



Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique

---

# Introduction à l'assimilation de données ; applications actuelles dans les projets du CERFACS

Joaquín Muñoz Sabater, Sébastien Massart, Sophie Ricci et Bertrand  
Bouriquet

*CERFACS, 42 avenue Coriolis  
31057, Toulouse, France  
massart@cerfacs.fr*

Technical Report TR/CMGC/08/129

Note:

**CERFACS**

**42, Avenue Gaspard Coriolis, 31057 Toulouse Cedex 1, FRANCE.**

**Tel. : 33 (0) 561 19 31 31 – Fax : 33 (0) 561 19 30 30**

**secretar@cerfacs.fr – <http://www.cerfacs.fr>**

# Introduction à l'assimilation de données ; applications actuelles dans les projets du CERFACS

Joaquín Muñoz Sabater, Sébastien Massart, Sophie Ricci et Bertrand Bouriquet

3 décembre 2008

# Table des matières

<b>1</b>	<b>Introduction à l'assimilation de données</b>	<b>3</b>
1.1	Pourquoi utiliser l'assimilation de données? . . . . .	3
1.1.1	La modélisation mathématique et ses limites . . . . .	3
1.1.2	L'objectif de l'assimilation de données. . . . .	3
1.1.3	L'assimilation de données et l'estimation linéaire optimale. . . . .	3
1.2	L'assimilation de données . . . . .	4
1.2.1	Concepts de base . . . . .	5
1.2.2	L'étape d'analyse . . . . .	7
1.2.3	L'exemple du pendule . . . . .	11
1.3	La philosophie de quelques méthodes d'assimilation de données . . . . .	14
1.3.1	Interpolation optimale (OI) . . . . .	14
1.3.2	Filtre de Kalman . . . . .	15
1.3.3	3D-VAR . . . . .	16
1.3.4	3D-VAR FGAT . . . . .	16
1.3.5	4D-VAR . . . . .	17
<b>2</b>	<b>Les outils pour l'assimilation de données</b>	<b>18</b>
2.1	PALM . . . . .	18
2.2	Les algorithmes de minimisation . . . . .	18
2.2.1	Méthode du gradient . . . . .	18
2.2.2	Gradient conjugué . . . . .	19
2.2.3	Perspectives . . . . .	20
2.3	L'assimilation de données pour le suivi d'espèces chimiques atmosphériques au CER- FACS . . . . .	20
2.3.1	Introduction . . . . .	20
2.3.2	Le schéma d'assimilation . . . . .	20
2.3.3	Perspectives . . . . .	22
2.4	Assimilation de données pour la neutronique au CERFACS . . . . .	22
2.4.1	Introduction . . . . .	22
2.4.2	La maquette MANARA . . . . .	23
2.4.3	La maquette KAFEINE . . . . .	24
2.5	Une approche pour l'assimilation appliquée à l'hydrologie et à l'hydraulique fluviale	24
2.5.1	Introduction . . . . .	24
2.5.2	Initialisation de l'humidité du sol pour les modèles de prévision numérique	25
2.5.3	Caractérisation de l'état des rivières . . . . .	25
<b>3</b>	<b>L'assimilation de données dans les Services de Prévision de Crues</b>	<b>28</b>
<b>A</b>	<b>Algorithmes de minimisation</b>	<b>31</b>
A.1	Méthode du gradient . . . . .	31
A.2	Gradient conjugué . . . . .	32
<b>B</b>	<b>Tableau-résumé de l'assimilation au CERFACS</b>	<b>32</b>
	<b>Bibliographie</b>	<b>33</b>

# 1 Introduction à l'assimilation de données

L'assimilation de données est une technique d'analyse mathématique dont le but est la description de l'état d'un système physique de façon la plus réaliste possible. Cette technique utilise les équations discrétisées de la physique ainsi que des observations sur le système. Historiquement, l'assimilation de données a été développée pour répondre aux besoins opérationnels de la prévision météorologique et des analyses de l'état de l'océan. Son application actuelle au CERFACS concerne d'autres champs aussi divers que l'hydrologie, la neutronique, la chimie ou la mécanique des structures.

