</b>	<b>37</b>
<b>9</b>	<b>1D-Var and other variational analysis systems</b>	<b>40</b>
<b>10</b>	<b>Four-dimensional variational assimilation (4D-Var)</b>	<b>41</b>
10.1	The four-dimensional analysis problem . . . . .	41
10.2	Theorem : minimization of the 4D-Var cost function . . . . .	42
10.3	Properties of 4D-Var . . . . .	43
10.4	Equivalence between 4D-Var and the Kalman Filter . . . . .	45
<b>11</b>	<b>Estimating the quality of the analysis</b>	<b>46</b>
11.1	Theorem : use of Hessian information . . . . .	46
11.2	Remarks . . . . .	47

<b>12 Implementation techniques</b>	<b>49</b>
12.1 Minimization algorithms and preconditioning . . . . .	49
12.2 Theorem : preconditioning of a variational analysis . . . . .	50
12.3 The incremental method . . . . .	51
12.4 The adjoint technique . . . . .	52
<b>13 Dual formulation of 3D/4D-Var (PSAS)</b>	<b>55</b>
<b>14 The extended Kalman filter (EKF)</b>	<b>57</b>
14.1 Notation and hypotheses . . . . .	57
14.2 Theorem : the KF algorithm . . . . .	58
14.3 Theorem : KF/4D-Var equivalence . . . . .	58
14.4 The Extended Kalman Filter (EKF) . . . . .	58
14.5 Comments on the KF algorithm . . . . .	59
<b>15 Conclusion</b>	<b>60</b>
<b>16 References</b>	<b>62</b>
<b>A Ecriture d'un code adjoint</b>	<b>64</b>
<b>B Principaux symboles utilisés</b>	<b>65</b>

# 1 Concepts de base

**Analyse.** Une analyse<sup>1</sup> est la production d'une image aussi fidèle que possible de l'état réel de l'atmosphère (ou n'importe quel système géophysique) à un instant donné, représentée dans un modèle par un ensemble de nombres. Le résultat, appelé aussi analyse, peut être utilisée simplement comme diagnostic global et auto-cohérent de l'atmosphère, ou comme condition initiale d'une prévision, comme un moyen de présenter des données mesurées de manière physiquement plus facile à comprendre, ou encore comme vérité de référence qui permettra de juger si une donnée est vraisemblable ou corrompte.

L'information objective permettant d'effectuer une analyse est un ensemble de données mesurées, qui sont des observations du système réel. Si l'état du modèle est complètement déterminé par les observations, l'analyse se réduit à un problème d'interpolation. Souvent, le problème d'analyse est sous-déterminé (au moins localement) par manque de données. Pour en faire un problème bien posé il faut prendre en compte des informations externes, par exemple une estimation a priori de l'état du modèle : climatologie, état trivial vraisemblable, prévision récente, ou analyse récente si l'on peut supposer que le système évolue lentement (hypothèse de persistance). L'imposition de contraintes physiques facilite aussi l'analyse.

*Assimilation.* L'assimilation de données est une technique d'analyse qui utilise des propriétés d'évolution du système considéré pour mieux l'estimer, en y accumulant l'information apportée par une séquence d'observations. C'est la dimension temporelle de ce processus qui distingue l'assimilation d'une simple analyse. Par exemple, l'assimilation des données d'une sonde fixe dans un fluide en mouvement permet intuitivement de connaître les propriétés de l'ensemble des particules qui traversent cette sonde au fil du temps, si l'on dispose d'un modèle fiable d'évolution des propriétés de ces particules. La théorie d'assimilation de données vise à formaliser ce type d'intuition : utiliser au mieux les données observées en fonction des propriétés de propagation de l'information dans le système physique modélisé.

Il y a 2 approches de l'assimilation de données : *séquentielle*, qui traite les observations du passé jusqu'à l'instant de l'analyse (utile dans les systèmes temps réel), ou *rétrospective*, dans laquelle les observations du futur (par rapport à chaque instant d'analyse) peuvent être utilisées, par exemple lors d'une réanalyse. Une autre classification importante porte sur le caractère temporellement *intermittent* ou *continu*. Dans une méthode intermittente, les observations peuvent être traitées par petits paquets, ce qui présente souvent des avantages techniques. Dans une méthode continue, les observations sont souvent traitées en bloc sur de longues périodes, et la séquence des analyses produites (appelée *trajectoire*) est temporellement continue, ce qui est physiquement plus réaliste (cf. fig.1. On peut trouver des compromis entre ces approches.

Il existe de nombreuses techniques d'assimilations. (figure 2). Elles diffèrent par leur coût numérique, leur degré d'optimalité, et leur adaptation aux contraintes d'assimilation en temps réel. La plupart sont expliquées dans ce cours.

ref : Daley 1991, Lorenc 1986, Ghil 1989.

---

<sup>1</sup> *analysis* en anglais se prononce 'anälizis', le verbe *to analyze* se prononce 'änalaiz'.

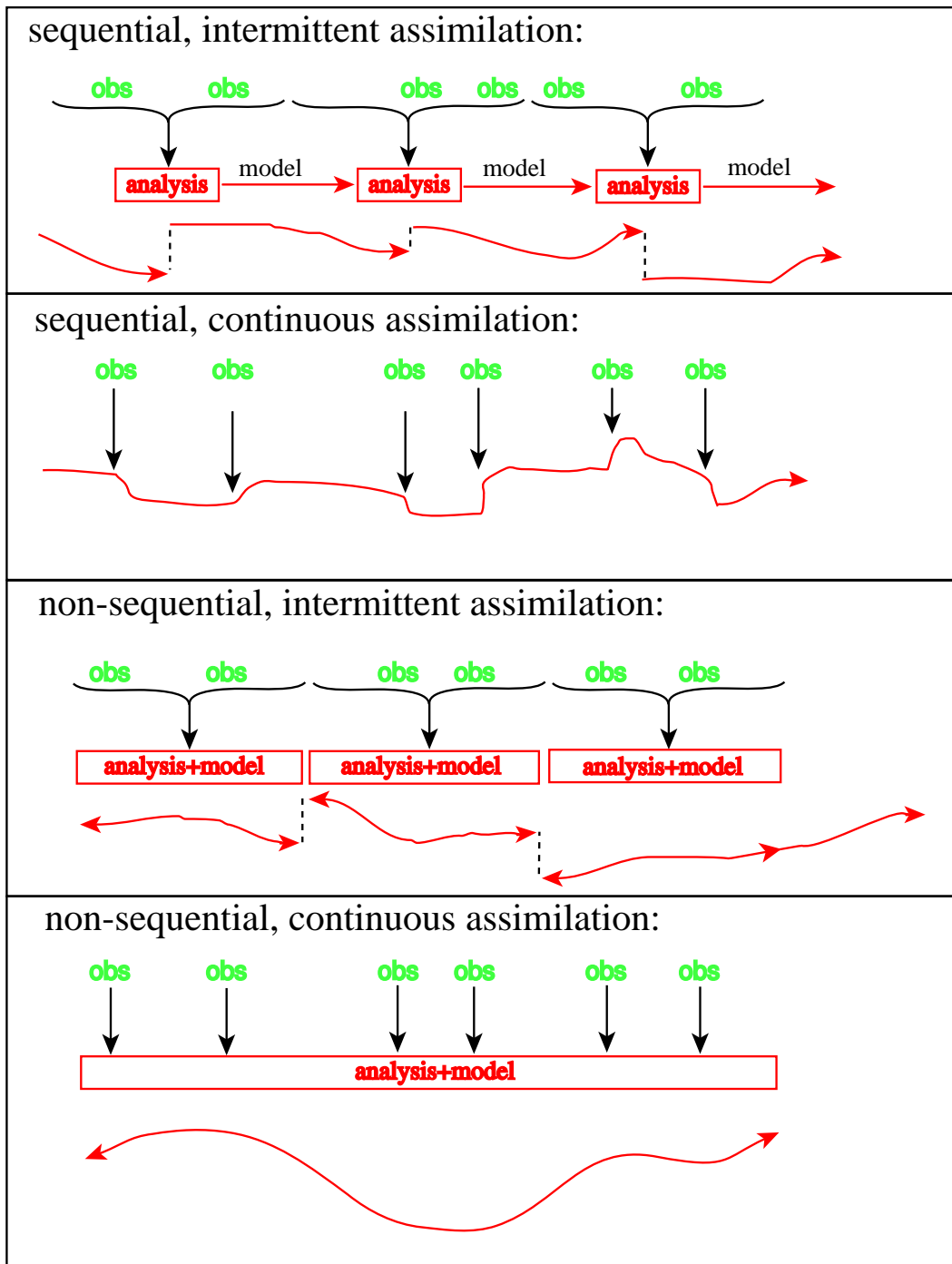


FIG. 1 – Représentation de 4 stratégies d’assimilation de données en fonction du temps : la manière de transformer une séquence d’observations en une séquence d’états assimilés peut être séquentielle ou continue.

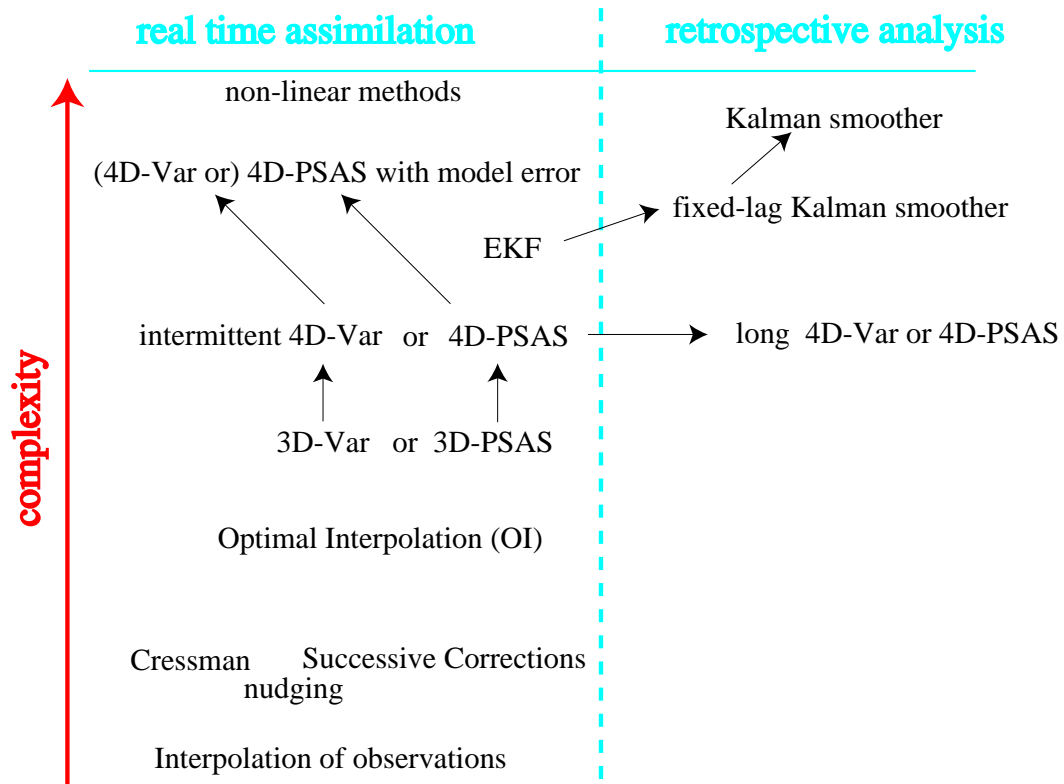


FIG. 2 – Histoire résumée des principaux algorithmes d’assimilation développés pour la météorologie et l’océanographie, et leur pertinence pour les applications temps réel. Actuellement, les principales méthodes sont 3D-Var, 4D-Var, et des versions simplifiées du filtre de Kalman.

## 1.1 La définition du modèle

Les concepts exposés ici sont transposables à toutes sortes de modèles : globaux ou à aire limitée, 3D, 2D, 1D ou scalaires. Ils sont aussi bien valables dans des modèles continus que discrets, mais les outils mathématiques nécessaires pour être rigoureux dans le cas continu étant parfois un peu complexes (notamment, théories des tenseurs et des distributions, espaces de Sobolev), on ne traitera ici que de problèmes discrets à dimension finie

Il existe souvent plusieurs manières de définir un modèle d'un système physique. Les champs eux-mêmes peuvent être discrétisés sur une grille (càd. par des moyennes des champs dans des boîtes), sur une base de fonctions spectrales, de composantes principales, d'éléments finis, etc. Le vent peut être représenté par ses composantes spatiales, sa force et son intensité, les champs de tourbillon et de divergence, les espèces chimiques par leur concentration en masse ou en volume, par leur rapport de saturation, etc. tout cela ne change pas le problème d'analyse, seulement sa représentation. Il est important de comprendre que le problème d'analyse peut être formulé dans n'importe quelle transformation inversible de l'état du modèle d'évolution : un choix astucieux des variables utilisées peut énormément simplifier le travail, notamment pour la formulation des statistiques d'erreurs nécessaires.

Un modèle de fluide ayant une résolution inférieure à la réalité, l'analyse ne pourra jamais être parfaitement réaliste. Dans la présentation des algorithmes d'analyse on parlera souvent de *l'état réel* du modèle. Cette expression fait référence à l'état du modèle qui est la meilleure représentation possible de la réalité à un instant donné, c'est ce que l'on cherche à estimer lors du processus d'analyse. Une observation sans erreur de mesure de cet état réel de modèle ne pourra généralement pas être identique à l'observation équivalente de la réalité, à cause du réalisme limité de la discrétisation du modèle. C'est ce que l'on appelle *l'erreur de représentativité*, conventionnellement incluse dans l'erreur d'observation, mais il faut se souvenir que c'est autant une propriété du modèle lui-même que de l'observation (par exemple, une station météorologique est généralement installée dans un pré, elle ne peut pas être représentative des détails de l'atmosphère près des arbres, rochers ou des bâtiments alentours).

## 1.2 Méthodes de base d'analyse de données

Fondamentalement, l'analyse de données consiste à déduire une estimation *satisfaisante* de l'état d'un modèle à partir d'observation. Dans le cas trivial où l'on dispose d'excellentes observations de toutes les variables du modèle, il suffit de substituer les valeurs observées dans l'état du modèle pour produire une analyse. Plus souvent, les observations ne coïncident pas avec cet état de modèle dans l'espace ou dans le temps. Si l'on définit le caractère *satisfaisant* de l'analyse simplement par (1) être proche des valeurs observées lorsqu'on en est spatialement ou temporellement proche, et (2) être continue dans l'espace et/ou le temps, le problème d'analyse se réduit à celui *d'interpolation*.

Différentes techniques d'interpolation plus ou moins simples existent, leur choix dépend étroitement des propriétés physiques ou mathématiques exigées de l'analyse, par exemple :

- voisin le plus proche (nearest neighbour) : pour chaque variable du modèle, trouver



l'observation la plus proche (au sens d'une métrique appropriée), et prendre sa valeur. Ceci produit une analyse généralement discontinue.

- interpolation linéaire : pour chaque variable du modèle, trouver les observations qui l'encadrent (au sens d'une topologie appropriée), construire une représentation localement linéaire du champ en cause, et prendre sa valeur au point de grille considéré. Par exemple, en 1D, pour un point d'abscisse  $x$ , trouver 2 observations  $(y_1, y_2)$  d'abscisses respectives  $(x_1, x_2)$  telles que  $x_1 \leq x < x_2$ , et prendre pour analyse en  $x$  la valeur  $y = (1 - \lambda)y_1 + \lambda y_2$  avec  $\lambda = (x - x_1)/(x_2 - x_1)$
- fonctions interpolantes : c'est une généralisation de l'interpolation linéaire : il faut se donner un espace de fonctions  $y_a(x)$  où  $x$  est la position,  $y$  est la valeur du champ,  $a$  un vecteur de coefficients dont dépend la fonction  $y$ . Pour chaque point  $x$  à analyser, il faut trouver  $n$  observations  $(x_i, y_i)$  dans un voisinage à définir de  $x$ , telles que  $y_a(x_i) = y_i$  pour tout  $i$ , et que l'ensemble de ces dernière équations en  $a$  admette une solution unique. Cette solution définit localement une fonction interpolante  $y_a$ , qui définit l'analyse par  $y = y_a(x)$ . Il est courant de prendre pour  $y_a$  des famille de polynômes, des sinus/cosinus, etc. les fonctions interpolantes les plus utiles sont les *splines*, qui fournissent avec un minimum de calcul une analyse avec un degré de régularité prédéfini (continuité des 1e et 2e dérivées, etc), et qui existent en une ou plusieurs dimensions.

Il faut se méfier des problèmes d'interpolation en ce sens qu'ils sont souvent d'une simplicité trompeuse. Par exemple, dans l'interpolation linéaire, il faut bien veiller au respect des inégalités pour le choix des observations, afin que la division soit toujours définie. Dans l'interpolation avec des fonctions polynômiales, rien ne garantit a priori que les extrema du champ analysé seront identiques aux extrema des valeurs observées, ce qui peut produire des résultats surprenant pour analyser des champs de concentrations chimiques par exemple, pour lesquels une valeur négative n'a pas de sens physique. On peut forcer l'interpolation à respecter de telles contraintes physiques, mais avec des précautions spéciales dans l'algorithme utilisé.

Un autre aspect subtil est la représentativité physique des valeurs interpolées : prenons le cas apparemment simple de l'interpolation linéaire d'observations sur la grille d'un modèle. Si la densité spatiale des observations est localement très inférieure à celle de la grille du modèle, l'analyse sera linéaire par morceaux, avec des points anguleux au voisinage de chaque observation, ce qui n'est pas forcément très physique, et peut être problématique si l'on veut ensuite calculer la dérivée seconde du champ analysé, puisqu'elle est localement très grande (par exemple pour calculer le champ de tourbillon à partir d'une analyse du champ de pression avec l'hypothèse d'équilibre géostrophique). Les points de grille situés très loin de toute observation seront analysés par moyenne d'observations distantes, ce qui peut être assez artificiel s'il n'y a pas de relation physique réelle entre des points distants.

Inversement, si la densité des observations est très supérieure à celle de la grille, chaque point de l'analyse résultera uniquement de l'interpolation de points d'observation très proches, donc sera très peu représentatif de l'ensemble des valeurs observées dans le voisinage ; les observations situées loin de tout point de grille (et il peut y en avoir beaucoup) ne seront pas du tout utilisées, bref l'analyse obtenue reflètera très mal la distribution des observations. Il peut être préférable de définir l'analyse comme la moyenne de toutes les observations situées plus près du point considéré que de tout autre point de grille.

En conclusion, si la densité relative des observations et de la grille modèle sont très va-

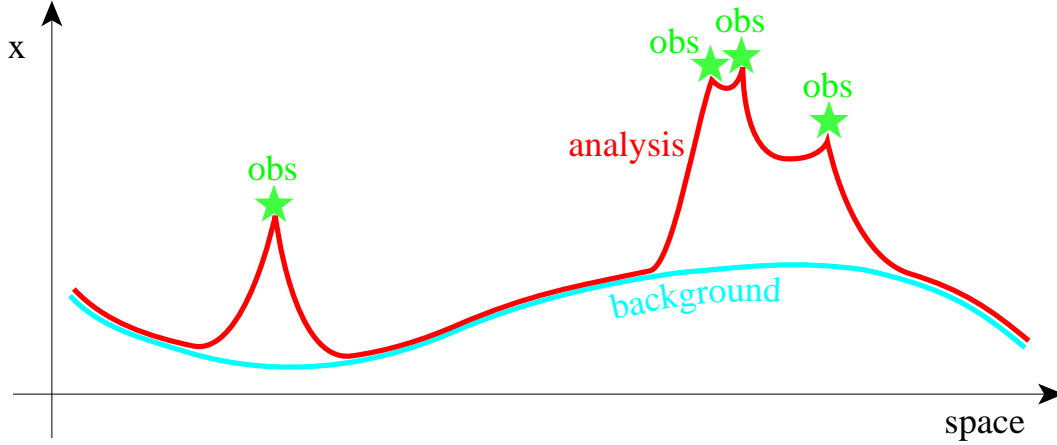


FIG. 3 – Exemple d’analyse de Cressman d’un champ unidimensionnel. Le champ d’ébauche  $\mathbf{x}_b$  est représenté par la courbe grise, les observations en vert. L’analyse (courbe noire) est produite par interpolation entre l’ébauche et les observations, au voisinage de chaque observation : plus elle est proche, plus son poids est élevé.

riables, il est difficile de définir un algorithme d’interpolation général et physiquement satisfaisant.

Lorsque l’analyse fait partie d’une assimilation séquentielle, on dispose à chaque instant d’une *ébauche* (background ou first guess), ce qui permet de résoudre une partie des problèmes ci-dessus en travaillant pas corrections (ou incréments) successives. Par exemple, avec l’algorithme de Cressman : Pour un modèle monovarié en points de grille, si l’on note  $\mathbf{x}_b$  l’ébauche,  $\mathbf{y}(i)$  un ensemble  $i = 1 \dots n$  d’observations du paramètre dont le champ est modélisé, un algorithme simple d’analyse de Cressman consiste à définir l’état du modèle  $\mathbf{x}_a$  en chaque point de grille  $j$  selon l’équation de mise à jour suivante :

$$\mathbf{x}_a(j) = \mathbf{x}_b(j) + \frac{\sum_{i=1}^n w(i, j)[\mathbf{y}(i) - \mathbf{x}_b(i)]}{\sum_{i=1}^n w(i, i)}$$

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

où  $d_{i,j}$  mesure la distance entre les points  $i$  et  $j$ .  $\mathbf{x}_b(i)$  est l’ébauche interpolée au point  $i$ . La *fonction de poids*  $w(i, j)$  si le point de grille  $j$  est confondu avec l’observation  $i$ . C’est une fonction décroissante de la distance si  $d_{i,j} < R$ , où  $R$  est une constante donnée (le rayon d’influence) au-delà de laquelle les observations n’ont pas de poids.