Ce document constitue une introduction à l'assimilation de données. Dans celui-ci, il est proposé une synthèse des méthodes d'assimilation et de leurs applications au CERFACS. Enfin, un tableau propose un récapitulatif de l'utilisation de ces méthodes par le SCHAPI et les Services de Prévision de Crues (SPC) à ce jour.

## 1.1 Pourquoi utiliser l'assimilation de données ?

### 1.1.1 La modélisation mathématique et ses limites

Une majorité des systèmes physiques qui nous intéressent sont d'une telle complexité que leur traduction par les équations d'un modèle mathématique discrétisé ne peut se faire que de façon approximative et à l'aide de plusieurs hypothèses simplificatrices. De plus, les estimations des modèles numériques sont contraintes par :

1. des conditions initiales,
2. des conditions aux limites du domaine physique en étude,
3. des paramètres qui décrivent les caractéristiques physiques de ce système.

Par exemple, pour un modèle décrivant les interactions entre les surfaces continentales et la couche limite atmosphérique, les conditions initiales sont données par la température et l'humidité initiale du sol. Le modèle numérique utilise un jeu de paramètres qui traduisent les propriétés physiques du sol (contenu en argile et sable, capacité de champ, coefficient de rugosité, etc.) et de la végétation alentours (résistance stomatique, conductance mésophylliale, etc.). L'estimation de ces paramètres est très couteuse à l'échelle de la parcelle et très complexe à plus grandes échelles.

En conséquence, aux erreurs causées par la modélisation mathématique s'ajoute une erreur non négligeable associée à la méconnaissance du jeu des conditions initiales, des conditions aux limites et des paramètres initiaux. Aussi on constate que la simulation numérique fournit fréquemment des résultats éloignés de la réalité physique du système.

### 1.1.2 L'objectif de l'assimilation de données.

Il est possible de diminuer l'incertitude des sorties des simulations numériques et d'améliorer notre connaissance sur le système physique, en exploitant d'autres sources d'information complémentaire sur le système. C'est l'objectif de l'assimilation de données. Généralement l'information complémentaire provient des observations (par des instruments sur la surface, par des capteurs, des satellites) ainsi que d'une estimation *a priori* de l'état du système physique d'étude.

Typiquement, l'assimilation de données permet d'optimiser directement les champs physiques estimés par le modèle mathématique (FIG. 1) ou de faire un recalage des paramètres d'entrée et/ou de la condition initiale pour s'approcher des observations (FIG. 2).

### 1.1.3 L'assimilation de données et l'estimation linéaire optimale.

La théorie de l'assimilation de données est fortement liée à la théorie d'estimation linéaire optimale. Nous allons nous appuyer sur ce lien pour expliquer en quoi consiste l'assimilation de données.

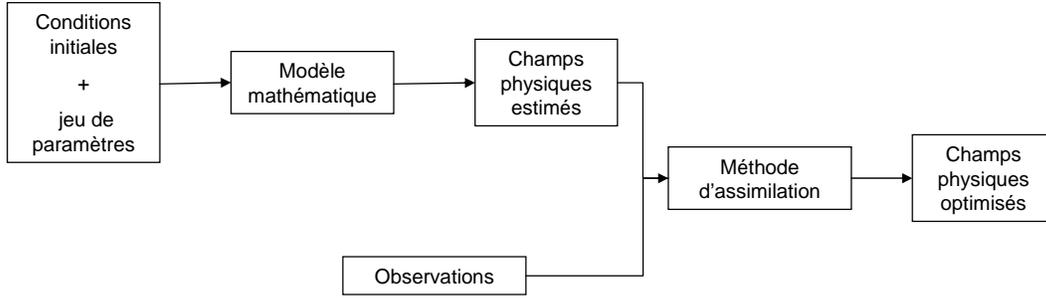


FIG. 1 – Schéma d’assimilation de données appliqué à l’optimisation directe sur les estimations brutes des champs physiques modélisés.

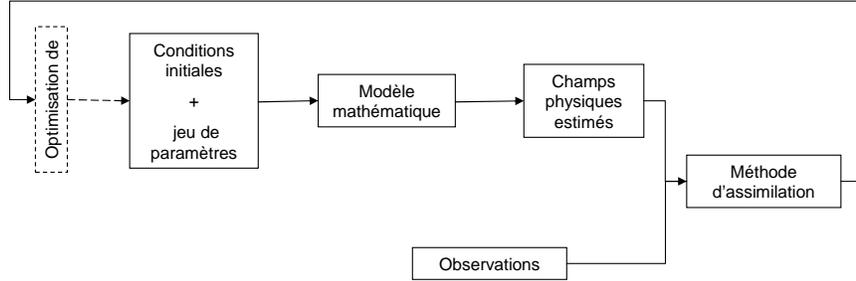


FIG. 2 – Principe du recalage de la condition initiale et des paramètres d’entrées d’un modèle mathématique pour l’assimilation de données.

Supposons par exemple que l’on dispose de deux informations sur la température au centre de Paris :

1. une valeur prévue par le modèle. On la note  $\mathbf{x}_b$ ,
2. une observation, notée  $\mathbf{y}$ ,

On se pose la question suivante : Compte tenu de ces deux informations, quelle est la meilleure estimation possible  $\mathbf{x}_a$  de la température au centre de Paris ? Trois cas possibles sont envisagés :

- a. Si le modèle est très précis :  $\mathbf{x}_a \sim \mathbf{x}_b$
- b. Si l’observation est très précise :  $\mathbf{x}_a \sim \mathbf{y}$
- c. Si le modèle et l’observation ont la même précision :  $\mathbf{x}_a \sim (\mathbf{y} + \mathbf{x}_b)/2$

On recherche ici le meilleur compromis entre les observations et le modèle, en tenant compte de leurs précisions respectives. Cela peut être formalisé par la théorie de l’estimation linéaire : l’état optimal  $\mathbf{x}_a$  est une moyenne pondérée entre le modèle  $\mathbf{x}_b$  et l’observation  $\mathbf{y}$ .  $\mathbf{x}_a$  s’écrit :

$$\mathbf{x}_a = (1 - \mathbf{K})\mathbf{x}_b + \mathbf{K}\mathbf{y} \quad (1)$$

où  $(1-\mathbf{K})$  et  $\mathbf{K}$  sont les poids respectifs de  $\mathbf{x}_b$  et de  $\mathbf{y}$ . La formulation de  $\mathbf{K}$  diffère selon les méthodes d’assimilation de données. Dans la section suivante, on présentera la formulation de  $\mathbf{K}$  telle que l’analyse  $\mathbf{x}_a$  soit optimale.

## 1.2 L’assimilation de données

Cette section présente les notions de base qui permettront à tout lecteur de s’initier aux techniques d’assimilation de données : les concepts, les méthodes d’assimilation et l’idée fondatrice derrière les méthodes utilisées actuellement au CERFACS. Ces concepts sont illustrés dans le cadre d’une étude d’assimilation appliquée à l’hydrologie. La mise en œuvre de ces concepts est présentée sur l’exemple simple du pendule oscillant en section 1.2.3.

### 1.2.1 Concepts de base

**Contrôle.** Supposons qu'à l'exutoire d'un bassin-versant on veuille contrôler à tout moment le débit de la rivière (noté  $\mathbf{Q}$ ) et l'humidité du sol aux alentours (notée  $\mathbf{w}$ ). Supposons de plus que ces deux variables décrivent complètement notre système. On rassemble ces variables dans le vecteur  $[\mathbf{Q}, \mathbf{w}]^T$  que l'on appelle *vecteur de contrôle* (noté  $\mathbf{x}$ ). De plus, on va supposer que l'on a un instrument qui nous permet de mesurer également le débit de la rivière (noté  $\mathbf{Q}_O$ ).