Il existe de nombreuses variantes de la méthode de Cressman. On peut redéfinir les fonctions de poids, par exemple  $\exp(-d_{i,j}^2/2R^2)$ . Les méthodes de *corrections successives* et de *nudging*<sup>2</sup> sont très semblables, les principales différences étant (1) que les poids lorsque  $i = j$  peuvent être inférieurs à 1, ce qui permet de doser le poids relatif des observations par rapport à l’ébauche, et (2) la dimension temporelle de l’assimilation est prise en compte, par une relaxation progressive du modèle vers les valeurs observées, ce qui produit une assimilation temporellement lisse. En théorie ces techniques peuvent être aussi performantes que n’importe

<sup>2</sup>en anglais, to nudge signifie ‘pousser en donnant un petit coup’, ou ‘donner du coude’.

que les méthodes décrites ci-après, malheureusement en pratique on dispose rarement d'un cadre théorique permettant de spécifier les poids des corrections de manière optimale : ces méthodes se caractérisent en fait par leur empirisme.

ref : Daley 1991.

Une autre classe de techniques d'analyses, très utilisées dans divers domaines d'ingénierie et notamment la géologie, est le *krigeage* (kriging en anglais). Le coeur de l'algorithme est identique à l'analyse de Cressman, mais la fonction de poids  $w$  est spécifiée à partir d'une étude préalable des propriétés statistiques du champ à analyser, l'idée étant que les erreurs existant dans l'ébauche ont toutes les chances d'avoir des structures proches des erreurs moyennes rencontrées dans l'histoire passée. Concrètement, la fonction  $w$  est spécifiée comme fonction d'autocovariance des erreurs, mesurées par l'écart moyen entre observations et ébauches :

$$w(i, j) = A \frac{1}{n} \sum_{(i,j)} [y(i) - \mathbf{x}_b(i)][y(j) - \mathbf{x}_b(j)]$$

où  $A$  est une constante de normalisation et  $n$  le nombre de couples  $(i, j)$  utilisés pour calculer  $w$ . En pratique, pour obtenir des statistiques stables, on est conduit à (a) accumuler la somme sur plusieurs cas d'analyse indépendants, (b) évaluer  $w$  comme un histogramme, donc en définissant des classes de  $i$  et  $j$  plutôt que toutes les combinaisons possibles, (c) faire l'hypothèse d'homogénéité des statistiques en ne faisant dépendre  $w$  que du vecteur de séparation  $i - j$  (dans ce cas  $w$  s'appelle un *variogramme*), voire (d) faire en plus l'hypothèse d'isotropie des statistiques, autrement dit que  $w$  ne dépend que de la distance entre  $i$  et  $j$ . Il existe plusieurs variantes de krigeages, certaines sont disponibles sous forme de logiciels en 'boîte noire'. La vertu du krigeage est d'adapter automatiquement les fonctions d'interpolation aux propriétés physiques du problème considéré.

Dans la suite, on va décrire une classe encore plus puissante d'algorithmes, qui sont, comme le krigeage, des *interpolations statistiques*.

### 1.3 Les avantages d'une approche statistique

Les techniques ci-dessus sont assez simples à mettre en oeuvre, mais sont souvent vulnérables aux problèmes suivants :

- si l'ébauche de l'analyse est déjà assez bonne qualité, on voudra éviter de la dégrader avec des observations entachées d'erreurs.
- on ne sait pas en général comment sélectionner la fonction  $w$ , autrement dit comment relaxer l'analyse vers l'ébauche lorsque l'on s'éloigne des observations. Ni comment spécifier  $w$  lorsque le modèle comporte plusieurs variables.
- une analyse doit idéalement respecter les propriétés fondamentales connues du système analysé : caractère lisse des champs, relations (*balance* en anglais) entre variables (équilibre hydrostatique, quasi-équilibre géostrophique, contraintes de positivité ou de non-sursaturation), etc. En effet, une analyse ne respectant pas ces propriétés n'a aucune chance d'être réaliste.

Ces contraintes sont pratiquement impossibles à respecter avec les approches empiriques décrites précédemment. Heureusement, un cadre théorique assez général existe, qui part des

idées suivantes, familières de ceux qui ont pratiqué les analyses à la main :

1. partir d'une ébauche de la meilleure qualité possible, par exemple une prévision récente qui fournisse une vue d'ensemble de l'état vraisemblable du système avant analyse,
2. faire une moyenne des observations si elles sont proches les unes des autres, et trouver un compromis entre les valeurs de l'ébauche et les valeurs observées, l'analyse devant respecter surtout les sources d'information réputées les plus fiables.
3. l'analyse doit être spatialement lisse comme la nature, lorsque l'on s'éloigne d'une observation, la relaxation vers l'ébauche doit se faire aux échelles de variabilité les plus fréquentes dans le système physique considéré.
4. l'analyse doit respecter autant que possible les propriétés les plus fréquemment respectées par le système physique, tout en autorisant des déviations par rapport à ces propriétés, si c'est ce qu'indiquent les observations.

En résumé, les informations à utiliser pour faire l'analyse sont l'ébauche, les observations, et les propriétés physiques connues du système, mais ni l'ébauche ni les observations ne sont totalement fiables, pour être la plus proche possible de la réalité l'analyse devra donc être une moyenne pondérée entre toutes ces informations. Il faut donc adopter un formalisme probabiliste. En pratique on *définira* l'analyse comme la solution d'un problème d'optimisation où l'on minimise l'erreur d'analyse moyenne (en un sens cohérent avec ce qui intéresse l'utilisateur), sous contrainte des informations disponibles.

ref : Lorenc 1986.

## 2 Vecteur d'état, espaces de contrôle et d'observation

### 2.1 Vecteur d'état

Il est important de commencer par bien définir l'espace de travail. En assimilation comme dans un modèle de prévision, la liste des nombres qui représentent l'état du modèle est une matrice colonne appelée *vecteur d'état*  $\mathbf{x}$ . La relation entre la réalité et les composantes de ce vecteur dépend du choix de la discrétisation employée, et de la base utilisée. Comme expliqué plus haut, on appelle état réel  $\mathbf{x}_t$  la valeur du vecteur d'état qui représente le mieux possible la réalité. L'ébauche (background, ou first guess) et l'analyse sont les valeurs du vecteur d'état notées  $\mathbf{x}_b$  et  $\mathbf{x}_a$ , respectivement.

### 2.2 Variable de contrôle

On ne souhaite pas forcément résoudre le problème d'analyse dans l'espace complet du vecteur d'état. Par exemple si certains paramètres du modèle sont difficiles à estimer de manière cohérente avec les autres paramètres (par exemple, les nuages ou la vitesse verticale dans un modèle atmosphérique sont bien plus délicats à estimer que le vent horizontal, la température et l'humidité de l'air). Ou simplement si les ressources informatiques disponibles obligent à réduire le domaine ou la résolution d'analyse. Dans ces cas, on préférera formuler le problème d'analyse comme la recherche d'une correction, ou *incrément*, de l'ébauche, cette correction  $\delta\mathbf{x}$  appartenant à un sous-espace appelé *espace de contrôle* :

$$\mathbf{x}_a = \mathbf{x}_b + \delta\mathbf{x}$$

On cherche donc la *variable de contrôle*  $\delta\mathbf{x}$  telle que  $\mathbf{x}_a$  soit aussi proche que possible de  $\mathbf{x}_t$ .

### 2.3 Observations

Le calcul d'une analyse utilise un ensemble de valeurs observées présentées dans un *vecteur d'observation*  $\mathbf{y}$ , qu'il faut savoir comparer au vecteur d'état. L'idéal serait d'observer chaque composante de la variable de contrôle : il suffirait alors de minimiser la distance entre  $\mathbf{x}_a$  et  $\mathbf{y}$ . En général les observations ne sont pas des paramètres du modèle, elles ne sont pas disposées sur la même grille, et tous les paramètres de l'espace de contrôle ne sont pas observés. La manière la plus commode d'effectuer la comparaison obs/modèle consiste à définir une application du modèle vers les observations appelée *opérateur d'observation*  $H$  (observation operator ou forward operator en anglais). Cet opérateur *simule* les valeurs  $H(\mathbf{x})$  qu'on observerait si le modèle était parfait, appelées *équivalents modèle* des observations. Un opérateur d'observation typique est, pour une mesure ponctuelle d'un paramètre du modèle en un point quelconque, l'interpolation des champs du modèle au point de mesure. Si  $H$  est linéaire, chacune de ses lignes correspond à une observation scalaire, et ses coefficients sont les poids de l'interpolation. Le nombre d'observations  $\dim \mathbf{y}$  est généralement différent, et souvent inférieur, au nombre de degrés de liberté du problème d'analyse  $\dim \delta\mathbf{x}$ , ce qui en fait un problème sous-déterminé en l'absence d'autres informations que celle observées à l'instant d'analyse.

Le problème de l'analyse consiste à utiliser les écarts entre observations et équivalents modèles :

$$\mathbf{y} - H(\mathbf{x})$$

qui est un vecteur appartenant à l'espace des observations.  $\mathbf{y} - H(\mathbf{x}_b)$  s'appelle *innovations*,  $\mathbf{y} - H(\mathbf{x}_a)$  s'appelle *résidus*. L'étude des innovations et résidus apporte des informations importantes sur les performances d'un système d'analyse ou d'assimilation de données.

### 3 La modélisation statistique des erreurs

Afin de prendre en compte l'incertitude liée à la présence d'erreurs inconnues dans les observations, l'ébauche et l'analyse, nous allons modéliser les caractéristiques de ces erreurs par des *fonctions de densité de probabilité* (probability density function, ou pdf), en prenant pour exemple les erreurs d'ébauche. (Une présentation rigoureuse ferait appel à la théorie des probabilités, ici on s'en tient à une présentation simplifiée).

Pour une ébauche  $\mathbf{x}_b$ , donnée en entrée d'une analyse, le vecteur qui la sépare de l'état réel est

$$\varepsilon_b = \mathbf{x}_b - \mathbf{x}_t$$

Si l'on répétait l'expérience d'analyse plusieurs fois sous les mêmes conditions, donc avec différentes réalisations des erreurs produites par des processus inconnus,  $\varepsilon_b$  serait différent à chaque fois. Pour un grand nombre de réalisations, les statistiques que l'on peut calculer sur  $\varepsilon$  convergent en pratique vers des valeurs qui ne dépendent que des conditions physiques de l'expérience, pas des réalisations particulières observées. Au moment de faire une analyse particulière, on ne peut pas savoir quelles erreurs exactes se produisent (par définition même d'une "erreur"), mais on peut connaître certaines statistiques de la pdf associée, notamment la moyenne et les covariances. On rappelle la définition des statistiques indispensables à bien connaître : pour un ensemble de  $i = 1 \dots n$  réalisations de variables scalaires  $u$  et  $v$ ,

**moyenne** : la moyenne de  $u$  est  $\bar{u} = 1/n \sum_{i=1}^n u_i$

**variance** : la variance de  $u$  est  $var(u) = 1/n \sum_{i=1}^n (u_i - \bar{u})^2$

**écart type** : l'écart type de  $u$  est  $\sigma(u) = \sqrt{var(u)}$  (standard deviation, or root mean square (rms) deviation)

**covariance** : la covariance entre  $u$  et  $v$  est  $cov(u, v) = 1/n \sum_{i=1}^n (u_i - \bar{u})(v_i - \bar{v})$

**corrélation** : lorsqu'elle existe, la corrélation entre  $u$  et  $v$  est  $\rho(u, v) = cov(u, v) / [\sigma(u)\sigma(v)]$

Les notions de moyenne et covariance se généralisent à des variables vectorielles : si  $\mathbf{u}$  est un vecteur,  $\bar{\mathbf{u}}$  est un vecteur dont les composantes sont les moyennes des composantes de  $\mathbf{u}$ , la covariance de  $\mathbf{u}$  est la matrice carrée

$$cov(\mathbf{u}) = \overline{(\mathbf{u} - \bar{\mathbf{u}})(\mathbf{u} - \bar{\mathbf{u}})^T}$$

dont les éléments sont des covariances (variances, sur la diagonale) entre tous les couples de composantes de  $\mathbf{u}$ . Si toutes ces composantes sont de variance non nulle, la matrice de covariance est symétrique, définie positive, et donc inversible. Diviser les lignes et les colonnes d'une matrice de covariance par le vecteur des variances produit une matrice de corrélations entre les composantes de  $\mathbf{u}$ .

#### 3.1 Modèles d'erreur

Les erreurs dont la modélisation est nécessaire pour la plupart des techniques d'assimilation sont les suivantes :

**erreurs d'ébauche** :  $\varepsilon_b = \mathbf{x}_b - \mathbf{x}_t$ , de moyenne  $\bar{\varepsilon}_b$  and de covariances  $\mathbf{B} = \overline{(\varepsilon_b - \bar{\varepsilon}_b)(\varepsilon_b - \bar{\varepsilon}_b)^T}$ .

Ce sont les erreurs d'estimation de l'ébauche, autrement dit la différence entre  $\mathbf{x}_b$  et l'état réel.

**erreurs d'observation :**  $\varepsilon_o = \mathbf{y} - H(\mathbf{x}_t)$ , de moyenne  $\overline{\varepsilon_o}$  et de covariances  $\mathbf{R} = \overline{(\varepsilon_o - \overline{\varepsilon_o})(\varepsilon_o - \overline{\varepsilon_o})^T}$ . Elles contiennent les erreurs du processus de mesure lui-même (erreurs instrumentales), les erreurs dans la modélisation de l'opérateur  $H$ , et les erreurs de représentativité, autrement dit les erreurs de discrétisation qui empêchent  $\mathbf{x}_t$  d'être une représentation parfaite de la réalité (notamment la variabilité du système réel non résolue par la grille du modèle)

**erreurs d'analyse :**  $\varepsilon_a = \mathbf{x}_a - \mathbf{x}_t$ , de moyenne  $\overline{\varepsilon_a}$ . Ce sont les erreurs d'estimation de l'analyse, que l'on cherche à minimiser, et dont une mesure de l'amplitude  $\|\varepsilon_a - \overline{\varepsilon_a}\|$  est fournie par la trace de la matrice de covariances d'erreur  $\mathbf{A}$ ,  $\text{Tr}(\mathbf{A}) = \|\varepsilon_a - \overline{\varepsilon_a}\|^2$ .

**erreurs de modélisation :** cf. la section sur le filtre de Kalman pour la description mathématique de ces erreurs. Elles interviennent dans des algorithmes d'assimilation avancés, et leurs caractéristiques sont encore mal comprises. En général, on les modélise comme les erreurs d'ébauche, à un facteur multiplicatif près.

Les moyennes de ces erreurs sont appelées *biais*, et résultent d'un défaut systématique dans le modèle, les observations, ou la manière de les utiliser.

### 3.2 Estimation pratique des covariances

Les statistiques d'erreur (moyennes et covariances utilisées dans les algorithmes qui suivent) représentent notre connaissance des incertitudes sur les données en entrée du processus d'analyse. Par définition, la valeur précise de ces erreurs à un instant donné est inconnue et inaccessible (si elles étaient connues il suffirait de les retrancher des données pour s'en débarrasser). En pratique, pour obtenir les statistiques voulues il existe trois types de techniques :

**l'hypothèse d'ergodicité** consiste à supposer que les erreurs conservent les mêmes statistiques pendant une certaine période (hypothèse de stationnarité) et/ou dans une certaine zone (hypothèse d'homogénéité). En effectuant des statistiques sur cette période et/ou zone on suppose donc que l'on a calculé les statistiques d'erreur correctes pour effectuer une analyse particulière. En un sens cela revient à faire une climatologie des erreurs. L'hypothèse d'ergodicité revient à dire que certaines caractéristiques des processus physiques générateurs d'erreur sont constantes. C'est traditionnellement le cas lorsqu'on étudie des fluides turbulents par exemple. Dans un système géophysique, cela revient à considérer que les sources de bruits, d'instabilité, les stations de mesure, etc, sont à peu près uniformément répartis. C'est généralement une hypothèse utile pour travailler en première approximation, mais qu'il faut savoir critiquer.

**les techniques d'ensemble** reviennent à effectuer un échantillonnage discret des sources d'erreurs connues, sur un grand nombre de réalisations simulées du système physique étudié, et à calculer les statistiques sur les résultats de ces simulations. C'est un peu plus réaliste que l'hypothèse d'ergodicité dans la mesure où l'on peut prendre en compte des propriétés physiques dans le calcul des erreurs. Par exemple, pour calculer des statistiques d'erreur d'ébauche dans une assimilation, on peut prendre en compte la maturation des erreurs au cours de la prévision qui a produit cette ébauche :

1. se donner un modèle statistique simple des erreurs d'estimation de l'analyse précédente,
2. tirer aléatoirement  $n$  réalisations de ces erreurs,
3. perturber l'analyse précédente en y ajoutant ces  $n$  erreurs possibles,



4. effectuer  $n$  prévisions à partir de ces analyses perturbées, (on peut aussi perturber aléatoirement les équations d'évolution du modèle lui-même),
5. calculer les statistiques d'erreur de l'ébauche sur cet ensemble de  $n$  prévisions.

On peut procéder de même en perturbant l'entrée d'un opérateur d'observation pour simuler des statistiques d'erreur d'observation, ou en perturbant les observations elles-mêmes pour simuler des erreurs d'analyse. C'est une technique commode en pratique, mais qui peut avoir des effets pervers si le modèle d'évolution lui-même est imparfait : par exemple, s'il sous-estime systématiquement la variabilité d'un paramètre donnée du système, les statistiques d'erreur ainsi simulées auront une faible dispersion, ce qui sera interprété par l'analyse comme une faible incertitude sur l'estimation de ce paramètre...qui ne sera donc jamais corrigé.

**les techniques stochastiques** consistent à fabriquer un algorithme d'estimation directe des statistiques d'erreur, avec un modèle d'évolution, des étapes de correction par les observations...de telles techniques sont très employées en finance, par exemple. L'une des plus simples est le *filtre de Kalman*, qui sera exposé ci-après.

En pratique ces méthodes permettent d'obtenir des informations sur les erreurs, mais pas de spécifier tous les paramètres voulus : pour un modèle de dimension  $n$ , la matrice de covariances d'erreurs d'ébauche comporte  $n(n+1)/2$  coefficients a priori indépendants et le biais en comporte  $n$ . Avec un modèle ou un jeu d'observation de grande dimension il faut donc réduire le nombre de degrés de liberté en supposant a priori que les erreurs obéissent à des modèles statistiques.

Exemples de techniques simples couramment employées :

**colocation** : la différence entre deux observations indépendantes en un point donné a pour variance la somme des variances d'erreur de ces deux observations. Si l'on en connaît une (par ex. par calibration industrielle de l'instrument), la différence fournit l'autre.

**validation croisée (cross-validation)** : effectuer deux analyses avec deux sous-ensembles indépendants de l'ensemble des observations. Les différences entre ces analyses sont une estimation des erreurs d'analyse.

**méthode observationnelle (Hollingsworth-Lönnberg method)** : comparer des ébauches à des observations, calculer les variogrammes de ces erreurs en fonction de la distance. En supposant que les erreurs d'observation n'ont pas de corrélation spatiale, les covariances calculées avec des séparations non nulles sont des covariances d'erreurs d'ébauche. En supposant que ces dernières sont de dérivée continue dans l'espace, leur prolongement par continuité pour une distance nulle fournit par différence la variance d'erreur d'observation

**covariances gaussiennes homogènes isotropes** : dans un système 2D au comportement raisonnablement uniforme, on suppose fréquemment que les corrélations d'erreur  $y$  sont de la forme  $\exp(-r^2/a^2)$  où  $r$  est la distance et  $a$  une constante (distance caractéristique). Il ne reste alors plus qu'à estimer  $a$  et le champ de variances d'erreurs. Plus généralement, pour un modèle homogène isotrope, la distance à laquelle les corrélations sont inférieures à un certain seuil (typiquement 0.5) est appelée *portée spatiale* des corrélations.

**méthode du NMC** : (nommée pour des raisons historiques, NMC est l'ancien nom du service météo national des USA). Dans un système de prévision, les différences entre couples

de prévisions produites consécutives, et valables à un même instant, ont des corrélations de différences proches des corrélations d'erreur de prévision (par ex. les prévisions faites hier et avant-hier et valables aujourd'hui à 12h).

ref : Hollingsworth et al 1986, Parrish and Derber 1992.

## 4 Interpolation statistique aux moindres carrés, ou BLUE

Dans cette section on présente l'équation fondamentale des techniques linéaires d'analyse objective, appelée BLUE (Best linear unbiased estimator, meilleur estimateur linéaire non-biaisé). C'est fondamentalement une simple généralisation multidimensionnelle de la régression linéaire : on cherche à prédire une correction linéaire de l'état du modèle à partir des observations. Le BLUE est à la base de la plupart des techniques d'analyses opérationnelles modernes en météorologie et océanographie.

### 4.1 Notations et hypothèses

La variable d'état du modèle est de dimension  $n$ , celle du vecteur d'observation est  $p$ . On note :

$\mathbf{x}_t$  état réel du modèle (de dimension  $n$ )

$\mathbf{x}_b$  ébauche du modèle (dimension  $n$ )

$\mathbf{x}_a$  analyse du modèle (dimension  $n$ )

$\mathbf{y}$  vecteur des observations (dimension  $p$ )

$H$  opérateur d'observation (va de la dimension  $p$  à  $n$ )

$\mathbf{B}$  matrice de covariances d'erreurs d'ébauche ( $\mathbf{x}_b - \mathbf{x}_t$ ) (dimension  $n \times n$ )

$\mathbf{R}$  matrice de covariance d'erreurs d'observation ( $\mathbf{y} - H[\mathbf{x}_t]$ ) (dimension  $p \times p$ )

$\mathbf{A}$  matrice de covariance d'erreurs d'analyse ( $\mathbf{x}_a - \mathbf{x}_t$ ) (dimension  $n \times n$ )

On fait les hypothèses suivantes :

**linéarisation de l'opérateur d'observation :** les variations de l'opérateur d'observation au voisinage de l'ébauche sont supposées linéaires : pour tout  $\mathbf{x}$  assez 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 :**  $\mathbf{B}$  et  $\mathbf{R}$  sont des matrices définies positives.

**erreurs non biaisées :** l'espérance des erreurs d'observation et d'ébauche est nulle :  $\overline{\mathbf{x}_b - \mathbf{x}_t} = \overline{\mathbf{y} - H(\mathbf{x}_t)} = 0$

**Erreurs non corrélées :** les erreurs d'observation et d'ébauche ont des corrélations mutuelles nulles :  $\overline{(\mathbf{x}_b - \mathbf{x}_t)(\mathbf{y} - H[\mathbf{y}_t])^T} = 0$

**Analyse linéaire :** on cherche une analyse définie par des corrections de l'ébauche en fonction linéaire des écarts entre observations et ébauche.

**Analyse optimale :** on cherche un état analysé aussi proche que possible de l'état réel au sens de la moyenne quadratique (autrement dit, une estimation de variance d'erreur minimale).

ref : Daley 1991, Lorenc 1986, Ghil 1989.

## 4.2 Théorème : équations d'analyse aux moindres carrés

(a) L' *estimateur linéaire optimal*, ou *analyse BLUE*, est défini par les équations d'interpolation suivantes :

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - H[\mathbf{x}_b]) \quad (\text{A1})$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \quad (\text{A2})$$

dans lesquelles l'opérateur  $\mathbf{K}$  s'appelle le *gain*, où *matrice des poids*, de l'analyse.

(b) La *matrice de covariances d'erreur d'analyse* est, pour tout  $\mathbf{K}$  :

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \quad (\text{A3})$$

Si de plus  $\mathbf{K}$  est le gain optimal au sens des moindres carrés, cette matrice vaut :

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} \quad (\text{A4})$$

(c) L'analyse BLUE optimale est obtenue de manière équivalente par solution du *problème d'optimisation variationnelle* suivant :

$$\begin{aligned} \mathbf{x}_a &= \text{Arg min } J \\ J(\mathbf{x}) &= (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - H[\mathbf{x}])^T \mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}]) \\ &= J_b(\mathbf{x}) + J_o(\mathbf{x}) \end{aligned} \quad (\text{A5})$$

où  $J$  s'appelle la *fonction coût* de l'analysis (ou *norme des résidus* (misfit), où *fonction de pénalité*),  $J_b$  est le *terme d'ébauche*,  $J_o$  est le *terme d'observation*.

(d) L'analyse  $\mathbf{x}_a$  est *optimale* : c'est la plus proche de l'état réel  $\mathbf{x}_t$  au sens de la norme quadratique.

(e) Si les pdfs d'erreurs d'ébauche et d'observation sont gaussiennes, alors  $\mathbf{x}_a$  est aussi un *estimateur de probabilité maximale* (maximum likelihood estimator) de  $\mathbf{x}_t$ .

### Démonstration:

Par translation de  $\mathbf{x}$  par  $\mathbf{x}_b$ , on peut supposer que  $H = \mathbf{H}$  donc que l'opérateur d'observation est ici linéaire. L'équation (A1) exprime mathématiquement que l'on cherche une analyse en fonction linéaire des écarts aux observations. L'expression de  $\mathbf{K}$  dans (A2) est bien posée parce que  $\mathbf{R}$  est une matrice définie positive, et  $\mathbf{H}\mathbf{B}\mathbf{H}^T$  est positive. Le problème de minimisation (A5) est bien posé parce que  $J_o$  est une fonction convexe et  $J_b$  est une fonction strictement convexe (c'est une forme quadratique).

L'équivalence entre les alinéas (a) et (c) du théorème provient de ce que le gradient de  $J$  est nul à l'optimum  $\mathbf{x}_a$  :

$$\begin{aligned} \nabla J(\mathbf{x}_a) = 0 &= 2\mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_b) - 2\mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}_a]) \\ 0 &= \mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_b) - \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}_b]) - \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}(\mathbf{x}_a - \mathbf{x}_b) \\ (\mathbf{x}_a - \mathbf{x}_b) &= (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}_b]) \end{aligned}$$

L'équivalence avec (A2) est facile (toutes les matrices inversées sont définies positives) :

$$\mathbf{H}^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}) = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})\mathbf{B}\mathbf{H}^T$$

$$\begin{aligned} &= \mathbf{H}^T + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{B} \mathbf{H}^T \\ \text{d'où } (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} &= \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} \end{aligned}$$

Les expressions (A3) et (A4) pour  $\mathbf{A}$  s'obtiennent en réécrivant l'équation d'analyse (A1) à partir des erreurs d'ébauche, d'analyse et d'observation : are obtained by rewriting the analysis

$$\begin{aligned} \varepsilon_b &= \mathbf{x}_b - \mathbf{x}_t \\ \varepsilon_a &= \mathbf{x}_a - \mathbf{x}_t \\ \varepsilon_o &= \mathbf{y} - H[\mathbf{x}_t] \\ \varepsilon_a - \varepsilon_b &= \mathbf{K}(\varepsilon_o - \mathbf{H}\varepsilon_b) \\ \varepsilon_a &= (\mathbf{I} - \mathbf{K}\mathbf{H})\varepsilon_b + \mathbf{K}\varepsilon_o \end{aligned}$$

En développant l'expression  $\varepsilon_a \varepsilon_a^T$  et en prenant sa moyenne, par linéarité de l'opérateur de moyenne on trouve l'expression générale (A3) ( $\varepsilon_b$  et  $\varepsilon_o$  sont non corrélés, donc leur covariance croisée est nulle). La forme (A4) s'obtient en substituant l'expression du  $\mathbf{K}$  optimal et en simplifiant les termes qui s'annulent.

Finalement pour prouver (A2) il faut prendre la matrice de covariance d'erreurs d'analyse (A3) et minimiser sa trace, c'est à dire la variance d'erreur totale : (noter que  $\mathbf{B}^T = \mathbf{B}$  et  $\mathbf{R}^T = \mathbf{R}$ )

$$\text{Tr}(\mathbf{A}) = \text{Tr}(\mathbf{B}) + \text{Tr}(\mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{K}^T) - 2\text{Tr}(\mathbf{B}\mathbf{H}^T\mathbf{K}^T) + \text{Tr}(\mathbf{K}\mathbf{R}\mathbf{K}^T)$$

C'est une fonction scalaire, continue et différentiable des coefficients de la matrice  $\mathbf{K}$ , et l'on trouve sa différentielle  $d_{\mathbf{K}}$  en prenant les termes au premier ordre par rapport à  $\mathbf{K}$  de la différence  $\text{Tr}(\mathbf{A})(\mathbf{K} + \mathbf{L}) - \text{Tr}(\mathbf{A})(\mathbf{K})$ , où  $\mathbf{L}$  est une matrice de perturbation arbitraire :

$$\begin{aligned} d_{\mathbf{K}}[\text{Tr}(\mathbf{A})]\mathbf{L} &= 2\text{Tr}(\mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{L}^T) - 2\text{Tr}(\mathbf{B}\mathbf{H}^T\mathbf{L}^T) + 2\text{Tr}(\mathbf{K}\mathbf{R}\mathbf{L}^T) \\ &= 2\text{Tr}(\mathbf{K}\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{L}^T - \mathbf{B}\mathbf{H}^T\mathbf{L}^T + \mathbf{K}\mathbf{R}\mathbf{L}^T) \\ &= 2\text{Tr}\{ [\mathbf{K}(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}) - \mathbf{B}\mathbf{H}^T]\mathbf{L}^T \} \end{aligned}$$

La dernière ligne montre que la dérivée est nulle quelle que soit  $\mathbf{L}$  si et seulement si  $(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})\mathbf{K}^T - \mathbf{H}\mathbf{B} = 0$ , ce qui est équivalent à

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

puisque  $(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})$  est supposé inversible.

Si les pdfs en jeu sont toutes gaussiennes, on peut modéliser les pdfs d'ébauche, d'observation et d'analyse respectivement par : ( $b$ ,  $o$  et  $a$  sont des facteurs de normalisation.)

$$\begin{aligned} \mathcal{P}_b(\mathbf{x}) &= b \exp \left[ \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) \right] \\ \mathcal{P}_o(\mathbf{x}) &= o \exp \left[ \frac{1}{2}(\mathbf{y} - H[\mathbf{x}])^T \mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}]) \right] \\ \mathcal{P}_a(\mathbf{x}) &= \mathcal{P}_b(\mathbf{x})\mathcal{P}_o(\mathbf{x})j \end{aligned}$$

Ce qui fournit les moyennes et covariances correctes pour les erreurs d'ébauche et d'observation, la pdf d'erreurs d'analyse est tout simplement le produit bayésien des deux sources d'information, les pdfs d'ébauche et d'observations. Cela peut s'écrire rigoureusement avec le théorème de Bayes, en formulant  $\mathcal{P}_a$  comme probabilité conditionnelle de  $\mathbf{x}$  étant données les observations et la pdf a priori fournie en ébauche). En prenant le logarithme de  $\mathcal{P}_a(\mathbf{x})$ , on trouve que l'état du modèle possédant la probabilité la plus élevée est celui qui minimise la fonction-coût  $J(\mathbf{x})$  écrite dans le théorème.

## 4.3 Commentaires

### 4.3.1 Cas de dégénérescence du BLUE

Les hypothèses de non-trivialité sont toujours vérifiées dans les problèmes d'analyse bien posés : si  $\mathbf{B}$  n'est pas définie positive, on peut restreindre l'espace de contrôle à l'orthogonal du noyau de  $\mathbf{B}$ , auquel cas l'analyse n'effectuera pas de correction des variables qui sont supposées parfaites dans l'ébauche. Si  $\mathbf{H}$  n'est pas une surjection, cela signifie que certaines observations sont redondantes, auquel cas le jeu d'observations peut être restreint à l'espace image de  $\mathbf{H}$ . Si  $\mathbf{R}$  n'est pas définie positive, l'expression (A2) de  $\mathbf{K}$  reste correcte (auquel cas l'analyse sera égale aux valeurs observées aux points  $H(\mathbf{x}_a)$ ), mais la forme variationnelle de l'analyse ne peut pas être utilisée. Il est même possible (avec certaines précautions algébriques) d'avoir des valeurs propres infinies dans  $\mathbf{R}$ , autrement dit d'avoir un  $\mathbf{R}^{-1}$  non défini positif, ce qui signifie que certaines observations ne possèdent aucun poids parce que leurs variances d'erreur sont infinies (cela peut être utile pour représenter le fait que certaines observations sont de très mauvaise qualité).

### 4.3.2 Traitement des biais

L'hypothèse d'absence de biais est souvent difficile à faire en pratique parce qu'il y a fréquemment des biais dans les champs d'ébauche \* (causés par un modèle de prévision imparfait) ou dans les observations, ou même dans l'opérateur d'observation. Les biais qui sont connus peuvent simplement être soustraits aux valeurs observées ou aux champs d'ébauche (opération de débiaisage, en anglais : bias correction). Dans ce cas, l'hypothèse d'absence de biais devient vraie pour les valeurs corrigées, et l'algèbre précédente s'applique sans modification.

Si on néglige de corriger des biais, l'analyse obtenue sera sous-optimale et contiendra des biais, ce qui ne l'empêche pas de les réduire par interpolation entre biais de l'ébauche et des observations. En pratique il est toujours important de surveiller les biais dans un système d'assimilation, par exemple en regardant des moyennes d'écart entre observations et ébauches. Mais il n'est pas toujours facile de décider si un biais 'observé'  $\overline{\mathbf{y} - \mathbf{H}\mathbf{x}_b}$  provient des ébauches ou des observations. Un biais peut apparaître comme une erreur systématique dans un paramètre, ou comme une fonction complexe de la date, de l'endroit, des conditions physiques locales, etc. auquel cas la distinction théorique entre biais et erreur aléatoire (qui est sensée être traitée par des matrices de covariance d'erreur, et non par une correction de biais) devient subtile... Fondamentalement, c'est notre propre capacité à prédire le signe de l'erreur qui différencie le biais de l'erreur aléatoire. Le BLUE réduit toutes les erreurs de ses données d'entrées, mais cette réduction est plus efficace pour les erreurs traitées comme biais (qui sont totalement effacées par la correction de biais) que comme erreur aléatoire (qui subit une minimisation au sens des moindres carrés, mais ne produit généralement pas une analyse parfaite). Il est donc intéressant en pratique de chercher à expliquer et corriger le plus possible les biais.

Dans un système d'assimilation, les biais sont plus problématique que lors d'analyses isolées, parce que les erreurs systématiques ont plus de temps pour s'accumuler et polluer

les analyses. Elles proviennent fréquemment de dérives lentes du modèle, qui s'accumulent lorsqu'il n'y a pas suffisamment d'observations pour les contrôler, ou d'erreurs de mesure qui finissent par faire dériver les analyses. Il est souvent instructif d'étudier le "climat" de l'assimilation en comparant les moyennes sur de longues périodes :

- d'une prévision très longue du modèle, c'est son "climat" préférentiel,
- des analyses, c'est le "climat" forcé par les observations,
- des ébauches ou de courtes prévisions du modèle, ce qui mesure la vitesse à laquelle le modèle a tendance à "oublier" les observations pour revenir à son propre climat.
- des incréments (analyses moins ébauche), ce qui indique comment les observations se "battent" avec les dérives du modèle.

La qualité de la correction de biais peut faire la différence entre un réseau d'observation utile ou nuisible à la qualité des analyses.

Des algorithmes plus ou moins automatiques de correction de biais sont en cours de développement, ils reviennent souvent à estimer les biais de modèle ou d'observation par des moyennes glissantes en temps réel au cours de l'assimilation. Mais il faut toujours contrôler avec des observations indépendantes si corriger les biais améliore le système, ou si au contraire cela l'encourage dans ses défauts.

### 4.3.3 Autres hypothèses qui affectent l'analyse BLUE

L'hypothèse d'*erreurs non corrélées* est souvent justifiée si les sources d'erreur dans les ébauches et les observations sont indépendantes. Mais il faut se méfier des techniques de prétraitement de certaines observations (satellites notamment) qui utilisent l'ébauche pour "améliorer" les observations (au sens où elles semblent plus cohérente avec l'ébauche), mais qui en fait introduisent dans les observations une partie des erreurs de l'ébauche. D'autant plus que l'on est souvent tenté de leur donner beaucoup de poids car elles semblent de très bonne qualité. Ce problème dit d'*inceste* peut dégrader énormément la qualité de l'analyse.

L'*hypothèse de linéarisation* n'est pas triviale et sera détaillée dans la section suivante.

Remarque : on peut réécrire les équations du BLUE en fonction des inverses des matrices de covariances d'erreur, ce que l'on appelle les matrices d'*information*. L'algèbre est un peu plus compliquée, mais elle fait apparaître clairement que l'information contenue dans l'analyse est, en un sens simple, la somme de l'information apportée par les observations et l'ébauche.

On peut aussi réécrire les équations du BLUE sous forme dite duale, où les calculs s'effectuent en majorité dans l'espace des observations  $\mathbf{y}$ . Un exemple est fourni par la forme PSAS ci-dessous.

## 4.4 Sur l'hypothèses linéaire tangente

L'hypothèse de *linéarisation de l'opérateur d'observation* est nécessaire pour produire une expression algébrique rigoureuse du  $\mathbf{K}$  optimal. En pratique,  $H$  n'est pas toujours linéaire,

mais on peut souvent le linéariser de manière satisfaisante au voisinage de l'ébauche :

$$H(\mathbf{x}) - H(\mathbf{x}_b) \simeq \mathbf{H}(\mathbf{x} - \mathbf{x}_b)$$

Sous cette hypothèse, la matrice  $\mathbf{K}$  étant une fonction continue de  $\mathbf{H}$ , l'analyse BLUE produira un  $\mathbf{x}_a$  qui est presque optimal.

On peut écrire cette *hypothèse linéaire tangente* sur  $H$  comme une formule de Taylor-Young au premier ordre au voisinage d'un état arbitraire  $\mathbf{x}$  pour une perturbation  $h$  :

$$H(\mathbf{x} + h) = H(\mathbf{x}) + \mathbf{H} h + \mathcal{O}(\|h\|^2)$$

avec  $\lim_{h \rightarrow 0} \mathcal{O}(\|h\|^2)h^{-2} = 0$ . Cette hypothèse n'est numériquement acceptable que si les variations de  $H$  d'ordre plus élevé sont négligeables, en particulier  $H$  doit être continu, pour toutes les perturbations considérées lors du processus d'analyse (perturbations dans l'espace de contrôle, de l'ordre de grandeur des erreurs d'ébauche). L'opérateur  $\mathbf{H}$  est la *différentielle*, où *dérivée première*, où opérateur *linéaire tangent (TL)*<sup>3</sup> de  $H$  au point  $\mathbf{x}$ . En pratique, cette définition mathématique de  $\mathbf{H}$  ne suffit pas toujours, puisque l'approximation

$$H(\mathbf{x} + h) - H(\mathbf{x}) \simeq \mathbf{H} h$$

doit être vraie, en des termes significatifs pour l'utilisateur, avec des valeurs finies de  $h$  qui dépendent de l'application considérée. Dans l'analyse BLUE, on veut que

$$\mathbf{y} - H(\mathbf{x}) \simeq \mathbf{y} - \mathbf{H}(\mathbf{x} - \mathbf{x}_b) + H(\mathbf{x}_b)$$

pour toutes les valeurs de  $\mathbf{x}$  rencontrées lors de la procédure d'analyse, notamment  $\mathbf{x} = \mathbf{x}_a$ ,  $\mathbf{x} = \mathbf{x}_t$ , ainsi que pour toutes les valeurs-test utilisées lors de la minimisation de  $J(\mathbf{x})$  si l'on résout la forme variationnelle de l'analyse<sup>4</sup>. Ce qui compte, c'est que la différence entre  $H(\mathbf{x}) - H(\mathbf{x}_b)$  et  $\mathbf{H}(\mathbf{x} - \mathbf{x}_b)$  doit être petite devant les erreurs d'observations typiques (définies par  $\mathbf{R}$ ), pour *toute* perturbation de l'état du modèle  $\mathbf{x} - \mathbf{x}_b$  de taille et forme cohérente avec les erreurs d'ébauche et avec l'amplitude des incréments d'analyse  $\mathbf{x}_a - \mathbf{x}_b$ .

Le problème de linéarisation de  $H$  ne dépend pas seulement des erreurs d'observation, il fait aussi intervenir les erreurs de l'ébauche  $\mathbf{x}_b$ , qui sont dans une assimilation séquentielle les erreurs de la prévision précédente, elles dépendent donc de la qualité du modèle et de l'échéance de prévision. En résumé, la validité de la linéarisation est fonction de l'ensemble du processus d'assimilation. Elle est meilleure dans un bon système, puisqu'alors les écarts  $\mathbf{x} - \mathbf{x}_b$  sont petits. Inversement, elle peut être inutilisable dans les problèmes d'assimilation difficiles. C'est souvent le cas en océanographie ou en assimilation de données satellitaires. Dans ces conditions, il vaut parfois mieux éviter de se lancer dans des techniques d'assimilation trop sophistiquées qui ne sont pas robustes vis à vis de fortes erreurs de linéarisation.

Le problème de linéarisation se pose aussi pour l'opérateur de prévision  $M$ , qui intervient dans le filtre de Kalman décrit ci-après. Là aussi, la validité de la linéarisation dépend du comportement d'ensemble du système d'assimilation : couverture en observations, qualité des mesures, résolution du modèle, faiblesses des paramétrisations physiques, etc.

<sup>3</sup>Les deux qualificatifs *tangent* et *linéaire* sont nécessaires :  $\mathbf{H}$  peut être linéaire sans satisfaire la formule de Taylor. Une fonction peut aussi être tangente à une autre sans pour autant être linéaire, si leur différence est en  $\mathcal{O}(\|h\|^2)$ , par exemple  $x^2$  et  $x^3$  sont tangente l'une de l'autre en  $x = 0$ .

<sup>4</sup>D'un point de vue qualitatif, toutes ces valeurs appartiennent à un voisinage de  $\mathbf{x}_b$  dont la forme et la taille dépendent des covariances d'erreur  $\mathbf{B}$  et  $\mathbf{R}$ .



## Analyse non-linéaire

L'hypothèse de linéarisation est parfois problématique, mais on ne sait en général calculer des équations d'analyse optimale qu'avec, par des manipulations d'algèbre linéaire. En pratique, on peut résoudre un problème d'assimilation faiblement non-linéaire, mais ce sera au prix de l'optimalité du résultat. La méthode *incrémentale* décrite plus loin est une approche empirique pour se rapprocher de l'optimalité malgré tout. Dans les problèmes fortement non-linéaires, des méthodes plus complexes, et souvent coûteuses, existent : le simplexe qui traite des variables à domaine de définition bornées, le recuit simulé pour minimiser des fonctions-coût à minima multiple, les méthodes génétiques, les réseaux de neurones... Il existe peut y avoir aussi des solutions très simple, comme un choix astucieux de variables modèle ou observées.

### 4.5 Le point de vue des probabilités conditionnelles

Il est instructif d'exprimer le problème d'analyse avec des probabilités conditionnelles, dites *bayésiennes*. On note  $P(x)$  la pdf a priori de l'état modèle ébauche, et  $P(y)$  la pdf des observations. Le but d'une analyse est de trouver le maximum de  $P(x | y)$ , probabilité conditionnelle de l'état modèle connaissant les observations. La probabilité conjointe de  $x$  et  $y$  (probabilité que les valeurs  $x$  et  $y$  se produisent simultanément)

$$P(x \wedge y) = P(x | y)P(y) = P(y | x)P(x)$$

autrement dit c'est la probabilité que  $x$  se produise pour un  $y$  donné, et vice versa. C'est le théorème de Bayes. Dans la procédure d'analyse on sait qu'une mesure a été effectuée avec la valeur  $y$ , donc  $P(y) = 1$  donc

$$P(x | y) = P(y | x)P(x)$$

which means that the analysis pdf is equal to the background pdf times the observation pdf  $P(y | x)$ . The latter peaks at  $y = H(x)$  but it is not a Dirac distribution because the observations are not error-free.