**Ébauche, observations et analyse.** L'objectif que nous poursuivons est la description optimale de l'état d'un système physique à partir :

1. d'une estimation *a priori* des variables qui décrivent le système étudié (généralement elle est donnée par une première simulation du modèle discrétisé). En assimilation de données on appelle aussi cet état *ébauche* et on le note  $\mathbf{x}_b$  ( $\mathbf{b}$  de l'anglais "background"). Dans notre exemple l'ébauche est constituée par deux variables,  $\mathbf{Q}$  et  $\mathbf{w}$ , et donc le vecteur d'ébauche est noté :  $[\mathbf{Q}_b, \mathbf{w}_b]^T$
2. des observations disponibles. On suppose que l'on dispose d'un instrument qui nous permet de mesurer le débit de la rivière (noté  $\mathbf{Q}_O$ ). Le vecteur qui rassemble toutes les observations est noté  $\mathbf{y}$ .

Le résultat du processus d'assimilation de données est appelé *analyse* et est noté  $\mathbf{x}_a$ . L'étape d'analyse consiste à mettre à jour les variables, les paramètres, les conditions initiales, etc., que l'on a choisi comme représentatifs de notre système physique d'étude pour s'approcher des observations disponibles. L'étape d'analyse prend aussi en compte toutes les sources d'erreur. Dans notre exemple, le vecteur d'analyse est  $[\mathbf{Q}_a, \mathbf{w}_a]^T$ .

**Opérateur d'observation.** Une étape cruciale de l'assimilation de données consiste à calculer l'écart entre l'ébauche et les observations. Cet écart sert par la suite à corriger la valeur des variables de contrôle. Usuellement, l'ébauche et les observations ne sont pas directement comparables car elles sont définies dans des espaces différents. Par exemple, imaginons que l'on veuille analyser la température de la surface des océans. D'une part on dispose de l'ébauche fournie par un modèle océanique (par exemple, un champ 3D de températures), et d'autre part on a des observations provenant de satellites défilants. Cependant, les capteurs des satellites ne mesurent pas des températures, mais des radiances. Il est donc indispensable d'utiliser un opérateur pour projeter les températures dans l'espace des radiances. Cet opérateur est appelé *opérateur d'observation*, il est à priori non-linéaire et on le note  $\mathcal{H}$  (sa formulation linéarisée est notée  $\mathbf{H}$ ). Ainsi, l'ébauche  $\mathbf{x}_b$  et les observations  $\mathbf{y}$  sont liées à travers l'expression :

$$\mathbf{y} = \mathcal{H}(\mathbf{x}_b) \quad (2)$$

où  $\mathcal{H}(\mathbf{x}_b)$  est la projection de l'état  $\mathbf{x}_b$  sur l'espace des observations. Dans notre exemple le débit est une variable de contrôle et observée au temps du modèle. En conséquence, l'opérateur d'observation est :  $\mathbf{H} = [1, 0]$

**Incertitudes.** Il est primordial de comprendre que l'assimilation de données effectue l'analyse en prenant en compte les erreurs affectées à chaque information décrivant le système. L'estimation de l'analyse est faite en considérant que l'ébauche et les observations ne sont pas parfaites et donc associées à une erreur (ou incertitude). L'erreur liée à l'ébauche s'écrit :

$$\boldsymbol{\epsilon}_b = \mathbf{x}_b - \mathbf{x}_t. \quad (3)$$

où  $\mathbf{x}_t$  est l'état vrai. L'état vrai étant inconnu, l'erreur d'ébauche est elle aussi inconnue. On en propose donc une estimation statistique. Dans notre exemple l'erreur d'ébauche est la différence

entre l'ébauche  $[\mathbf{Q}_b, \mathbf{w}_b]^T$  et les vrais débits et humidités inconnus.