L'intérêt de l'approche probabiliste est qu'elle permet de traiter des problèmes comportant des probabilités non-gaussiennes (bien que cela invalide l'équation (A2) de calcul de  $Krm$ ). C'est utilisé pour le contrôle qualité des observations, dans lequel on suppose que les observations comportent une probabilité de "grosses erreurs", autrement dit d'erreurs rares, mais de grande amplitude qui ne suivent pas une modélisation gaussienne, par exemple à cause de défaillances instrumentales. Prendre le logarithme d'une telle pdf produit une fonction-coût non gaussienne, qui donne moins de poids à une observation dans l'analyse qu'une fonction-coût quadratique, pour des états modèle très incohérents avec l'observation.

ref : Lorenc 1986.

### 4.6 Coût numérique de l'analyse aux moindres carrés

Dans les modèles météorologiques opérationnels, la dimension de l'état du modèle (de l'espace de contrôle, en fait)  $\mathbf{x}$  est de l'ordre de  $n = 10^7$ , celle du vecteur d'observations (le

nombre de scalaires observés) est de l'ordre de  $p = 10^5$  toutes les 6 heures. Le problème d'analyse est donc mathématiquement sous-déterminé (mais il peut être localement surdéterminé là où la densité d'observation est supérieure à la résolution du modèle). Dans les applications pratiques il est important de bien comprendre les dimensions des opérateurs utilisés pour calculer l'analyse. (figure 4). L'analyse BLUE nécessite en principe de spécifier les matrices  $\mathbf{B}$  et  $\mathbf{R}$  (ou leurs inverses dans la forme variationnelle) qui contiennent respectivement  $n^2/2$  et  $p^2/2$  coefficients distincts, c'est à dire autant de statistiques à estimer. L'estimation d'une seule variance ou covariance converge comme la racine carrée du nombre de réalisations, autrement dit il en faut au moins une centaine par coefficient. Le calcul explicite de  $\mathbf{K}$  requiert l'inversion d'une matrice de taille  $p \times p$ , opération dont le coût asymptotique est de l'ordre de  $p^2 \log(n)$ . La minimisation exacte de la fonction-coût  $J$  nécessite en principe  $n+1$  évaluations de la fonction coût et de son gradient, en supposant que  $J$  est quadratique et qu'il n'y a pas d'erreur de troncature numérique dans les calculs de minimisation (par ex. par un algorithme de gradient conjugué). C'est évidemment trop coûteux à calculer exactement en dehors de problèmes académiques de très petite dimension. En pratique, il faut effectuer des approximations, ce sont les techniques expliquées dans les sections qui suivent.

## 4.7 Conclusion

Nous avons vu deux méthodes de définir le problème d'analyse optimale :

- soit on suppose que l'on connaît les covariances d'ébauche et d'analyse, on trouve alors les équations d'analyse en minimisant les variances d'erreur d'analyse,
- soit on suppose que les pdf d'erreur d'ébauche et d'analyse sont gaussiennes, et l'analyse est l'état dont la probabilité est maximale.

Ces deux approches conduisent aux mêmes deux algorithmes, mutuellement équivalents :

- le calcul direct de la matrice de gain  $\mathbf{K}$ ,
- ou la minimisation d'une fonction-coût quadratique.

En pratique ces algorithmes ont des propriétés numériques très différentes, leur équivalence disparaît dès qu'une des hypothèses sous-jacentes n'est pas vérifiée, comme la linéarisation.

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$

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

$$\mathbf{H} \mathbf{B} \mathbf{H}^T$$

$$\mathbf{J}(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x})$$

FIG. 4 – Dimensions schématiques des matrices et vecteurs intervenant dans un problème d'analyse typique où il y a plus de degrés de liberté dans le modèle que d'observations.

## 5 Exemple scalaire simple d'analyse BLUE

Supposons que l'on veut estimer la température  $T_t$  dans une pièce avec un thermomètre non biaisé dont la variance d'erreur de mesure  $\sigma_o^2$  est connue. Il mesure  $T_o$ , sans autre information on peut uniquement dire que la température est  $T_o$  avec une incertitude d'écart-type  $\sigma_o$ .

Maintenant supposons que l'on dispose en plus d'une autre estimation, appelée ébauche,  $T_b$ , de variance d'erreur  $\sigma_b^2$  (par exemple, c'est la température indiquée sur le thermostat du radiateur, mais on ne lui fait pas très confiance).

Intuitivement  $T_o$  et  $T_b$  peuvent être combinées pour fournir une meilleure estimation, ou *analyse*) de  $T_t$ . On cherche l'analyse sous forme d'une moyenne pondérée des 2 informations :

$$T_a = kT_o + (1 - k)T_b$$

que l'on peut aussi écrire  $T_a = T_b + k(T_o - T_b)$ , autrement dit on cherche une correction linéaire de l'ébauche  $T_b$ .

La variance d'erreur de l'estimation est :

$$\sigma_a^2 = (1 - k)^2\sigma_b^2 + k^2\sigma_o^2$$

où l'on a supposé que les erreurs d'ébauche et d'observation sont non corrélées. On va choisir d'utiliser le  $k$  qui minimise la variance d'erreur d'analyse :

$$k = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$$

ce qui est équivalent à minimiser (figure 5)

$$J(T) = J_b(T) + J_o(T) = \frac{(T - T_b)^2}{\sigma_b^2} + \frac{(T - T_o)^2}{\sigma_o^2}$$

- Si la mesure est infiniment mauvaise, ( $\sigma_o \gg \sigma_b$ ),  $k = 0$  et l'analyse est égale à l'ébauche.
- si l'observation est infiniment bonne ( $\sigma_o \ll \sigma_b$ ),  $k = 1$  et l'analyse est égale à l'observation.
- Si les 2 ont la même qualité,  $\sigma_o = \sigma_b$ ,  $k = 1/2$  et l'analyse est la moyenne arithmétique de  $T_o$  et  $T_b$  en guise de compromis.
- On a toujours  $0 \leq k \leq 1$ , autrement dit l'analyse est une moyenne pondérée convexe entre ébauche et observation.

Ces considérations sont schématisées sur la figure 6.

Il est instructif de regarder la variance d'erreur d'analyse pour le  $k$  optimal :

$$\frac{1}{\sigma_a^2} = \frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2}$$

ou

$$\sigma_a^2 = \frac{\sigma_o^2}{1 + (\sigma_o/\sigma_b)^2} = \frac{\sigma_b^2}{1 + (\sigma_b/\sigma_o)^2} = (1 - k)\sigma_b^2$$

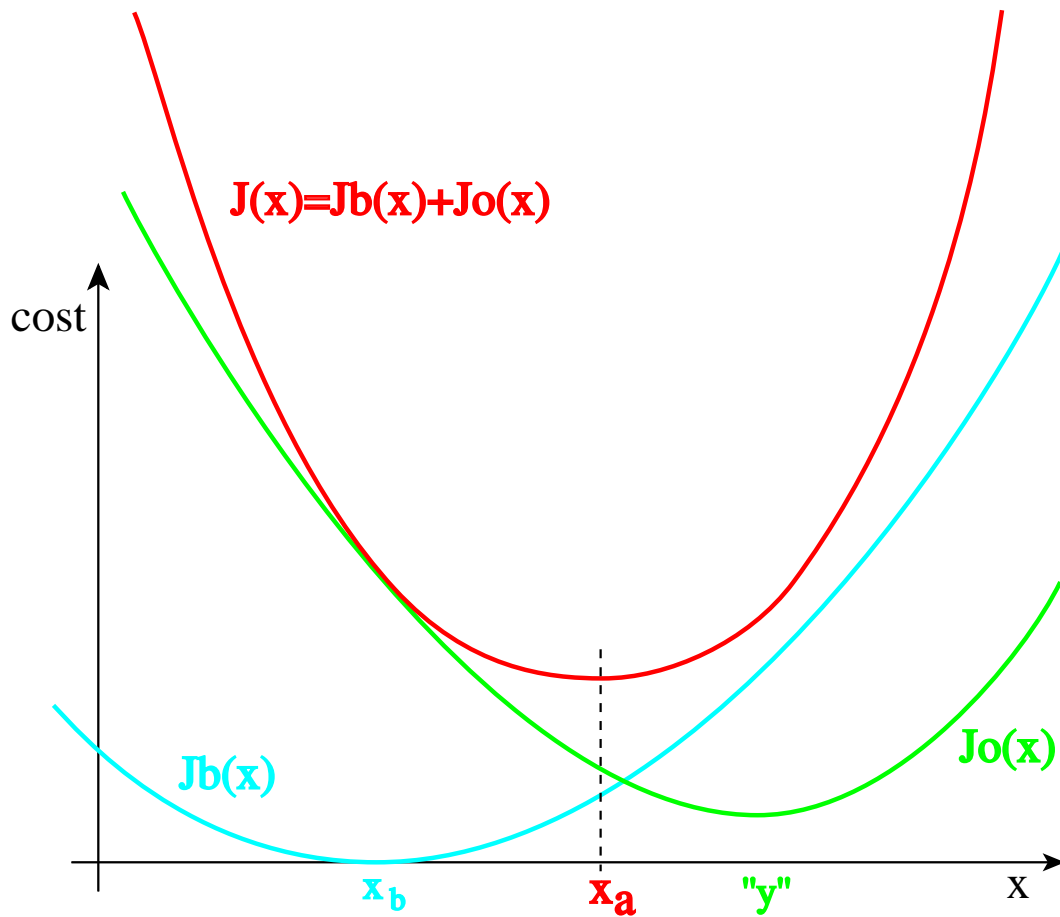


FIG. 5 – Représentation schématique scalaire de la forme variationnelle de l'analyse BLUE : les fonctions coût  $J_b$  et  $J_o$  tendent à 'attirer' l'analyse vers l'ébauche  $x_b$  et l'observation  $y$ , le minimum de leur somme est l'analyse optimale, quelque part entre les deux.

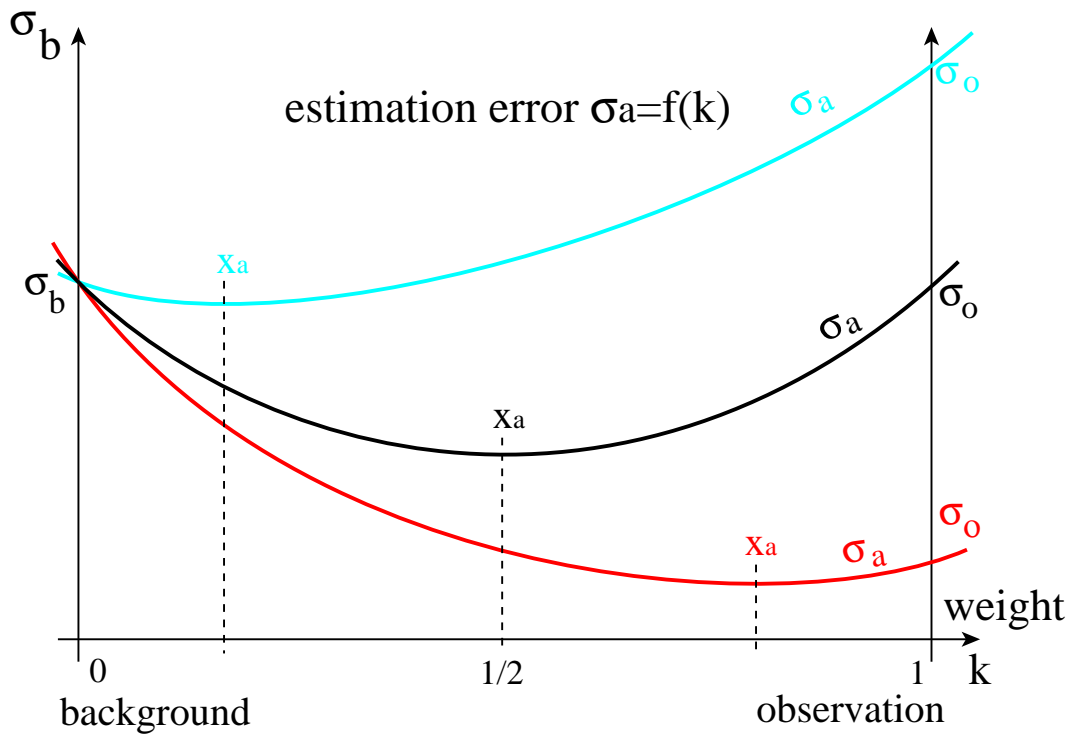


FIG. 6 – Schematic representation of the variations of the estimation error  $\sigma_a$ , and of the optimal weight  $k$  that determines the analysis  $\mathbf{x}_a$ , for various relative amplitudes of the background and observation standard errors ( $\sigma_b$ ,  $\sigma_o$ ).

qui montre que la variance d'erreur d'analyse est toujours inférieure au minimum des variances d'erreur d'ébauche et d'observation, et son propre minimum est atteint lorsque ces deux sont égales, auquel cas la variance d'erreur d'analyse est la moitié de la variance d'erreur d'ébauche (si on avait  $n$  observation, la variance d'erreur d'analyse tendrait vers zéro — autrement dit, la perfection ! — en  $1/(n + 1)$ )

## 6 Modèles de covariances d'erreurs

Les covariances d'erreurs d'ébauche et d'observation utilisées dans l'analyse déterminent comment l'information observée est spatialisée, et à quel point l'analyse 'collera' aux valeurs observées, notamment dans quelle mesure l'information observée est spatialement lissée ou non.

La forme de l'incrément d'analyse obtenu avec une seule observation s'appelle *fonction de structure*. Pour un nombre quelconque d'observations, l'incrément d'analyse est donc une combinaison linéaire de fonctions de structures.

Les corrélations d'erreur d'observation sont souvent supposées nulles. C'est compréhensible pour des mesures effectuées par des instruments différents, donc affectés par des erreurs physiques indépendantes. Il faut se méfier des erreurs de représentativité, par exemple des erreurs d'interpolation, qui sont généralement corrélées pour des observations très proches les unes des autres. Spécifier des corrélations d'erreur d'observation non-nulles aboutit à décroître le poids individuel de chaque observation concernée dans l'analyse (par rapport à des obs non corrélées), et à accroître le poids dans l'analyse des *différences* entre les valeurs observées, autrement dit aux gradients. Numériquement, il ne faut jamais mettre dans l'analyse d'observations trop corrélées (90% en valeur absolue), sous peine d'obtenir des artefacts numériques dans l'analyse, voire des problèmes de calcul des équations d'analyse.

On peut estimer objectivement des corrélations d'erreur d'observation en étudiant les corrélations avec des observations co-localisées dont on est sûr qu'elles sont non corrélées. Il existe aussi des techniques statistiques (mesurant l'optimalité de l'analyse pour différentes spécifications des corrélations d'erreur d'observation).

Les variances d'erreur d'observation peuvent être estimées par des observations colocalisées, par la connaissance de la qualité instrumentale des mesures utilisées, ou comme sous-produit de l'estimation des variances d'erreur d'ébauche.

Les variances d'erreur d'ébauche sont des estimations des erreurs dans la prévision précédente de l'assimilation. Le filtre de Kalman les estime directement (avec un coût de calcul élevé). En première approximation, on peut simplement spécifier une fraction de la variance climatologique des paramètres du modèle.

Les corrélations d'erreur d'ébauche sont les plus importantes à bien spécifier. Elles déterminent la forme des incréments d'analyse (l'incrément d'analyse avec 1 seule observation scalaire est un champ proportionnel à  $\mathbf{B}\mathbf{H}^T$ , autrement dit à 1 colonne de  $\mathbf{B}$  si l'on observe l'un des paramètres du modèle). Typiquement, on voudra que ces corrélations permettent de spatialiser les observations ponctuelles aussi loin qu'il existe des corrélations d'erreur significatives, mais pas plus loin que les limites physiques qui existent dans le système considéré (par exemple, une côte proche dans un modèle d'océan).

Les corrélations entre variables, appelées *balance*, doivent refléter autant que possible les relations physiques qui existent entre ces variables : équilibre hydrostatique, équation de continuité, équilibre géostrophique, etc... ceci afin d'éviter de créer des analyses qui n'ont aucune chance d'exister dans la nature, et aussi pour tirer profit des observations d'un paramètre



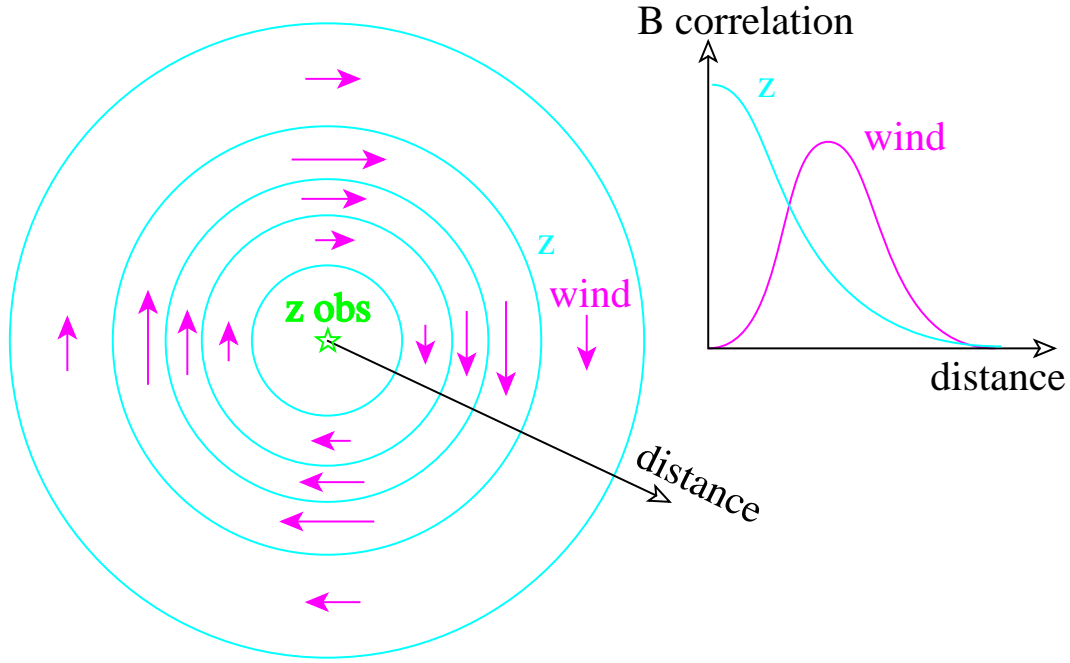


FIG. 7 – Exemple de fonctions de structure couramment utilisées en météorologie : autocorrélation horizontale de pression isotrope quasi-gaussienne, corrélation avec le vent qui respecte partiellement l'équilibre géostrophique, c'est à dire que le vent transverse est maximal là où le gradient de pression est maximal. Le sens du vent dépend de l'hémisphère considéré, l'intensité du couplage pression/vent est plus faible près du sol ou sous les tropiques.

pour corriger les autres paramètres du modèle.

L'expérience actuelle avec les systèmes d'analyse d'atmosphère et d'océan indiquent que la spécification des corrélations d'erreur d'ébauche est le facteur le plus important d'amélioration des algorithmes d'analyse (après la présence d'observations abondantes, bien entendu).

ref : Courtier et al 1998, Bouttier et al 1997.

## 6.1 Technique d'estimation of error covariances

Il existe de nombreuses méthodes, certaines très complexes. On décrit ici la méthode observationnelle, dite aussi de Hollingsworth-Lönnberg, qui est assez délicate à mettre en oeuvre mais dont le principe est très instructif. Le principe ( figure 8) consiste à calculer un histogramme, ou variogramme, de covariances d'erreurs d'ébauche, en fonction de la distance mutuelle des couples de points considérés, tirés dans un ensemble d'observations supposées avoir des statistiques d'erreurs homogènes : pour un ensemble de réalisations d'un couple de points d'observation  $i$  et  $j$ , la covariance des écarts (obs – ébauche) (innovations)  $c(i, j)$  est

$$\begin{aligned}
 c(i, j) &= \overline{(\mathbf{y}_i - \mathbf{H}_i \mathbf{x}_b)(\mathbf{y}_j - \mathbf{H}_j \mathbf{x}_b)^T} \\
 &= \overline{[(\mathbf{y}_i - \mathbf{H}_i \mathbf{x}_t) + (\mathbf{H}_i \mathbf{x}_t - \mathbf{H}_i \mathbf{x}_b)] [(\mathbf{y}_j - \mathbf{H}_j \mathbf{x}_t) + (\mathbf{H}_j \mathbf{x}_t - \mathbf{H}_j \mathbf{x}_b)]^T} \\
 &= \overline{(\mathbf{y}_i - \mathbf{H}_i \mathbf{x}_t)(\mathbf{y}_j - \mathbf{H}_j \mathbf{x}_t)^T} + \mathbf{H}_i (\mathbf{x}_t - \mathbf{x}_b)(\mathbf{x}_t - \mathbf{x}_b)^T \mathbf{H}_j^T +
 \end{aligned}$$

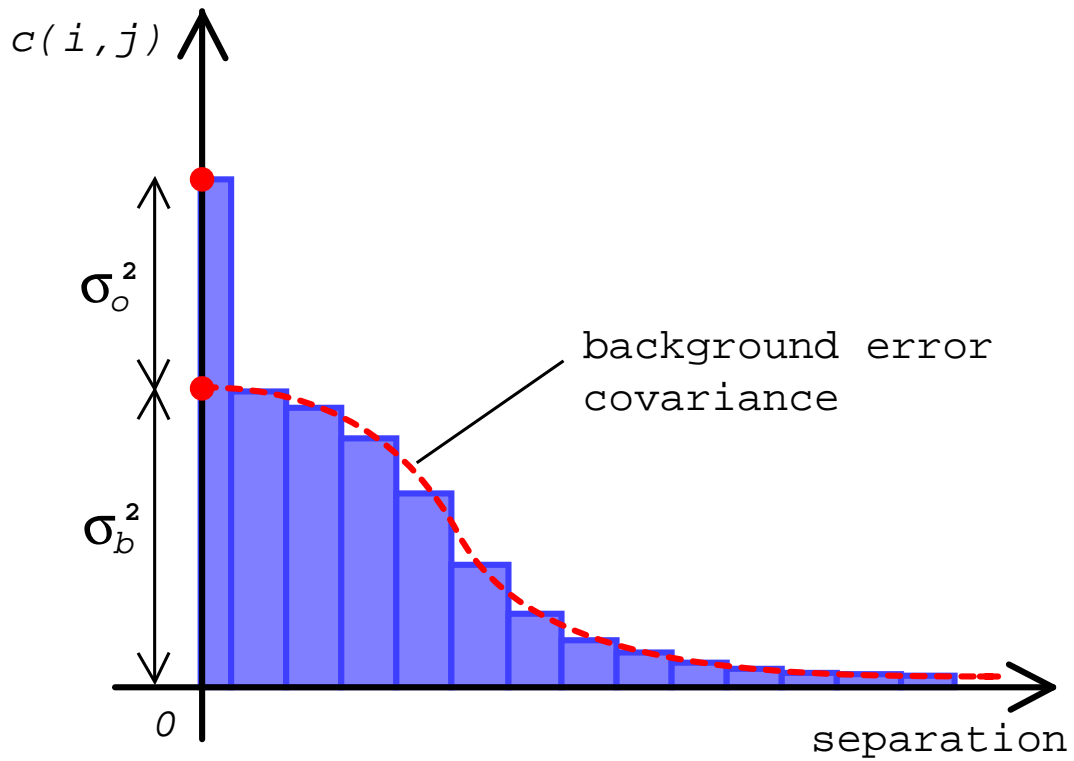


FIG. 8 – Représentation schématique de la méthode observationnelle. Les statistiques (obs – ébauche) d’un système d’assimilation donné sont stratifiées en fonction de la distance, l’interception à l’origine de l’histogramme ainsi obtenu indique les variances d’erreurs moyennes d’ébauche et d’observation pour le système d’assimilation et les ponts d’observation considérés.

$$\begin{aligned}
& + \overline{(\mathbf{y}_i - \mathbf{H}_i \mathbf{x}_t)(\mathbf{x}_t - \mathbf{x}_b)^T \mathbf{H}_j^T} + \overline{\mathbf{H}_i(\mathbf{x}_t - \mathbf{x}_b)(\mathbf{y}_j - \mathbf{H}_j \mathbf{x}_t)^T} \\
= & \mathbf{R}_{ij} + \mathbf{H}_i \mathbf{B} \mathbf{H}_j^T + 0 + 0
\end{aligned}$$

En supposant qu'il n'y a pas de corrélation entre erreurs d'observation et d'ébauche, les deux derniers termes de la 2e ligne sont nuls. Le premier terme est la covariance d'erreur d'observation entre  $i$  and  $j$ , le second terme est la covariance d'erreur d'ébauche interpolée en ces points. En résumé,

$$c(i, j) = \mathbf{R}_{ij} + \mathbf{H}_i \mathbf{B} \mathbf{H}_j^T$$

- si  $i = j$ ,  $c(i, i) = \sigma_o^2(i) + \sigma_b^2(i)$ , la somme des variances d'erreurs d'observation et d'ébauche.
- si  $i \neq j$  et que l'on suppose qu'il n'y a pas de corrélations d'erreurs d'observations,  $c(i, j) = \text{cov}_b(i, j)$ , la covariance d'erreur d'ébauche entre  $i$  et  $j$ .
- avec les mêmes hypothèses, si  $i$  et  $j$  sont très proches, alors  $\lim_{i \rightarrow j} c(i, j) = \sigma_b^2(i)$ , ce qui donne par continuité de  $c(i, j)$  en  $i = j$  la valeur de  $\sigma_b^2(i)$ .
- Par différence on en déduit  $\sigma_o^2(i) = c(i, i) - \sigma_b^2(i)$  puis les corrélations d'erreur d'ébauche avec  $c(i, j)/\sigma_b^2(i)$

ref : Hollingsworth and Lönnberg, 1986.

ref : Courtier et al 1998.

## 7 Optimal interpolation (OI) analysis

The OI is an algebraic simplification of the computation of the weight  $\mathbf{K}$  in the analysis equations :

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - H[\mathbf{x}_b]) \quad (\text{A1})$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \quad (\text{A2})$$

The equation (A1) can be regarded as a list of scalar analysis equations, one per model variable in the vector  $\mathbf{x}$ . For each model variable the analysis increment is given by the corresponding line of  $\mathbf{K}$  times the vector of background departures ( $\mathbf{y} - H[\mathbf{x}_b]$ ). The fundamental hypothesis in OI is : *For each model variable, only a few observations are important in determining the analysis increment.* It is implemented as follows :

1. For each model variable  $\mathbf{x}(i)$ , select a small number  $p_i$  of observations using empirical selection criteria.
2. Form the corresponding list of  $p_i$  background departures  $(\mathbf{y} - H[\mathbf{x}_b])_i$ , the  $p_i$  background error covariances between the model variable  $\mathbf{x}(i)$  and the model state interpolated at the  $p_i$  observation points (i.e. the relevant  $p_i$  coefficients of the  $i$ -th line of  $\mathbf{B}\mathbf{H}^T$ ), and the  $p_i \times p_i$  background and observation error covariance submatrices formed by the restrictions of  $\mathbf{H}\mathbf{B}\mathbf{H}^T$  and  $\mathbf{R}$  to the selected observations.
3. Invert the  $p_i \times p_i$  positive definite matrix formed by the restriction of  $(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})$  to the selected observations (e.g. by an *LU* or Choleski method),
4. Multiply it by the  $i$ -th line of  $\mathbf{B}\mathbf{H}^T$  to get the necessary line of  $\mathbf{K}$ .

It is possible to save some computer time on the matrix inversion by solving directly a symmetric positive linear system, since we know in advance the vector of departures to which the inverse matrix will be applied. Also, if the same set of observations is used to analyse several model variables, then the same matrix inverse (or factorization) can be reused.

In the OI algorithm it is necessary to have the background error covariances  $\mathbf{B}$  as a model which can easily be applied to pairs of model and observed variables, and to pairs of observed variables. This can be difficult to implement if the observation operators are complex. On the other hand, the  $\mathbf{B}$  matrix needs not be specified globally, it can be specified in an ad hoc way for each model variable, as long as it remains locally positive definite. The specification of  $\mathbf{B}$  usually relies on the design of empirical autocorrelation functions (e.g. Gaussian or Bessel functions and their derivatives), and on assumed amounts of balance constraints like hydrostatic balance or geostrophy.

The selection of observations should in principle provide all the observations which would have a significant weight in the optimal analysis, i.e. those which have significant background error covariances  $\mathbf{B}\mathbf{H}^T$  with the variable considered. In practice, background error covariances are assumed to be small for large separation, so that only the observations in a limited geometrical domain around the model variable need to be selected. For computational reasons it may be desirable to ensure that only a limited number of observations are selected each time, in order to keep the matrix inversions cheap. Two common strategies for observation selection are pointwise selection (figure 9) and box selection (figure 10)

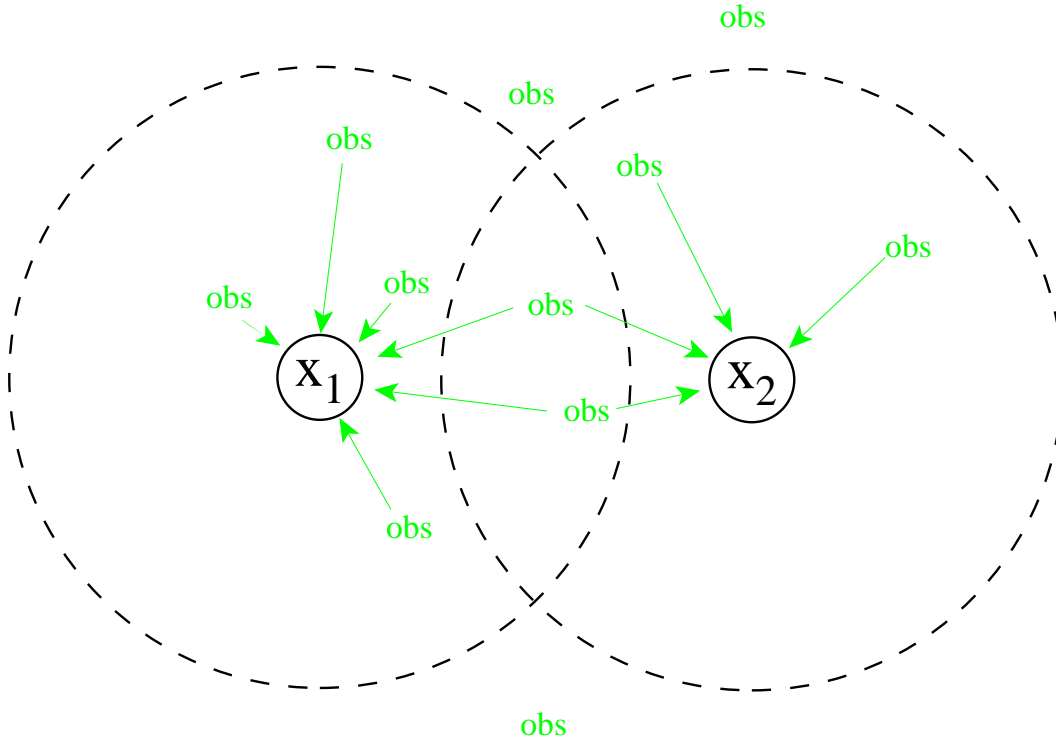


FIG. 9 – One OI data selection strategy is to assume that each analysis point is only sensitive to observations located in a small vicinity. Therefore, the observations used to perform the analysis at two neighbouring points  $x_1$  or  $x_2$  may be different, so that the analysis field will generally not be continuous in space. The cost of the analysis increases with the size of the selection domains.

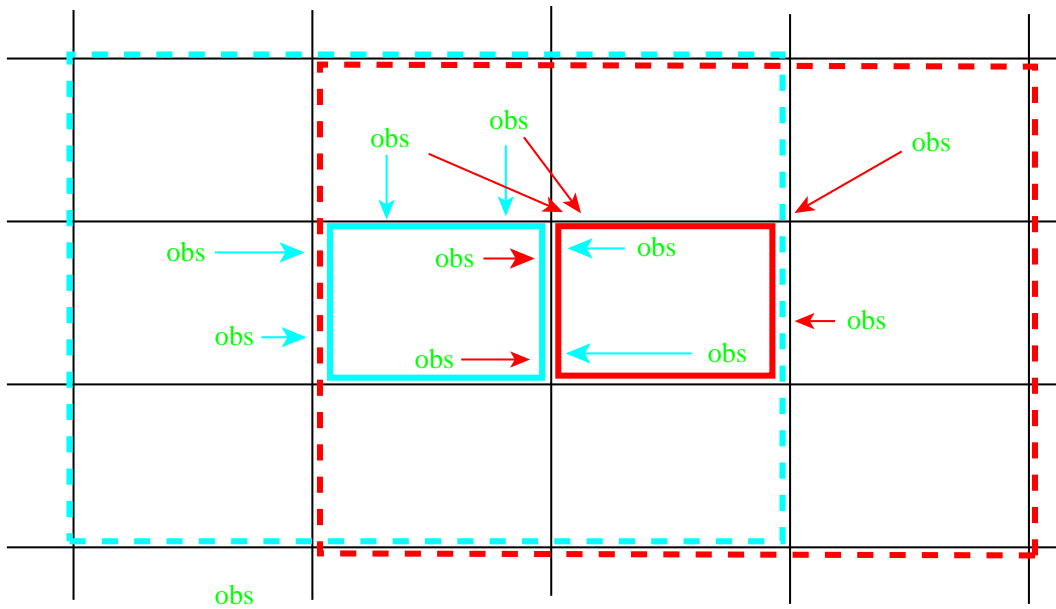


FIG. 10 – A slightly more sophisticated and more expensive OI data selection is to use, for all the points in an analysis box (black rectangle), all observations located in a bigger selection box (dashed rectangle), so that most of the observations selected in two neighbouring analysis boxes are identical.

The advantage of OI is its simplicity of implementation and its relatively small cost if the right assumptions can be made on the observation selection.

A drawback of OI is that spurious noise is produced in the analysis fields because different sets of observations (and possibly different background error models) are used on different parts of the model state. Also, it is impossible to guarantee the coherence between small and large scales of the analysis.

ref : Lorenc 1981.

## 8 Three-dimensional variational analysis (3D-Var)

The principle of 3D-Var is to avoid the computation (A2) of the gain  $\mathbf{K}$  completely by looking for the analysis as an approximate solution to the equivalent minimization problem defined by the cost function  $J$  in (A5). The solution is sought iteratively by performing several evaluations of the cost function

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - H[\mathbf{x}])^T \mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}])$$

and of its gradient

$$\nabla J(\mathbf{x}) = 2\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) - 2\mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}])$$

in order to approach the minimum using a suitable descent algorithm. The approximation lies in the fact that only a small number of iterations are performed. The minimization can be stopped by limiting artificially the number of iterations, or by requiring that the norm of the gradient  $\|\nabla J(\mathbf{x})\|$  decreases by a predefined amount during the minimization, which is an intrinsic measure of how much the analysis is closer to the optimum than the initial point of the minimization. The geometry of the minimization is suggested in figure 11.

In practice, the initial point of the minimization, or *first guess*, is taken equal to the background  $\mathbf{x}_b$ . This is not compulsory, however, so it is important to distinguish clearly between the terms *background* (which is used in the definition of the cost function) and *first guess* (which is used to initiate the minimization procedure). If the minimization is satisfactory, the analysis will not depend significantly on the choice of first guess, but it will always be sensitive to the background.

A significant difficulty with 3D-Var is the need to design a model for  $\mathbf{B}$  that properly defines background error covariances for all pairs of model variables. In particular, it has to be symmetric positive definite, and the background error variances must be realistic when expressed in terms of observation parameters, because this is what will determine the weight of the observations in the analysis.

The popularity of 3D-Var stems from its conceptual simplicity and from the ease with which complex observation operators can be used, since only the operators and the adjoints of their tangent linear need to be provided<sup>5</sup>. Weakly non-linear observation operators can be used, with a small loss in the optimality of the result. As long as  $J$  is strictly convex, there is still one and only one analysis.

In most cases the observation error covariance matrix  $\mathbf{R}$  is block-diagonal, or even diagonal, because there is no reason to assume observation error correlations between independent observing networks, observing platforms or stations, and instruments, except in some special cases. It is easy to see that a block-diagonal  $\mathbf{R}$  implies that  $J_o$  is a sum of  $N$  scalar cost-functions  $J_{o,i}$ , each one defined by a submatrix  $\mathbf{R}_i$  and the corresponding subsets  $H_i$  and  $\mathbf{y}_i$  of the observation operators and values :

$$J_o(\mathbf{x}) = \sum_{i=1}^N J_{o,i}(\mathbf{x})$$

---

<sup>5</sup> whereas OI requires a background error covariance model between each observed variable and each model variable.

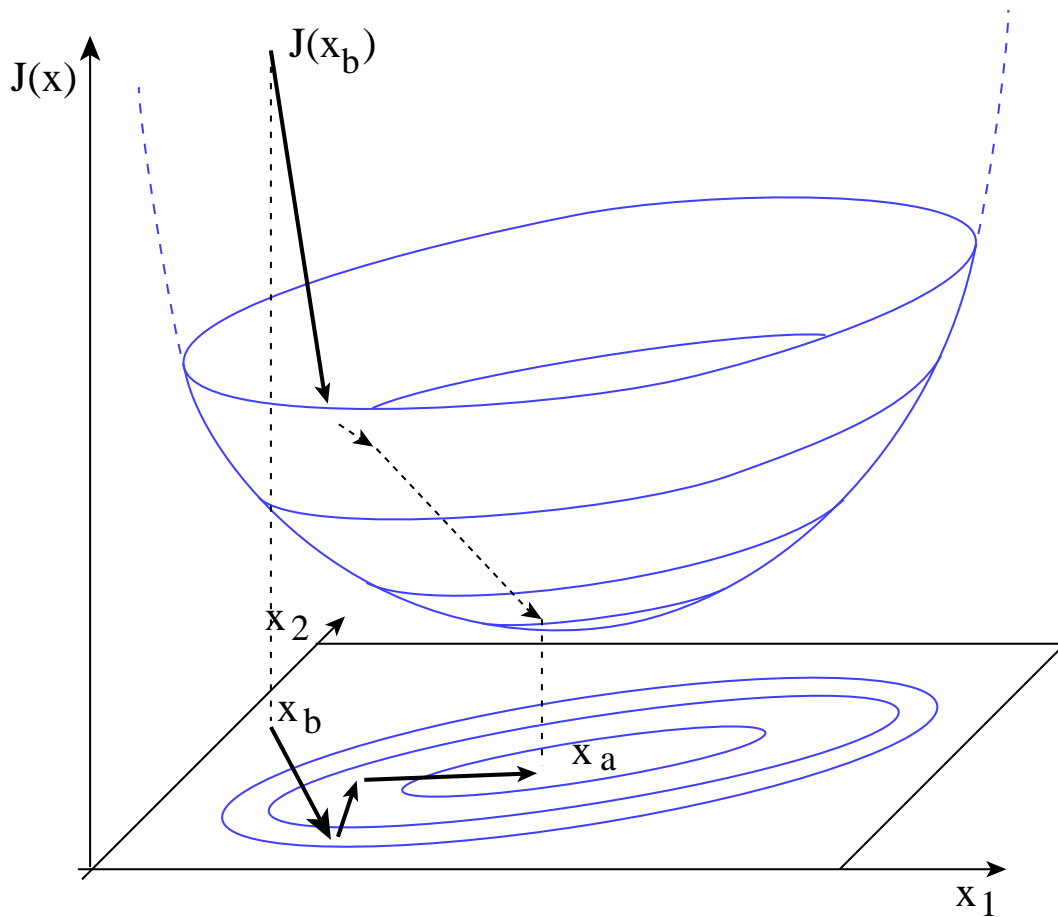


FIG. 11 – Schematic representation of the variational cost-function minimization (here in a two-variable model space) : the quadratic cost-function has the shape of a paraboloid, or bowl, with the minimum at the optimal analysis  $\mathbf{x}_a$ . The minimization works by performing several line-searches to move the control variable  $\mathbf{x}$  to areas where the cost-function is smaller, usually by looking at the local slope (the gradient) of the cost-function.



$$J_{o,i}(\mathbf{x}) = (\mathbf{y}_i - H_i[\mathbf{x}])^T \mathbf{R}_i^{-1} (\mathbf{y}_i - H_i[\mathbf{x}])$$

The gradient  $\nabla J_o$  can be similarly decomposed. The breakdown of  $J_o$  is a useful diagnostic tool of the behaviour of 3D-Var in terms of each observation type : the magnitude of each term measures the misfit between the state  $\mathbf{x}$  and the corresponding subset of observations. It can also simplify the coding of the computations of  $J_o$  and its gradient<sup>6</sup>.

Another advantage is the ability to enforce external weak (or *penalty*) constraints, such as balance properties, by putting additional terms into the cost function (usually denoted  $J_c$ ). However, this can make the preconditioning of the minimization problem difficult.

ref : Parrish and Derber 1992, Courtier et al 1998.

---

<sup>6</sup>Actually the whole  $J_o$  can be decomposed into as many elementary cost functions as there are observed parameters, by redefining the observation space to be the eigenvectors of  $\mathbf{R}$ .

## 9 1D-Var and other variational analysis systems

The essence of the 3D-Var algorithm is to rewrite a least-squares problem as the minimization of a cost-function. The method was introduced in order to remove the local data selection in the OI algorithm, thereby performing a global analysis of the 3-D meteorological fields, hence the name. Of course, the technique has been applied equally well to other problems in which the control variable is much smaller. A very successful example is the satellite data retrieval problem, in which the 1D-Var algorithm performs a local analysis of one atmospheric column (the model state) at the location of each satellite sounding such as TOVS radiances or microwave measurements. Similar variational techniques have been applied to the retrieval of surface wind fields from a collection of scatterometer ambiguous wind measurements or to the analysis of land surface properties in a numerical weather prediction model (in this case the control variable is more or less a column of the 3-D model, but the time dimension is taken into account as in 4D-Var). Except 1D-Var, these methods have no established name yet.

ref : Eyre 1987.

## 10 Four-dimensional variational assimilation (4D-Var)

4D-Var is a simple generalization of 3D-Var for observations that are distributed in time. The equations are the same, provided the observation operators are generalized to include a forecast model that will allow a comparison between the model state and the observations at the appropriate time.

Over a given time interval, the analysis being at the initial time, and the observations being distributed among  $n$  times in the interval, we denote by the subscript  $i$  the quantities at any given observation time  $i$ . Hence,  $\mathbf{y}_i$ ,  $\mathbf{x}_i$  and  $\mathbf{x}_{ti}$  are the observations, the model and the true states at time  $i$ , and  $\mathbf{R}_i$  is the error covariance matrix for the observation errors  $\mathbf{y}_i - H_i(\mathbf{x}_{ti})$ . The observation operator  $H_i$  at time  $i$  is linearized as  $\mathbf{H}_i$ . The background error covariance matrix  $\mathbf{B}$  is only defined at initial time, the time of the background  $\mathbf{x}_b$  and of the analysis  $\mathbf{x}_a$ .

### 10.1 The four-dimensional analysis problem

In its general form, it is defined as the minimization of the following cost function :

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + \sum_{i=0}^n (\mathbf{y}_i - H_i[\mathbf{x}_i])^T \mathbf{R}_i^{-1} (\mathbf{y}_i - H_i[\mathbf{x}_i])$$

which can be proven, like in the three-dimensional case detailed previously, to be equivalent to finding the maximum likelihood estimate of the analysis subject to the hypothesis of Gaussian errors.