L'erreur liée aux observations s'écrit :

$$\boldsymbol{\epsilon}_y = \mathbf{y} - \mathcal{H}(\mathbf{x}_t). \quad (4)$$

où  $\boldsymbol{\epsilon}_y$  est une variable aléatoire qui représente l'écart entre les observations et l'état vrai du système au travers de l'opérateur d'observation  $\mathcal{H}$ .

**Matrices d'erreurs.** Les techniques d'assimilation de données s'attachent à réduire l'écart de l'état du système à l'état vrai. L'information pertinente pour ces techniques est contenue dans les variances et les covariances des erreurs d'ébauche et d'observations. Les covariances des erreurs liées à l'ensemble des variables de contrôle s'écrivent sous forme matricielle, cette matrice notée  $\mathbf{B}$  est la matrice de covariance d'erreur d'ébauche :

$$\mathbf{B} = \begin{bmatrix} \text{VAR}(\epsilon_1) & \dots & \text{COV}(\epsilon_1, \epsilon_n) \\ \vdots & \text{VAR}(\epsilon_j) & \vdots \\ \text{COV}(\epsilon_n, \epsilon_1) & \dots & \text{VAR}(\epsilon_n) \end{bmatrix} \quad (5)$$

Notons que les éléments diagonaux de la matrice  $\mathbf{B}$  correspondent aux variances de l'erreur liée aux variables de contrôle, tandis que les termes non-diagonaux représentent les covariances entre les erreurs liées à deux variables de contrôle distinctes.

Dans notre exemple, si l'on suppose que l'erreur commise pour le débit est de 10% et pour l'humidité de 4%, la matrice d'erreur d'ébauche s'écrit :

$$\mathbf{B} = \begin{bmatrix} (0.1 \cdot \mathbf{Q})^2 & (0.1\mathbf{Q} \cdot 0.04\mathbf{w}) \\ (0.1\mathbf{Q} \cdot 0.04\mathbf{w}) & (0.04 \cdot \mathbf{w})^2 \end{bmatrix} \quad (6)$$

Les covariances des erreurs liées aux observations s'écrivent aussi sous forme matricielle, cette matrice notée  $\mathbf{R}$  est la matrice de covariance d'erreur d'observation :

$$\mathbf{R} = \begin{bmatrix} \text{VAR}(\epsilon_1) & \dots & \text{COV}(\epsilon_1, \epsilon_p) \\ \vdots & \text{VAR}(\epsilon_j) & \vdots \\ \text{COV}(\epsilon_p, \epsilon_1) & \dots & \text{VAR}(\epsilon_p) \end{bmatrix} \quad (7)$$

Les éléments diagonaux de la matrice  $\mathbf{R}$  correspondent aux variances de l'erreur liée aux observations aux points d'observation. Les termes non-diagonaux représentent les covariances entre les erreurs liées à des observations distinctes. Dans notre exemple, cette matrice se réduit à un scalaire car on n'a qu'une observation :  $\mathbf{R} = [\sigma_{\mathbf{Q}_0}^2]$ .

Dans les matrices (5) et (7) les indices 1,2,...,j,...,p,.. font référence aux variables de contrôle ou aux observations. Noter que la matrice  $\mathbf{B}$  est de dimension  $n \times n$  (avec  $n$  le nombre de variables de contrôle) et la matrice  $\mathbf{R}$  est de dimension  $p \times p$  (avec  $p$  le nombre d'observations). En général  $n \neq p$ .

**Erreur d'analyse.** De la même façon que l'on a associé une erreur à l'ébauche et aux observations, l'analyse est elle-même entachée d'une erreur. Cette erreur est définie comme l'écart entre l'analyse  $\mathbf{x}_a$  et l'état vrai  $\mathbf{x}_t$ . C'est cette erreur que l'assimilation de données s'attache à réduire :

$$\boldsymbol{\epsilon}_a = \mathbf{x}_a - \mathbf{x}_t \quad (8)$$

On note  $\mathbf{A}$  la matrice de covariance d'erreur de l'analyse, cette matrice contient toute l'information sur les incertitudes des variables analysées. Etant donné que cette matrice exprime les erreurs des variables de contrôle corrigées par l'analyse, sa dimension est la même que celle de la matrice  $\mathbf{B}$  ( $n \times n$ ). Pour notre exemple, cette matrice s'écrit :