The **4D-Var analysis**, or **four-dimensional variational assimilation problem**, is by convention defined as the above minimization problem subject to the strong constraint that the sequence of model states  $\mathbf{x}_i$  must be a solution of the model equations :

$$\forall i, \mathbf{x}_i = M_{0 \rightarrow i}(\mathbf{x})$$

where  $M_{0 \rightarrow i}$  is a predefined model forecast operator from the initial time to  $i$ . 4D-Var is thus a nonlinear constrained optimization problem which is very difficult to solve in the general case. Fortunately it can be greatly simplified with two hypotheses :

**Causality.** The forecast model can be expressed as the product of intermediate forecast steps, which reflects the causality of nature. Usually it is the integration of a numerical prediction model starting with  $\mathbf{x}$  as the initial condition. If the times  $i$  are sorted, with  $\mathbf{x}_0 = \mathbf{x}$  so that  $M_0$  is the identity, then by denoting  $M_i$  the forecast step from  $i - 1$  to  $i$  we have  $\mathbf{x}_i = M_i \mathbf{x}_{i-1}$  and by recurrence

$$\mathbf{x}_i = M_i M_{i-1} \cdots M_1 \mathbf{x}$$

**Tangent linear hypothesis.** The cost function can be made quadratic by assuming, on top of the linearization of  $H_i$ , that the  $M$  operator can be linearized, i.e.

$$\mathbf{y}_i - H_i M_{0 \rightarrow i}(\mathbf{x}) \simeq \mathbf{y}_i - H_i M_{0 \rightarrow i}(\mathbf{x}_b) - \mathbf{H}_i \mathbf{M}_{0 \rightarrow i}(\mathbf{x} - \mathbf{x}_b)$$

where  $\mathbf{M}$  is the *tangent linear (TL) model*, i.e. the differential of  $M$ . For a discussion of this hypothesis, refer to the section on the tangent linear hypothesis, in which the remarks made on  $H$  apply similarly to  $M$ . It explains that the realism of the TL hypothesis depends not only on the model, but also on the general characteristics of the assimilation system, including notably the length of the 4D-Var time interval.

The two hypotheses above simplify the general minimization problem to an unconstrained quadratic one which is numerically much easier to solve. The first term  $J_b$  of the cost function is no more complicated than in 3D-Var and it will be left out of this discussion. The evaluation of the second term  $J_o$  would seem to require  $n$  integrations of the forecast model from the analysis time to each of the observation times  $i$ , and even more for the computation of the gradient  $\nabla J_o$ . We are going to show that the computations can in fact be arranged in a much more efficient way.

## 10.2 Theorem : minimization of the 4D-Var cost function

The evaluation of the 4D-Var observation cost function and its gradient,  $J_o(\mathbf{x})$  and  $\nabla J_o(\mathbf{x})$ , requires one direct model integration from times 0 to  $n$  and one suitably modified *adjoint integration* made of transposes of the tangent linear model timestepping operators  $\mathbf{M}_i$ .

### Démonstration:

The first stage is the direct integration of the model from  $\mathbf{x}$  to  $\mathbf{x}_n$ , computing successively at each observation time  $i$  :

1. the forecast state  $\mathbf{x}_i = M_i M_{i-1} \cdots M_1 \mathbf{x}$ ,
2. the “normalized departures”  $d_i = \mathbf{R}_i^{-1}(\mathbf{y}_i - H_i[\mathbf{x}_i])$  which are stored,
3. the contributions to the cost function  $J_{oi}(\mathbf{x}) = (\mathbf{y} - H_i[\mathbf{x}_i])^T d_i$
4. And finally  $J_o(\mathbf{x}) = \sum_{i=0}^n J_{oi}(\mathbf{x})$ .

To compute  $\nabla J_o$  it is necessary to perform a slightly complex factorization :

$$\begin{aligned}
 -1/2 \nabla J_o &= -1/2 \sum_{i=0}^n \nabla J_{oi} \\
 &= \sum_{i=0}^n \mathbf{M}_1^T \cdots \mathbf{M}_i^T \mathbf{H}_i^T d_i \\
 &= \mathbf{H}_0^T d_0 + \mathbf{M}_1^T [\mathbf{H}_1^T d_1 + \mathbf{M}_2^T [\mathbf{H}_2^T d_2 + \dots + \mathbf{M}_n^T \mathbf{H}_n^T d_n] \cdots]
 \end{aligned}$$

and the last expression is easily evaluated from right to left using the following algorithm :

1. initialize the so-called *adjoint variable*  $\tilde{\mathbf{x}}$  to zero at final time :  $\tilde{\mathbf{x}}_n = 0$
2. for each time step  $i - 1$  the variable  $\tilde{\mathbf{x}}_{i-1}$  is obtained by adding the *adjoint forcing*  $\mathbf{H}_i^T d_i$  to  $\tilde{\mathbf{x}}_i$  and by performing the *adjoint integration* by multiplying the result by  $\mathbf{M}_i^T$ , i.e.  $\tilde{\mathbf{x}}_{i-1} = \mathbf{M}_i^T (\tilde{\mathbf{x}}_i + \mathbf{H}_i^T d_i)$
3. at the end of the recurrence, the value of the adjoint variable  $\tilde{\mathbf{x}}_0 = -1/2 \nabla J_o(\mathbf{x})$  gives the required result.

The terminology employed in the algorithm reflects the fact that the computations look like the integration of an *adjoint model* backward in time with a timestepping defined by the transpose timestepping operators  $\mathbf{M}_i^T$  and an external forcing  $\mathbf{H}_i^T d_i$ , which depends on the

distance between the model trajectory and the observations. In this discrete presentation it is just a convenient way of evaluating an algebraic expression<sup>7</sup>

### 10.3 Properties of 4D-Var

When compared to a 3-D analysis algorithm in a sequential assimilation system, 4D-Var has the following characteristics :

- it works only under the assumption that the model is perfect. Problems can be expected if model error are large.
- it requires the implementation of the rather special  $\mathbf{M}_i^T$  operators, the so-called adjoint model. This can be a lot of work if the forecast model is complex.
- in a real-time system it requires the assimilation to wait for the observations over the whole 4D-Var time interval to be available before the analysis procedure can begin, whereas sequential systems can process observations shortly after they are available. This can delay<sup>8</sup> the availability of  $\mathbf{x}_a$ .
- if  $\mathbf{x}_a$  is used as the initial state for a forecast, then by construction of 4D-Var one is sure that the forecast will be completely consistent with the model equations and the four-dimensional distribution of observations until the end of the 4D-Var time interval  $n$  (the *cutoff time*). This makes intermittent 4D-Var a very suitable system for numerical forecasting (figure 12).
- 4D-Var is an optimal assimilation algorithm over its time period thanks to the following theorem. It means that it uses the observations as well as possible, even if  $\mathbf{B}$  is not perfect, to provide  $\mathbf{x}_a$  in a much less expensive way than the equivalent Kalman Filter. For instance, the coupling between advection and observed information is illustrated in figure 13.

### 10.4 Equivalence between 4D-Var and the Kalman Filter

Over a given time interval, under the assumption that the model is perfect, with the same input data (initial background and its covariances  $\mathbf{B}$ , distribution of observations and their covariances  $\mathbf{R}_i$ ), the 4D-Var analysis at the end of the time interval is equal to the Kalman filter analysis at the same time.

This theorem is discussed in more details in the section about the Kalman filter algorithm, with a discussion of the pros and cons of using 4D-Var.

A special property of the 4D-Var analysis in the middle of the time interval is that it uses all the observations simultaneously, not just the ones before the analysis time. It is said that 4D-Var is a *smoothing* algorithm<sup>9</sup>.

---

<sup>7</sup>In a continuous (in time) presentation, the concept of adjoint model could be carried much further into the area of differential equations. However, this is not relevant to real models where the adjoint of the discretized model must be used, instead of the discretization of a continuous adjoint model. The only relevant case is if some continuous operators have a simple adjoint : then, with a careful discretization that preserves this property, the implementation of the discrete transpose operators can be simplified.

<sup>8</sup>Some special implementations of 4D-Var can partly solve this problem.

<sup>9</sup>Equivalent to the *Kalman smoother* algorithm which is a generalization of the Kalman filter, but at a much smaller cost.

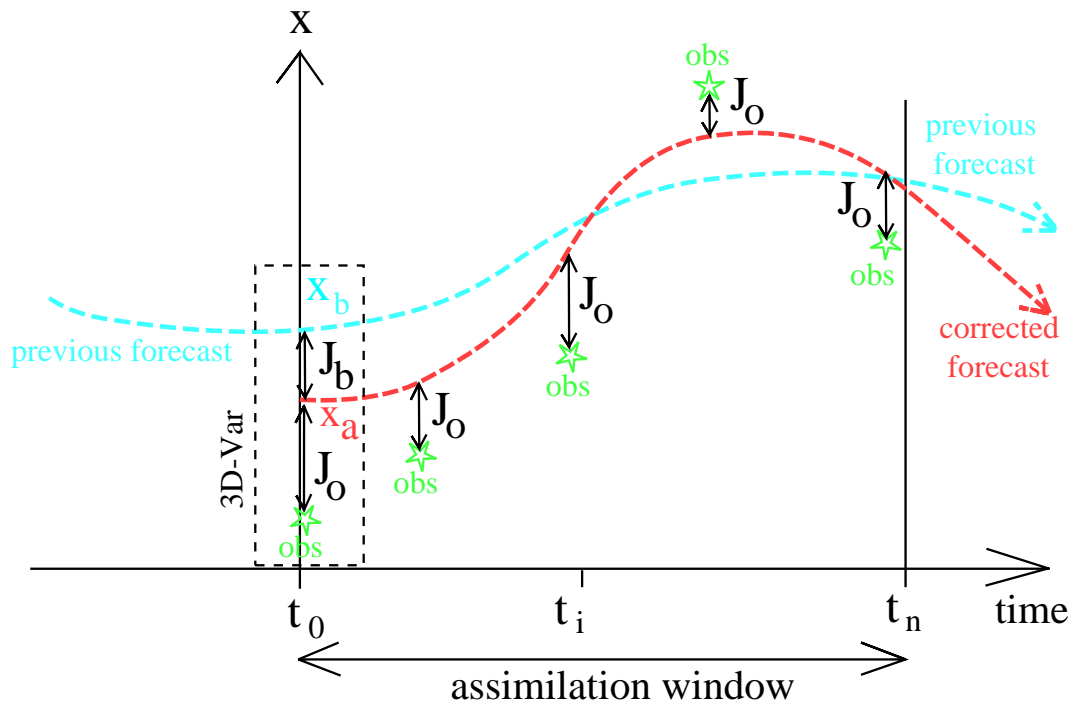


FIG. 12 – Example of 4D-Var intermittent assimilation in a numerical forecasting system. Every 6 hours a 4D-Var is performed to assimilate the most recent observations, using a segment of the previous forecast as background. This updates the initial model trajectory for the subsequent forecast.

ref : Talagrand and Courtier 1987, Thépaut and Courtier 1991, Rabier and Courtier 1992, Lacarra and Talagrand 1988, Errico et al 1993.

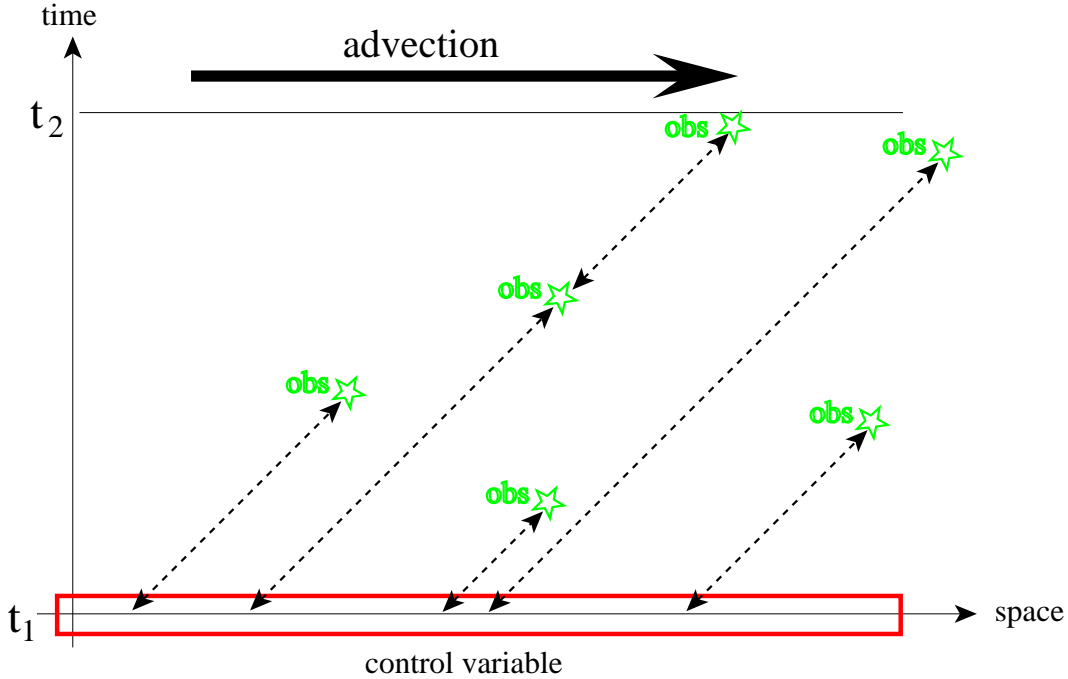


FIG. 13 – Example of propagation of the information by 4D-Var (or, equivalently, a Kalman filter) in a 1-D model with advection (i.e. transport) of a scalar quantity. All features observed at any point within the 4D-Var time window ( $t_1, t_2$ ) will be related to the correct upstream point of the control variable by the tangent linear and adjoint model, along the characteristic lines of the flow (dashed).

## 11 Estimating the quality of the analysis

It is usually an important property of an analysis algorithm that it should be able to provide an estimate of the quality of its output. If there is no observation the quality is obviously that of the background. In a sequential analysis system the knowledge of the analysis quality is useful because it helps in the specification of the background error covariances for the next analysis, a problem called *cycling* the analysis. If the background is a forecast, then its errors are a combination of analysis and model errors, evolved in time according to the model dynamics. This is explicitly represented in the Kalman filter algorithm.

If the analysis gain  $\mathbf{K}$  has been calculated, e.g. in an OI analysis, then the analysis error covariance matrix is provided by the equation

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \quad (\text{A3})$$

which reduces to  $\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}$  (A4) in the unlikely case where  $\mathbf{K}$  has been computed exactly.

In a variational analysis procedure, the error covariances of the analysis can be inferred from the matrix of second derivatives, or *Hessian*, of the cost function thanks to the following result :

## 11.1 Theorem : use of Hessian information

The Hessian of the cost function of the variational analysis is equal to twice the inverse of the analysis error covariance matrix :

$$\mathbf{A} = (1/2J'')^{-1}$$

### Démonstration:

The Hessian is obtained by differentiating  $J$  twice with respect to the control variable  $\mathbf{x}$  :

$$\begin{aligned} J(\mathbf{x}) &= (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - H[\mathbf{x}])^T \mathbf{R}^{-1} (\mathbf{y} - H[\mathbf{x}]) & (A5) \\ \nabla J &= 2\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) - 2\mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}]) \\ J'' = \nabla \nabla J &= 2(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) \end{aligned}$$

Now we express the fact that  $\nabla J(\mathbf{x}_a) = 0$  and we insert the true model state  $\mathbf{x}_t$  into the equation :

$$\begin{aligned} 0 &= \mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_b) - \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}_a]) \\ &= \mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_t + \mathbf{x}_t - \mathbf{x}_b) - \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}_t] + \mathbf{H}[\mathbf{x}_t - \mathbf{x}_a]) \\ &= \mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_t) - \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}(\mathbf{x}_t - \mathbf{x}_a) - \mathbf{B}^{-1}(\mathbf{x}_b - \mathbf{x}_t) - \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}_t]) \end{aligned}$$

Hence

$$(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})(\mathbf{x}_a - \mathbf{x}_t) = \mathbf{B}^{-1}(\mathbf{x}_b - \mathbf{x}_t) + \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - H[\mathbf{x}_t])$$

When it is multiplied on the right by its transpose, and the expectation of the result is taken, the right-hand side then contains two terms that multiply  $(\mathbf{x}_b - \mathbf{x}_t)(\mathbf{y} - H[\mathbf{x}_t])^T$  which is zero because we assume background and observation errors are uncorrelated. The remaining terms lead to, successively :

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

which proves the result.

## 11.2 Remarks

A simple, geometrical illustration of the relationship between the Hessian and the quality of the analysis is provided in figure 14. In a multidimensional problem, the same interpretation is valid along cross-sections of the cost-function.

If the linearization of the observation operator  $H$  can be performed exactly, the cost function  $J$  is exactly quadratic and  $J''$  does not depend on the value of the analysis :  $\mathbf{A}$  can be determined as soon as  $J$  is defined, even before the analysis is actually carried out<sup>10</sup>. If the linearization is not exact,  $J''(\mathbf{x})$  is not constant. It may depend a lot on  $\mathbf{x}$ , even if  $J$  itself does not look very different from a quadratic function. For instance, if  $J$  is continuously differentiable but not strictly convex, there are points at which  $J'' = 0$ . If  $\nabla J$  is not continuous, then there are points at which  $J''$  is not defined at all. It means that  $H$  must be exactly linear in order to be able to calculate  $\mathbf{A}$  using the Hessian. In practice  $J$  must be modified to use the tangent linear of  $H$ , which can be acceptable in a close vicinity of  $\mathbf{x}_a$ .

<sup>10</sup>Actually, neither  $\mathbf{A}$  nor  $\mathbf{K}$  depend on the values of the background or of the observations.



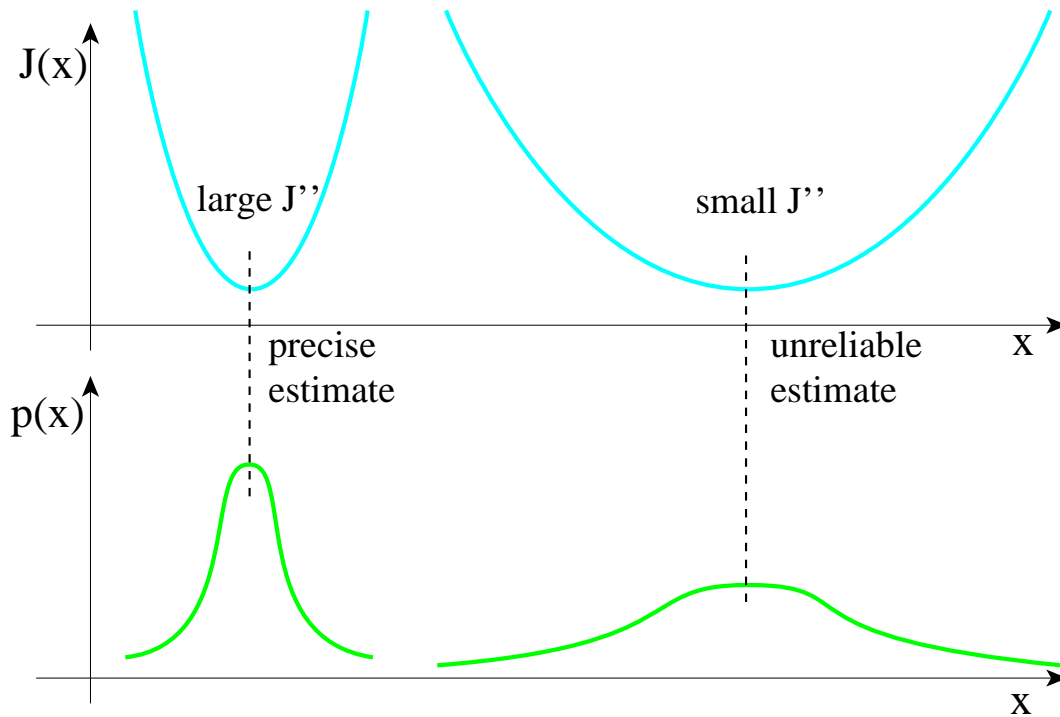


FIG. 14 – Illustration in a one-dimensional problem of the relationship between the Hessian and the quality of the analysis. In one dimension, the Hessian is the second derivative, or convexity, of the cost-function of the variational analysis : two examples of cost-functions are depicted in the upper panel, one with a strong convexity (on the left), the other with a weaker one (on the right). If the cost-function is consistent with the pdfs involved in the analysis problem, the Hessian is a measure of the sharpness of the pdf of the analysis (depicted in the lower panel). A sharper pdf (on the left) means that the analysis is more reliable, and that the probability of the estimated state to be the true one is higher.

The identity  $\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$  shows clearly how the observed data increases the inverse error covariances, also called *information matrices*.

ref : Rabier and Courtier 1992.

## 12 Implementation techniques

In most practical applications, numerical cost is an important issue. As shown above, there is a variety of analysis methods available. It does not imply that any of these is the best; they should be regarded as a choice of several compromises between numerical cost, statistical optimality and physical realism of the assimilation system. The sections below describe other features of the analysis algorithms which can be used to further cut down on the numerical cost, without sacrificing too much on the sophistication of the analysis method itself. They are discussed here in the framework of 3D-Var, but they can be applied equally well (with a few adaptations) to all related algorithms : 1D-Var, 4D-Var, PSAS or the Kalman filter.

### 12.1 Minimization algorithms and preconditioning

In a variational analysis system a cost function has to be minimized, usually using an iterative descent algorithm. The cost of the analysis is proportional to the number of evaluations of the cost function and its gradient<sup>11</sup>, called the number of *simulations*. When the state itself  $\mathbf{x}$  is updated, an *iteration* is performed. Each iteration may require one or more simulations, depending on the minimizing algorithm used. Hence the technical implementation of a variational analysis can be summarized as a *simulator* operator :

$$\mathbf{x} \mapsto J(\mathbf{x}), \nabla J(\mathbf{x})$$

How to use the simulator to minimize the cost function is a well-developed area of mathematics (called *optimization*, a part of *numerical analysis*). With the analysis methods described above, the cost function will be a scalar function of a real vector in a euclidean space; in most applications it will be quadratic and  $\mathbf{x}$  will be unconstrained. There are several ready-to-use algorithms that do the minimization, called *minimizers*. An obvious method, the *steepest descent* method, is to update  $\mathbf{x}$  by adding a correction that is proportional to  $-\nabla J(\mathbf{x})$ . This is usually not very efficient, and more popular algorithms are the *conjugate gradient* and *quasi-Newton* methods. They are still being improved. There are more specialized algorithms for situations where  $J$  is not quadratic or  $\mathbf{x}$  is bounded, e.g. *simulated annealing* or the *simplex*, although such methods can be very expensive. The *incremental* method described below can also be regarded as a particular minimizer. A detailed description of the main minimizing algorithms can be found in dedicated mathematical books. Among the important theoretical results are the optimality properties of the conjugate gradient method in the case of an exactly quadratic cost function, and its equivalence with a Lanczos method for determining eigenvectors of the Hessian matrix. Also, the quasi-Newton methods can be regarded as a preconditioning of the cost function using accumulated information about the second derivatives.