$$\mathbf{A} = \begin{bmatrix} \sigma_{\mathbf{Q}_a}^2 & \text{COV}(\mathbf{Q}_a, \mathbf{w}_a) \\ \text{COV}(\mathbf{w}_a, \mathbf{Q}_a) & \sigma_{\mathbf{w}_a}^2 \end{bmatrix} \quad (9)$$

**Le modèle de propagation.** L'assimilation de données est fréquemment appliquée à des systèmes dynamiques dont l'évolution en temps est régie par un modèle de propagation. Ce modèle de propagation est une représentation mathématique des lois de la physique discrétisées en espace et en temps. Le modèle prédit l'état du système au temps  $i + 1$  en fonction de l'état au temps  $i$  et d'un jeu de conditions initiales et de paramètres physiques. Ce modèle de propagation est noté  $\mathcal{M}$ .

Soit  $\mathbf{x}$  l'état du système au temps  $i$ ,  $\mathbf{x}$  peut par exemple représenter le débit évalué à différents points d'une rivière par un modèle hydrodynamique. Le modèle de propagation permet d'exprimer l'état du système au temps  $i + 1$  :

$$\mathbf{x}^{i+1} = \mathcal{M}(\mathbf{x}^i) \quad (10)$$

Dans la suite de ce document, nous ferons l'hypothèse que le modèle de propagation est linéaire (on le note  $\mathbf{M}$ ) et qu'il n'introduit pas d'erreur entre les instants  $i$  et  $i + 1$ , c'est à dire que l'on le considère parfait entre ces deux instants.

Dans le contexte de l'assimilation de données, le modèle de propagation permet de propager l'état du système ainsi que la matrice de covariance d'erreur d'ébauche  $\mathbf{B}$ . Ainsi la matrice  $\mathbf{B}$  évaluée à l'instant  $i$  peut être propagée à l'instant  $i + 1$  selon la loi de propagation des variances :

$$\mathbf{B}^{i+1} = \mathbf{M}\mathbf{B}^i\mathbf{M}^T \quad (11)$$

Le tableau 1 rassemble les concepts abordés ainsi que leur notation.

La figure 3 présente schématiquement l'utilisation des éléments que l'on a introduit jusqu'à ici pour produire une analyse.

### 1.2.2 L'étape d'analyse

L'étape d'analyse consiste à identifier une solution optimale pour la description du système en utilisant au mieux le modèle mathématique, les observations, l'ébauche et les erreurs sur ces informations. Cette solution dépend de la méthode choisie. On peut classer les méthodes d'assimilation en divers groupes, notamment en les distinguant selon le traitement de la dimension temporelle.

Le premier groupe de méthodes consiste à réajuster le modèle à chaque fois qu'une observation est disponible. On parle alors de *méthodes d'assimilation séquentielle*. L'autre groupe consiste à modifier la condition initiale au début d'une fenêtre de temps pour ajuster au mieux l'état du système aux observations disponibles sur cette fenêtre. On parle de *méthodes d'assimilation variationnelles* (voir FIG. 4).

symbole	nom
$\mathbf{x}$	vecteur de contrôle
$\mathbf{x}_t$	état réel
$\mathbf{x}_b$	vecteur d'ébauche
$\mathbf{x}_a$	vecteur d'analyse
$\mathbf{y}$	vecteur des observations
$\mathcal{H} / \mathbf{H}$	opérateur d'observation / version linéarisée
$\epsilon_b$	erreur d'ébauche
$\epsilon_y$	erreur d'observation
$\epsilon_a$	erreur d'analyse
$\mathbf{B}$	matrice d'erreur d'ébauche
$\mathbf{R}$	matrice d'erreur des observations
$\mathbf{A}$	matrice d'erreur d'analyse
$\mathcal{M} / \mathbf{M}$	modèle de propagation / version linéarisée

TAB. 1 – notation de base pour l'assimilation de données.

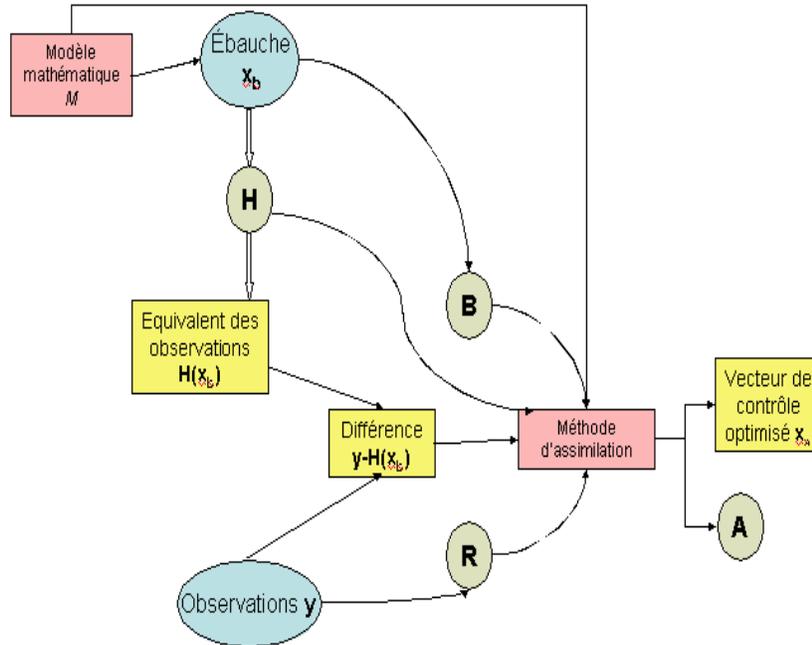


FIG. 3 – Schéma d'utilisation du modèle numérique  $\mathcal{M}$ , de l'ébauche  $\mathbf{x}_b$ , des observations  $\mathbf{y}$  et des matrices d'erreurs  $\mathbf{B}$  et  $\mathbf{R}$  pour la production d'une analyse  $\mathbf{x}_a$  et son erreur  $\mathbf{A}$  en assimilation de données.

**Assimilation séquentielle.** Ce type de méthodes s'appuie sur la théorie statistique du BLUE (de l'anglais "Best Linear Unbiased Estimator"). Cette approche formule un estimateur de  $\mathbf{x}_a$  comme une combinaison linéaire du vecteur d'ébauche  $\mathbf{x}_b$  et du vecteur des observations  $\mathbf{y}$  ([6], [75], [68]). Cet estimateur s'écrit :

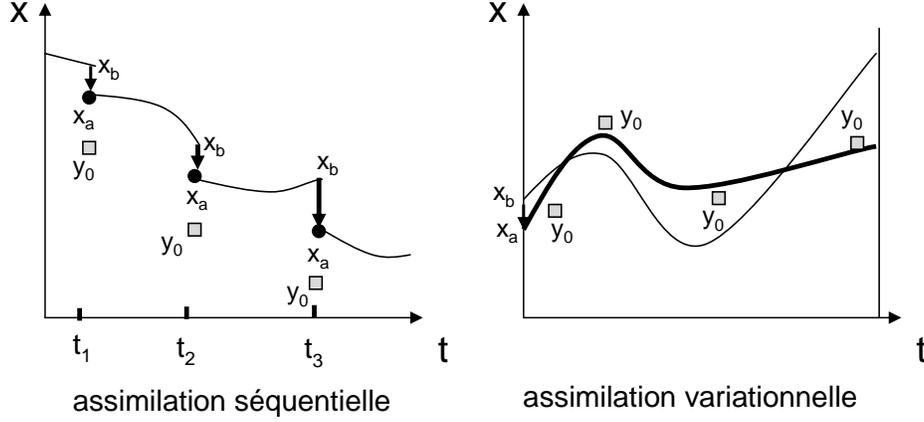


FIG. 4 – Méthodes d’assimilation séquentielles et méthodes d’assimilation variationnelles. Les carrés représentent les observations, les flèches représentent l’incrément d’analyse,  $\mathbf{x}_b$  est l’ébauche,  $\mathbf{x}_a$  l’analyse. La courbe en trait fin représente l’estimation du modèle, la courbe en trait gras représente la trajectoire optimisée après l’assimilation des observations.