The main aspect of  $J$  that affects the performance of conventional minimizers (assuming  $J$  is quadratic or almost) is its *condition number*. This quantity measures the ellipticity of the iso-surfaces of  $J$ , and it describes the difficulty of minimization problem (or *ill-conditioning*)

---

<sup>11</sup>Some minimization algorithms also use information about the second derivative of the cost function, which requires the coding of the *second-order adjoint* of its components.

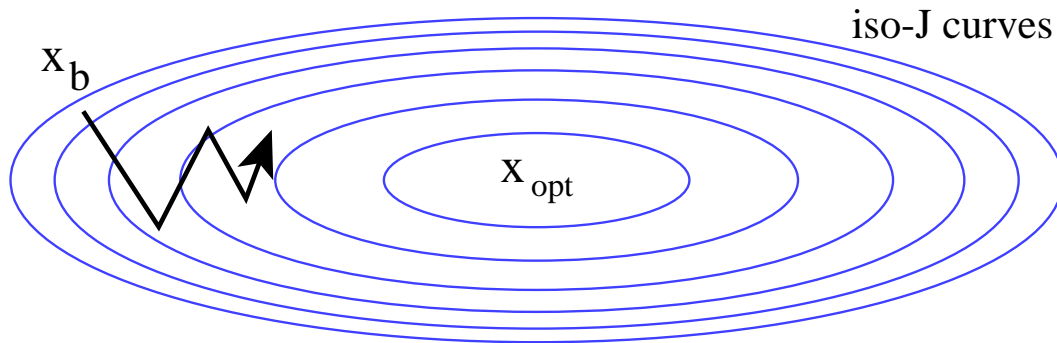


FIG. 15 – Illustration of the so-called narrow valley effect : in a plane of the control variable space where the convexity of the cost-function depends a lot on direction, the isolines are narrow ellipses, and in most places the gradient of the cost function is nearly orthogonal to the direction of the minimum  $\mathbf{x}_{opt}$ , which means that minimization algorithms will tend to waste many iterations zig-zagging slowly towards the optimum.

due to the gradient  $\nabla J$  not pointing accurately toward the minimum (figure 15). In this case minimizers have trouble converging, a phenomenon called the *narrow valley* effect.

**Condition number.** The condition number of  $J$  is defined to be the ratio between the largest and the smallest eigenvalue of  $J''$ . The largest the number, the more ill-conditioned the problem is.

If the condition number is equal to one, i.e.  $J''$  is proportional to  $\mathbf{I}$ , the cost function is said to be *spherical* and the minimum can be found in one iteration because  $-\nabla J(\mathbf{x}_b)$  points directly toward the minimum.

In the general case,  $J$  is elliptic, but it is possible to define a change of minimization space called *preconditioning* that decreases the condition number. The idea is to present the minimizer with a problem that is not the minimization of  $J(\mathbf{x})$ , but another easier problem from which  $\mathbf{x}_a$  can be obtained easily. The mapping between both problems is defined as follows using a *preconditioner* operator  $L$  :

## 12.2 Theorem : preconditioning of a variational analysis

If  $L$  is an invertible operator, an equivalent rewriting of the minimization problem :

$$\begin{aligned} \mathbf{x}_a &= \text{Arg min } J \\ J : \mathbf{x} &\mapsto J(\mathbf{x}), \nabla J(\mathbf{x}) \\ \text{with the initial point } \mathbf{x}_{ini} &= \mathbf{x}_b \end{aligned}$$

is the *preconditioned* problem :

$$\begin{aligned} \chi_a &= \text{Arg min } \hat{J} \\ \hat{J} : \chi &\mapsto \hat{J}(\chi) = J(L\chi), \nabla \hat{J}(\chi) = L^T \nabla J(L\chi) \\ \text{with the initial point } \chi_{ini} &= L^{-1} \mathbf{x}_b \end{aligned}$$

The solution is given by  $\mathbf{x}_a = L\chi_a$ .

The proof is left as an exercise. In 3D-Var, a simple and efficient preconditioner is the symmetric left-hand square root of  $\mathbf{B}$ , i.e. a matrix<sup>12</sup>  $L$  such that  $\mathbf{B} = LL^T$ . In this case one can show that

$$\hat{J}(\chi) = \chi^T \chi + J_o(L\chi)$$

i.e. the  $J_b$  term is now the canonical inner product. An ideal preconditioner would of course be provided by the symmetric square root of the Hessian matrix. Some sophisticated minimizer packages allow the user to provide his own preconditioner to the code, which can take the form of a clever specification of the inner product.

ref : Gilbert and Lemaréchal 1989.

## 12.3 The incremental method

The incremental method is a relatively empirical technique designed to reduce the cost of solving a predefined variational problem, e.g. by reducing the resolution of the increments.

In the introduction it was explained how the control variable could be made smaller than the model state by requiring that the increments can only be non-zero in a subspace of the model. In this case there is no guarantee that the analysis verifies any optimality condition in the full model space. For instance, OI solves the problem separately in a set of subspaces (defined by the observation selection), but the result is not as optimal as a global least-squares analysis. With 3D- or 4D-Var it is usually not affordable to solve the variational problem at the full model resolution. However, it is expected that most of the complexity of the analysis is in the synoptic scales, because this is where most background errors are expected to be. If the increments are right at the synoptic scales, then one can expect the smaller scales to be more or less forced to be realistic features by the model dynamics. It is undesirable, though, to completely neglect the small scales in the analysis procedure because they are important in the comparison of the observations with the background state. In other words, one is looking for a

<sup>12</sup>The symmetric square root is not unique, it is defined modulo an orthogonal matrix.

low-resolution correction to a high-resolution background. The incremental method described below has been designed for this particular problem. Mathematically, it can be thought the approximation of a large problem by a sequence of smaller problems. However, there is no proof of the convergence of the general procedure<sup>13</sup>

In the incremental method some high-resolution versions of the cost function, the observation operator and the model state are considered, denoted respectively  $(J_h, H_h, \mathbf{x}_h)$ . We are trying to minimize  $J_h(\mathbf{x}_h)$ . One or several successive approximations to this problem are solved successfully. Each one is an *inner loop* that tries to *update* a high-resolution state  $\mathbf{x}_{h,i}$  into another one  $\mathbf{x}_{h,i+1}$  that is more optimal (in the first update,  $\mathbf{x}_{h,i} = \mathbf{x}_b$ ). The input information to the inner loop is given by the high-resolution departures :

$$d_{h,i} = \mathbf{y} - H_h(\mathbf{x}_{h,i})$$

and by a low-resolution version  $\mathbf{x}_i$  of  $\mathbf{x}_{h,i}$  defined by a conversion operator  $S_{h \rightarrow l}$  :

$$\mathbf{x}_i = S_{h \rightarrow l}(\mathbf{x}_{h,i})$$

It is natural to linearize the low-resolution observation operator  $H$  in the vicinity of  $\mathbf{x}_i$  which is the best currently available estimate of the analysis<sup>14</sup>, which yields a linearized observation operator  $\mathbf{H}_i$  that depends on the update index  $i$ , and defined as the differential of  $H$  in the vicinity of  $\mathbf{x}_i$  :

$$H(\mathbf{x}) \simeq \mathbf{H}_i(\mathbf{x} - \mathbf{x}_i) + H(\mathbf{x}_i)$$

However, for consistency with the high-resolution problem, one also requires that the low-resolution is kept consistent with the high-resolution one for  $\mathbf{x} = \mathbf{x}_i$ , so that the linearized departures used at low resolution will be calculated as

$$\mathbf{y} - H(\mathbf{x}) \simeq \mathbf{y} - [\mathbf{H}_i(\mathbf{x} - \mathbf{x}_i) + H_h(\mathbf{x}_{h,i})] = d_{h,i} - \mathbf{H}_i(\mathbf{x} - \mathbf{x}_i)$$

so that the low-resolution cost-function to minimize in the inner loop is

$$J_i(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + [d_{h,i} - \mathbf{H}_i(\mathbf{x} - \mathbf{x}_i)]^T \mathbf{R}^{-1}[d_{h,i} - \mathbf{H}_i(\mathbf{x} - \mathbf{x}_i)]$$

which is exactly quadratic. Its minimum is  $\mathbf{x}_{i+1}$  which can in turn be used to update the high-resolution state using a (possibly nonlinear) ad hoc conversion operator  $S_{l \rightarrow h}$  :

$$\mathbf{x}_{h,i+1} = \mathbf{x}_{h,i} + S_{l \rightarrow h}(\mathbf{x}_{i+1}) - S_{l \rightarrow h}(\mathbf{x}_i)$$

which ensures that the high-resolution state is not modified if the inner loop minimization does not change the state. From  $\mathbf{x}_{h,i+1}$  the new high-resolution departures  $d_{h,i+1}$  can be calculated and used to define the next low-resolution problem. If  $d_{h,i+1} = d_{h,i}$  then the high- and low-resolution problems are fully consistent with each other and the whole algorithm has converged. However, it not guaranteed that there is a convergence at all. This is why one must be careful about the physical implications of changing the resolution. The intuitively important assumption for convergence (this can be proven in simplified systems) is that

$$H_h \circ S_{l \rightarrow h}(\mathbf{x}_{i+1}) - H_h \circ S_{l \rightarrow h}(\mathbf{x}_i) \simeq \mathbf{H}_i(\mathbf{x}_{i+1} - \mathbf{x}_i)$$

<sup>13</sup>It is possible to guarantee convergence for some special forms of the incremental algorithm.

<sup>14</sup>One would rather like to use a low-resolution version of the linearized high-resolution  $H_h$  in the vicinity of  $\mathbf{x}_{h,i}$ , but it would be more expensive than the technique described here.

i.e. the changes in the model equivalents of the observations should be similar at high and low resolutions. If for instance they are of opposite signs, one can expect the model state at high resolution to go *away* from the observations during the procedure until it is stopped by the  $J_b$  term — a not very desirable behaviour. Whether this is a genuine problem is still an area of research. History has shown so far that 3D-Var with a simple incremental formulation and a rather low resolution of the inner loops can be much better than an OI algorithm at full resolution, for a similar numerical cost.

ref : Courtier, Thépaut and Hollingsworth, 1994.

## 12.4 The adjoint technique

As shown in the explanation of the 4D-Var method, some computational savings can be achieved by a suitable ordering of the algebraic operations, in order to reduce the size and number of the matrix multiplications involved. For minimization problems in particular, when the derivative of a scalar function with respect to a large vector needs to be evaluated (e.g.  $J_o$ ), it is advantageous to use the chain rule *backwards*, i.e. from the scalar function to the input vector. Algebraically this means replacing a set of matrices by their transposes, hence the name of *adjoint* technique. The definition of the adjoint depends on the scalar products<sup>15</sup> used :

**Adjoint operator.** By definition, given a linear operator  $A$  going from a space  $E$  to a space  $F$ , and scalar products  $\langle \cdot, \cdot \rangle_E$ , and  $\langle \cdot, \cdot \rangle_F$  in these respective spaces, the *adjoint* of  $A$  is the linear operator  $A^*$  such that for any vectors  $(x, y)$  in the suitable spaces,

$$\langle Ax, y \rangle_F = \langle x, A^*y \rangle_E$$

### Important remarks on the adjoints.

- **Riesz theorem** : The adjoint always exists and it is unique, assuming spaces of finite dimension<sup>16</sup>. Hence, coding the adjoint does not raise questions about its existence, only questions of technical implementation.
- In the meteorological literature, the term *adjoint* is often improperly used to denote the adjoint of the tangent linear of a non-linear operator. One must be aware that discussions about the “existence of the adjoint” usually address the existence of the tangent linear operator (or the acceptability of using the adjoint of an improper tangent-linear in order to minimize a 4D-Var cost-function). As explained above, the adjoint itself always exist.
- In general, the adjoint depends on the definition of spaces  $E$  and  $F$ . For instance, a canonical injection (i.e.  $Ax = x$  with  $E$  being a subspace of  $F$ ) is not necessarily self-adjoint although  $A$  does not involve any arithmetic operation.
- In general, the adjoint depends on the choice of scalar products, even if  $E = F$ . For instance, a symmetric matrix may not be self-adjoint if the scalar product is not the canonical product (see below).

---

<sup>15</sup>or : *inner products*

<sup>16</sup>It is actually true for all continuous operators in Hilbert spaces, but this is outside the scope of this paper.

**Theorem : adjoint and scalar product change.** The operator  $A : E \rightarrow F$  being identified with its matrix, and the scalar products  $\langle \cdot, \cdot \rangle_E$  and  $\langle \cdot, \cdot \rangle_F$  being identified with their symmetric positive definite matrices  $E$  and  $F$  such that e.g.  $\langle x, x \rangle_E = x^T E x$ , the matrix of the adjoint of  $A$  is

$$A^* = E^{-1} A^T F$$

The proof is obvious from the definition of  $A^*$  :

$$\langle Ax, y \rangle_F = x^T A^T F y = \langle x, A^* y \rangle_E = x^T E A^* y$$

and noting that  $E$  is invertible. In most practical cases (such as in the rest of this paper) the implicit scalar product used is the *canonical inner product*<sup>17</sup>, so that the transpose is the adjoint :  $A^T = A^*$ . However, one must take care whenever another scalar product is used, because it has implications on the coding of the adjoint : the scalar product coefficients or their inverses must be used according to the above equation.

**Adjoint of a sequence of operators.** Like the transpose, the adjoint of a product of operators is the product of the adjoints in the reverse order. The scalar product matrices cancel out each other, so that if  $A = A_1 A_2 \cdots A_n$  is a sequence of operators, its adjoint is

$$(A_1 A_2 \cdots A_n)^* = E^{-1} A_n^T \cdots A_2^T A_1^T F$$

which shows that, even if the scalar products are not the canonical inner product, in most of the adjoint coding it can be considered that **the adjoint is the transpose**. The guidelines for practical adjoint coding are detailed in an appendix.

ref : Errico and Vukićević 1992.

---

<sup>17</sup>or : *inner dot product*, or : *euclidean product*.



### 13 Dual formulation of 3D/4D-Var (PSAS)

The 3D-Var formulation (A5) can be rewritten into a form called PSAS (Physical Space Assimilation System)<sup>18</sup> which is equivalent in the linear case only. The idea is to notice that the expression

$$\mathbf{x}_a - \mathbf{x}_b = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$

can be split as the following two equalities

$$\begin{aligned}\mathbf{w}_a &= (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}_b) \\ \mathbf{x}_a - \mathbf{x}_b &= \mathbf{B}\mathbf{H}^T\mathbf{w}_a\end{aligned}$$

where  $\mathbf{w}_a$  has the same dimension as  $\mathbf{y}$  and can be regarded as a kind of “increment” in observation space<sup>19</sup>, whereas  $\mathbf{B}\mathbf{H}^T$  is a smoothing term that maps the increment from observation to model space. The aim is to solve the analysis problem in terms of  $\mathbf{w}$  rather than in model space. One way is to solve for  $\mathbf{w}$  the linear system

$$(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})\mathbf{w} = \mathbf{y} - \mathbf{H}\mathbf{x}_b$$

which can be regarded as the dual of the OI algorithm. Another way is to find a cost function that  $\mathbf{w}_a$  minimizes, for instance

$$F(\mathbf{w}) = \mathbf{w}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})\mathbf{w} - 2\mathbf{w}^T(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$$

which is a quadratic cost function. The practical PSAS analysis algorithm is as follows :

1. Calculate the background departures  $\mathbf{y} - \mathbf{H}\mathbf{x}_b$
2. Minimize  $F(\mathbf{w})$ . Some possible preconditionings are given by the symmetric square root of  $\mathbf{R}$  or  $\mathbf{H}\mathbf{B}\mathbf{H}^T$ .
3. Multiply the minimum  $\mathbf{w}_a$  by  $\mathbf{B}\mathbf{H}^T$  to obtain analysis increments.
4. Add the increments to the background  $\mathbf{x}_b$ .

A 4-D generalization of PSAS is obtained by a suitable redefinition of the space  $\mathbf{w}$  to be a concatenation  $(\mathbf{w}_1, \dots, \dots, \mathbf{w}_i, \dots)$  of all the values  $\mathbf{w}_i$  at all observation time steps  $i$ . Then  $\mathbf{H}$  must be replaced by an operator that uses the tangent linear model  $\mathbf{M}_i$  to map the initial model state to the observation space at each time step  $i$ , i.e.  $(\mathbf{H}_1\mathbf{M}_1, \dots, \mathbf{H}_i\mathbf{M}_i, \dots)$ . The factorization of the cost function evaluation using the adjoint method is applied to the computation of the term  $\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{w}$ , so that the evaluation of the 4D-PSAS cost function  $F(\mathbf{w})$  is as follows :

1. Calculate the departures  $\mathbf{y}_i - \mathbf{H}_i\mathbf{M}_i(\mathbf{x}_b)$  for each time step, (this needs only be done once)
2. Integrate the adjoint model from final to initial time, starting with a null model state, adding the forcing  $\mathbf{H}_i^T\mathbf{w}_i$  at each observation timestep,
3. Multiply the resulting adjoint variable at initial time by  $\mathbf{B}$ , which yields  $\mathbf{B}\mathbf{H}\mathbf{w}$ ,

---

<sup>18</sup>The misleading name PSAS was introduced for historical reasons and is widely used, probably because it sounds like the US slang word *pizzazz*.

<sup>19</sup>Note, though, that it does not have the right physical dimensions. The actual increment in observation space is  $\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{w}_a$ , and a precise physical interpretation of  $\mathbf{w}_a$  is difficult.

4. Integrate the tangent-linear model, starting with  $\mathbf{B}\mathbf{H}\mathbf{w}$  as model state, storing the state times  $\mathbf{H}_i$  at each observation timestep. The collection of the stored values is  $\mathbf{H}\mathbf{B}\mathbf{H}^T\mathbf{w}$ .
5. Add  $\mathbf{w}^T\mathbf{R}\mathbf{w}$  and  $\mathbf{w}^T(\mathbf{y} - \mathbf{H}\mathbf{x}_b)$  (both obtained by sums of already computed quantities) to obtain  $F(\mathbf{w})$ .

More comments on the 4D-PSAS algorithm are provided in Courtier (1997). The PSAS algorithm is equivalent to the *representer* method (Bennett and Thorburn 1992).

As of today it is still unclear whether PSAS is superior or not to the conventional variational formulations, 3D and 4D-Var. Here are some pros and cons :

- PSAS is only equivalent to 3D/4D-Var if  $\mathbf{H}$  is linear, which means that it cannot be extended to weakly non-linear observation operators.
- However, most implementations of 3D/4D-Var are incremental, which means that they do rely on a linearization of  $\mathbf{H}$  anyway : they include non-linearity through incremental updates, which can be used identically in an incremental version of PSAS.
- It is awkward to include a  $J_c$  term in PSAS for constraints expressed in model space.
- Background error models can be implemented directly in PSAS as the  $\mathbf{B}$  operator. In 3D/4D-Var they need to be inverted (unless they are factorized and used as preconditioner).
- The size of the PSAS cost function is determined by the number of observations  $p$  instead of the dimension of the model space  $n$ . If  $p \ll n$  then the PSAS minimization is done in a smaller space than 3D/4D-Var. In a 4D-Var context,  $p$  increases with the length of the minimization period whereas  $n$  is fixed, so that this apparent advantage of PSAS may disappear.
- The conditioning of a PSAS cost function preconditioned by the square root of  $\mathbf{R}$  is identical to that of 3D/4D-Var preconditioned by the square root of  $\mathbf{B}$ . However the comparison may be altered if more sophisticated preconditionings are used, or if one square root or the other is easier to specify.
- Both 3D/4D-Var and PSAS can be generalized to include model errors. In 3D/4D-Var this means increasing the size of the control variable, which is not the case in PSAS, although the final cost of both algorithms looks the same.

ref : Bennett and Thornburn 1992, Courtier 1997.

## 14 The extended Kalman filter (EKF)

The Kalman Filter and its extended version (EKF) are developments of the least-squares analysis method in the framework of a sequential data assimilation, in which each background is provided by a forecast that starts from the previous analysis. It is adapted to the real-time assimilation<sup>20</sup> of observations distributed in time into a forecast model  $M$ .

The analysis equations of the linear Kalman Filter are exactly the ones already described in the least-squares analysis theorem. The notation is the same, except that the background (i.e. forecast) and analysis error covariance matrices are now respectively denoted  $\mathbf{P}_f$  and  $\mathbf{P}_a$ . The background state  $\mathbf{x}_b$  is a forecast denoted  $\mathbf{x}_f$ .

### 14.1 Notation and hypotheses

They are the same as in the least-squares analysis theorem, except that :

- the background and analysis error covariance matrices  $\mathbf{B}$  and  $\mathbf{A}$  are respectively replaced by  $\mathbf{P}_f$  and  $\mathbf{P}_a$  to denote the fact that the background is now a forecast.
- The time index  $i$  of each quantity is denoted by the suffix  $\bullet(i)$ . The model forecast operator from dates  $i$  to  $i + 1$  is denoted by  $M_{i \rightarrow i+1}$
- **forecast errors** : the deviation of the forecast prediction from the true evolution,  $M_{i \rightarrow i+1}[\mathbf{x}_t(i)] - \mathbf{x}_t(i + 1)$ , is called the *model error*<sup>21</sup> and we assume that it is **not biased**<sup>22</sup> and that the **model error covariance matrix**  $\mathbf{Q}(i)$  is known.
- **uncorrelated analysis and model errors** : the analysis errors  $\mathbf{x}_a(i) - \mathbf{x}_t(i)$  and model errors of the subsequent forecast  $M_{i \rightarrow i+1}[\mathbf{x}_t(i)] - \mathbf{x}_t(i + 1)$  are assumed to be mutually uncorrelated.
- **linearized forecast operator** : the variations of the model prediction in the vicinity of the forecast state are assumed to be a linear function of the initial state : for any  $\mathbf{x}(i)$  close enough to  $\mathbf{x}_a(i)$ ,  $M_{i \rightarrow i+1}[\mathbf{x}(i)] - M_{i \rightarrow i+1}[\mathbf{x}_a(i)] = \mathbf{M}_{i \rightarrow i+1}[\mathbf{x}(i) - \mathbf{x}_a(i)]$  where  $\mathbf{M}$  is a linear operator.

---

<sup>20</sup>The word *filter* characterizes an assimilation techniques that uses only observation from the past to perform each analysis. An algorithm that uses observations from both past and future is called a *smoother*. 4D-Var can be regarded as a smoother. Observation smoothing can be useful for non-real time data assimilation, e.g. reanalysis, although the idea has not been used much yet. The Kalman filter has a smoother version called *Kalman smoother*.

<sup>21</sup>Or *modelling error*.

<sup>22</sup>This is equivalent to assuming that the background errors are unbiased, so it is not really a new hypothesis.

## 14.2 Theorem : the KF algorithm

Under the specified hypotheses the optimal way (in the least squares sense) to assimilate sequentially the observations is given by the *Kalman filter algorithm* defined below by recurrence over the observation times  $i$  :

State forecast	$\mathbf{x}_f(i+1) = M_{i \rightarrow i+1} \mathbf{x}_a(i)$	(KF1)
Error covariance forecast	$\mathbf{P}_f(i+1) = \mathbf{M}_{i \rightarrow i+1} \mathbf{P}_a(i) \mathbf{M}_{i \rightarrow i+1}^T + \mathbf{Q}(i)$	(KF2)
Kalman gain computation	$\mathbf{K}(i) = \mathbf{P}_f(i) \mathbf{H}^T(i) [\mathbf{H}(i) \mathbf{P}_f(i) \mathbf{H}^T(i) + \mathbf{R}(i)]^{-1}$	(KF3)
State analysis	$\mathbf{x}_a(i) = \mathbf{x}_f(i) + \mathbf{K}(i) [\mathbf{y}(i) - H(i) \mathbf{x}_f(i)]$	(KF4)
Error covariance of analysis	$\mathbf{P}_a(i) = [\mathbf{I} - \mathbf{K}(i) \mathbf{H}(i)] \mathbf{P}_f(i)$	(KF5)

and the analyses are the sequences of  $\mathbf{x}_a(i)$ .

### Démonstration:

The forecast equation (KF1) just translates the fact that we use the model  $M$  to evolve the model state, starting from the previous optimal analysis  $\mathbf{x}_a$ . The equation (KF2) is obtained by first subtracting  $\mathbf{x}_t(i+1)$  from (KF1) and using the linearity of the forecast operator :

$$\mathbf{x}_f(i+1) - \mathbf{x}_t(i+1) = \mathbf{M}_{i \rightarrow i+1} [\mathbf{x}_a(i) - \mathbf{x}_t(i)] + [M_{i \rightarrow i+1} \mathbf{x}_t(i) - \mathbf{x}_t(i+1)]$$

Multiplying it on the right by its transpose and taking the expectation of the result yields, by definition,  $\mathbf{P}_f(i+1)$  on the left-hand side, and on the right-hand side four terms. Two of these are  $\mathbf{M}_{i \rightarrow i+1} \mathbf{P}_a(i) \mathbf{M}_{i \rightarrow i+1}^T$  and  $\mathbf{Q}(i)$  by definition. The remaining two terms are cross-correlations between the analysis error  $\mathbf{x}_a(i) - \mathbf{x}_t(i)$  and the model error for  $M_{i \rightarrow i+1}$ , which are assumed to be zero. This means that  $\mathbf{P}_f(i+1)$  provided by (KF2) is the background error covariance matrix for the analysis at time  $i+1$ .

The equations (KF3), (KF4) and (KF5) are simply the least-square analysis equations (A2), (A1) and (A4) that were proven above, using  $\mathbf{P}_f(i)$  as background errors, and assuming that  $\mathbf{K}$  is computed optimally.

## 14.3 Theorem : KF/4D-Var equivalence

Over the same time interval  $[0, N]$  assuming that  $\mathbf{Q} = 0$  (i.e. the model is perfect), and that both algorithms use the same data (notably,  $\mathbf{P}_f(0)$  is the initial background error covariance matrix), then there is equality between

1. the final analysis  $\mathbf{x}_a(N)$  produced by the above Kalman filter algorithm, and
2. the final value of the optimal trajectory estimated by 4D-Var, i.e.  $M_{0 \rightarrow N} \mathbf{x}_a(0)$ .

This theorem means that the KF verifies the four-dimensional least-squares optimality theory expressed by the 4D-Var cost function, although it is defined by a sequence of 3-D analyses, whereas 4D-Var solves the 4-D problem globally.

## 14.4 The Extended Kalman Filter (EKF)

The Kalman filter algorithm can be generalized to non-linear  $H$  and  $M$  operators, although it means that neither the optimality of the analysis nor the equivalence with 4D-Var hold in

that case. If  $H$  is non-linear,  $\mathbf{H}$  can be defined as its tangent linear in the vicinity of  $\mathbf{x}_b$ , as discussed in a previous section. Similarly, if  $M$  is non-linear, which is the case of most meteorological and oceanographical models,  $\mathbf{M}$  can be defined as the *tangent linear forecast model* in the vicinity of  $\mathbf{x}_a$ , i.e. we assume that for any likely initial state  $\mathbf{x}(i)$  (notably  $\mathbf{x}_t(i)$ ),

$$M_{i \rightarrow i+1}[\mathbf{x}(i)] - M_{i \rightarrow i+1}[\mathbf{x}_a(i)] \simeq \mathbf{M}_{i \rightarrow i+1}[\mathbf{x}(i) - \mathbf{x}_a(i)]$$

and the realism of this hypothesis must be appreciated using physical arguments, as already discussed about the observation operator and 4D-Var. If  $H$  and/or  $M$  are non-linear, the algorithm written above is called the *Extended Kalman Filter*. Note that the linearization of  $M$  interacts with the model errors in a possibly complicated way, as can be seen from the proof of equation (KF2) above. If non-linearities are important, it may be necessary to include empirical correction terms in the equation, or to use a more general stochastic prediction method such as an *ensemble prediction* (or *Monte Carlo*) method, which yields an algorithm known as the *Ensemble Kalman Filter*.

## 14.5 Comments on the KF algorithm

The input to the algorithms is : the definition of the model and the observation operator, the initial condition for  $(\mathbf{x}, \mathbf{P})$  when the recurrence of the filter is started<sup>23</sup>, the sequence of observations  $\mathbf{y}$ , and the sequence of model and observation error covariance matrices  $(\mathbf{Q}, \mathbf{R})$ . The output is the sequence of estimates  $(\mathbf{x}_a, \mathbf{P}_a)$  of the model state and its error covariance matrix. The organization of the KF assimilation looks like a coupled stream of estimations of model states and error covariances (figure 16).

The variational form of the least-squares analysis can be used in the analysis step of the Kalman filter, instead of the explicit equations written above.

The numerical cost of the KF or EKF is that of the analysis itself, plus the estimation of the analysis error covariances, discussed in a specific section, plus the (KF2) covariance forecast equation which requires  $n$  forecasts of the tangent linear model ( $n$  being the dimension of the model state) to build the operator  $\mathbf{M}$ . The storage cost itself is significant, since each  $\mathbf{P}$  matrix is  $n \times n$  (only a half can be stored since they are symmetric) and in (KF5) the  $\mathbf{KH}$  matrix must be evaluated and stored too (unless the variational form is used, in which case  $n$  evaluations of the gradient of the cost function must be performed to build the Hessian which must then be inverted). It means that the cost of the KF is much larger than 4D-Var, even with small models. The algorithm should rather be regarded as a reference in the design of more approximate assimilation algorithms which are being developed nowadays. It is still not clear what is the best way to approximate the KF, and the answer will probably be application-dependent.

There are many similarities between 4D-VAR and the EKF and it is important to understand the fundamental differences between them :

- 4D-VAR can be run for assimilation in a realistic NWP framework because it is computationally much cheaper than the KF or EKF.

---

<sup>23</sup>Note that it is not well known whether, after a long time, the analysis ceases or not to depend significantly on the way the KF is initialized.

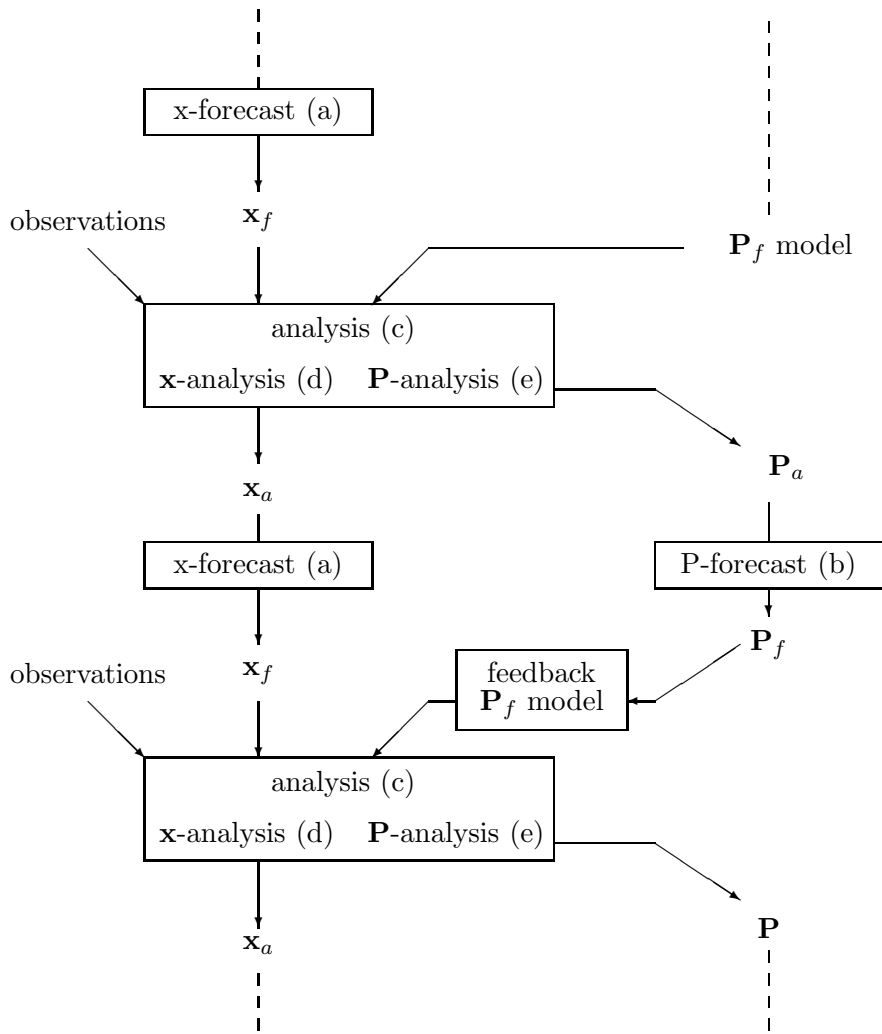


FIG. 16 – The organization of computations in a KF or EKF assimilation.

- 4D-VAR is more optimal than the (linear or extended) KF inside the time interval for optimization because it uses all the observations at once, i.e. it is not sequential, it is a smoother.
- unlike the EKF, 4D-VAR relies on the hypothesis that the model is perfect (i.e.  $\mathbf{Q} = 0$ ).
- 4D-VAR can only be run for a finite time interval, especially if the dynamical model is non-linear, whereas the EKF can in principle be run forever.
- 4D-VAR itself does not provide an estimate of  $\mathbf{P}_f$ , a specific procedure to estimate the quality of the analysis must be applied, which costs as much as running the equivalent EKF.

ref : Ghil 1989, Lacarra and Talagrand 1988, Errico et al 1993.

## 15 Conclusion

This presentation of analysis algorithms has been centered on the algebra of the least-squares analysis method. However one shall not forget the importance of other issues like observation screening and physical consistency of the assimilation, including bias correction, which can be of great importance for the quality of the assimilation system taken as a whole.

The recent trend in data assimilation is to combine the advantages of 4D-Var and the Kalman filter techniques. In a real-time assimilation system, 4D-Var over a short time interval is a very efficient analysis method. A Hessian estimation method can provide a good estimate of the analysis error covariance matrix. A simplified version of the extended Kalman filter forecast step is then used (SKF) to estimate the forecast error covariances at the time of the next analysis, which must then be combined with an empirical, more static model of the background error covariances. It is hoped that a good compromise between these algorithms can be achieved. There can be some constructive interactions with the problems of ensemble prediction, and specific studies of analysis quality like sensitivity studies and observation targeting. These new methods provide many by-products which still remain to be used as diagnostic tools for improving the assimilation and forecast system.

## 16 References

The references below have been chosen because of their educational value. It does not necessarily mean that they are the original historical references. These can usually be found in the reference lists of the paper below.

- Bennett, A. and M. Thornburn, 1992 : The generalized inverse of a non-linear quasi-geostrophic ocean circulation model. *J. Phys. Oceanogr.*, **3**, 213-230.
- Derber, J. and F. Bouttier, 1999 : A reformulation of the background error covariance in the ECMWF global data assimilation system. *Accepted for publication in Tellus*.
- Courtier, P., J.-N. Thépaut and A. Hollingsworth, 1994 : A strategy for operational implementation of 4D-VAR, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367-1387.
- Courtier, P., 1997 : Dual formulation of four-dimensional variational assimilation. *Quart. J. Roy. Met. Soc.*, **123**, 2449-2461.
- Courtier, P., E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier and M. Fisher, 1998 : The ECMWF implementation of three-dimensional variational assimilation (3D-Var). Part 1 : formulation. *Quart. J. Roy. Met. Soc.*, **124**, 1783-1807.
- Daley, R., 1991 : Atmospheric Data Analysis. *Cambridge Atmospheric and Space Science Series*, Cambridge University Press. ISBN 0-521-38215-7, 457 pages.
- Errico, R. and T. Vukićević, 1992 : Sensitivity Analysis using an Adjoint of the PSU-NCAR Mesoscale Model. *Mon. Wea. Rev.*, **120**, 1644-1660.
- Errico, R., T. Vukićević and K. Raeder, 1993 : Examination of the accuracy of a tangent linear model. *Tellus*, **45A**, 462-477.
- Eyre, J., 1987 : Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation : theory and simulation for TOVS. *Quart. Jour. Roy. Met. Soc.*, **113**.
- Ghil, M., 1989 : Meteorological Data Assimilation for Oceanographers. Part I : Description and Theoretical Framework. *Dyn. of Atm. and Oceans*, **13**, 171-218.
- Gilbert, J.-C. and C. Lemaréchal, 1989 : Some numerical experiments with variable storage quasi-Newton algorithms. *Mathematical Programming*.
- Hollingsworth, A., and P. Lönnberg, 1986 : The statistical structure of short-range forecast errors as determined from radiosonde data. Part I : The wind field. *Tellus*, **38A**, 111-136.
- Hollingsworth, A., D. Shaw, P. Lönnberg, L. Illari, K. Arpe and A. Simmons, 1986 : Monitoring of observation and analysis quality by a data-assimilation system. *Mon. Wea. Rev.*, **114**, 1225-1242.
- Lacarra, J.-F., and O. Talagrand, 1988 : Short-Range Evolution of Small Perturbations in a Barotropic Model. *Tellus*, **40A**, 81-95.
- Lorenc, A., 1986 : Analysis methods for numerical weather prediction. *Quart. J. Roy. Met. Soc.*, **112**, 1177-1194.
- Lorenc, A., 1981 : A Global Three-Dimensional Multivariate Statistical Interpolation Scheme. *Mon. Wea. Rev.*,



- Parrish, D. and J. Derber, 1992 : The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747-1763.
- Rabier, F., and P. Courtier, 1992 : Four-dimensional assimilation in the presence of baroclinic instability. *Quart. J. Roy. Meteor. Soc.*, **118**, 649-672.
- Talagrand, O. and P. Courtier, 1987 : Variational assimilation of meteorological observations with the adjoint vorticity equation. I : Theory. *Quart. J. Roy. Meteor. Soc.*, **113**, 1311-1328.
- Thépaut, J.-N. and P. Courtier, 1991 : Four-dimensional data assimilation using the adjoint of a multilevel primitive-equation model. *Quart. J. Roy. Meteor. Soc.*, **117**, 1225-1254.
- Vukicevic, T., 1991 : Nonlinear and Linear Evolution of Initial Forecast Errors. *Mon. Wea. Rev.*, **119**, 1602-1611.

## A Écriture d'un code adjoint

A partir du code d'un opérateur linéaire (dit *code direct*), construire le code adjoint ne présente pas de réelle difficulté. Il existe des logiciels (payants) pour effectuer cette tâche, mais ils posent souvent des problèmes d'efficacité. Lorsque les produits scalaires en jeu sont canoniques, l'adjoint est simplement la transposée du code direct. Une méthode, simple à coder mais chère dans des grands espaces, consiste à construire la matrice (dite jacobienne) de l'opérateur, et à appliquer sa transposée dans le code adjoint. L'autre méthode consiste à décomposer le code direct en séquence d'opérateurs linéaires élémentaires (environ un par ligne de code active), et à construire le code adjoint comme séquence inverse des transposées de ces opérateurs (la transposée d'une composition d'opérateurs étant la composée des opérateurs transposés, mais en ordre inverse). Il est essentiel de bien définir les espaces d'entrée et de sortie de chaque opérateur. La déclaration d'une nouvelle variable dans le code direct correspond à l'injection (dite canonique) de l'espace de travail dans un espace plus grand. Inversement, la fin de vie d'une variable dans le code direct (par exemple, une variable locale lorsque l'on quitte un sous-programme) correspond à une projection dans l'espace orthogonal à cette variable. Projections et injections canoniques sont adjointes l'une de l'autre. Exemple : lorsque l'espace de travail est  $(a, b)$  et que l'on cesse d'utiliser  $b$ , cela correspond à l'opérateur linéaire

$$(a) = (1 \ 0) \begin{pmatrix} a \\ b \end{pmatrix}$$

dont l'adjoint est :

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} (a)$$

en Fortran, cela correspond à l'opération  $\mathbf{b}=0$ . Si on l'oublie, on risque d'avoir un  $b$  mal initialisé dans le code.

Une affectation  $\mathbf{a}=\mathbf{b}$  est, en toute généralité, l'opération

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} (b)$$

dont l'adjoint est

$$(b) = (11) \begin{pmatrix} a \\ b \end{pmatrix}$$

autrement dit c'est la ligne de code  $\mathbf{b}=\mathbf{b}+\mathbf{a}$ . Si dans le code direct c'est la dernière ligne où l'on utilise  $b$ , il faut faire précéder cette ligne du code adjoint de  $\mathbf{b}=0$ . Autrement dit, c'est équivalent à  $\mathbf{b} = \mathbf{a}$ , mais en pratique il est recommandé de **toujours coder les adjoints d'opérations sous forme additive, et de toujours initialiser les variables nouvellement allouées à zéro dans le code adjoint**. Cela permet de coder l'adjoint en travaillant 'ligne à ligne', sans se demander tout le temps si chaque ligne de code est la dernière ou non où l'on utilise une variable (les langages informatiques faisant rarement apparaître dans le code source l'endroit où une variable cesse d'exister, alors qu'on voit souvent où une variable est déclarée pour la première fois).

Autre exemple : l'adjoint de  $\mathbf{a}=\mathbf{s}*\mathbf{b}+\mathbf{t}*c$  ( $\mathbf{s}$  et  $\mathbf{t}$  sont des constantes) est

```

a=0      ! là où a est définie pour la première fois dans le code adjoint
b=0      ! .. b ..
c=0      ! .. c ..
...
b=b+s*a
c=c+t*a

```

Si  $a, b, c$  sont les arguments d'entrée d'un sous-programme, leur valeur est initialisée par le programme appelant.

Autre exemple : l'adjoint de  $a=s*a+t*c$  est

```

c=0
...
a=s*a      ! et non a=a+s*a !
c=c+t*a

```

puisque  $a$  est à la fois dans l'espace d'entrée et de sortie de l'opération.

L'adjoint de

```

if (a>0) then
  a=2*a
endif

```

n'existe pas, puisque cet opérateur n'est pas linéaire en  $a$ . Il faut le linéariser d'abord. La linéarisation d'un opérateur non-linéaire fait apparaître des coefficients qui ne font pas partie de l'espace de contrôle, mais dépendent de l'état du modèle, par exemple le linéarisé de  $b = a^2$  est  $b' = 2a'a$ . Il faut donc stocker la valeur de  $a$  et la fournir au bon endroit du code direct, et adjoint  $a' = a' + 2ab'$ .

## B Principaux symboles utilisés

$\mathbf{x}$  vecteur d'état du modèle

$\mathbf{x}_t$  valeur vraie de l'état du modèle, autrement dit analyse parfaite

$\mathbf{x}_b$  ébauche de l'état du modèle

$\mathbf{x}_a$  analyse de l'état du modèle

$\mathbf{y}$  vecteur d'observations

$H$  opérateur d'observation (application de l'espace de  $\mathbf{x}$  dans celui de  $\mathbf{y}$ , qui fournit les équivalents modèle des valeurs observées)

$\mathbf{H}$  opérateur d'observation linéarisé (au voisinage d'un état modèle prédéfini)

$\mathbf{B}$  matrice des covariances d'erreurs d'ébauche (erreurs d'estimation de  $\mathbf{x}_b$ )

**A** matrice des covariances d'erreur d'analyse (erreurs d'estimation de  $\mathbf{x}_a$ )  
**R** matrice des covariances d'erreur dites d'observation (des quantités  $\mathbf{y} - H(\mathbf{x}_t)$ )  
**K** matrice de gain (ou : des poids) de l'analyse  
**I** opérateur identité dans l'espace courant  
 $J$  fonction coût de l'analyse variationnelle  
 $J_b$  terme d'ébauche de la fonction coût  
 $J_o$  terme d'observation de la fonction coût  
 $J_c$  terme de pénalisation de la fonction coût