$$\mathbf{x}_a = \mathbf{C} \cdot \begin{bmatrix} \mathbf{x}_b \\ \mathbf{y} \end{bmatrix} \quad (12)$$

où  $\mathbf{C}$  est une constante linéaire à déterminer. Dans cette approche, on émet deux hypothèses majeures :

1. l’opérateur d’observation  $\mathcal{H}$ , qui fait le lien entre le vecteur de contrôle et le vecteur d’observations (éq. (2)), est linéaire;
2. l’estimateur  $\mathbf{x}_a$  est non biaisé, c’est à dire, son espérance mathématique est nulle ( $E\{\epsilon_a\} = 0$ ).

On fait de plus l’hypothèse suivante : les erreurs sur les observations sont indépendantes (la covariance entre les erreurs liées à deux observations est nulle). Sous cette hypothèse, la solution du BLUE est obtenue en minimisant la variance de  $\mathbf{x}_a$ . Ce qui conduit à la solution :

$$\mathbf{x}_a = \mathbf{x}_b + \underbrace{\mathbf{K}[\mathbf{y} - \mathbf{H}(\mathbf{x}_b)]}_{\Delta \mathbf{x}_a} \quad (13)$$

$$\text{avec} \quad \mathbf{K} = \mathbf{B}\mathbf{H}^T[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]^{-1} \quad (14)$$

Le BLUE permet de plus de formuler les statistiques sur l’erreur liée à l’analyse. La matrice de covariance d’erreur d’analyse  $\mathbf{A}$  s’écrit [6] :

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B} \quad (15)$$

Les équations (13), (14) et (15) fournissent la solution optimale pour les variables de contrôle d’un système linéaire avec des incertitudes suivant des statistiques gaussiennes. Cette solution est optimale au sens où elle est l’état de minimum de variance et donc l’état le plus certain. Dans l’équation (13) on note que l’analyse  $\mathbf{x}_a$  est la somme de l’ébauche et d’un terme correctif  $\Delta \mathbf{x}_a$  appelé *incrément d’analyse*. Ce terme est proportionnel à l’écart entre l’ébauche et les observations, appelé *vecteur d’innovation*. Le facteur de proportionnalité est la matrice  $\mathbf{K}$  qui dépend seulement des statistiques d’erreur d’ébauche (matrice  $\mathbf{B}$ ) et du vecteur des observations (matrice  $\mathbf{R}$ ). La matrice  $\mathbf{K}$  est appelée *matrice de gain*. Considérons à présent les deux cas limites de l’équation de la matrice de gain :

1. si l’on fait entièrement confiance au modèle, alors il n’y a pas d’erreur sur les variables d’ébauche et donc  $\mathbf{B}$  est la matrice nulle, ce qui implique que  $\mathbf{K}$  est aussi la matrice nulle. Dans ce cas, la solution de l’analyse est identique à l’ébauche :  $\mathbf{x}_a = \mathbf{x}_b$  ;

2. si les observations sont parfaites, alors il n'y a pas d'erreur sur les variables observées, donc  $\mathbf{R}$  est la matrice nulle et  $\mathbf{K} = \mathbf{H}^{-1}$  de sorte que :  $\mathbf{x}_a = \mathbf{H}^{-1}\mathbf{y}$  (à condition que la matrice  $\mathbf{H}$  soit inversible). Dans cette situation, l'analyse est la projection des observations dans l'espace modèle.

**Assimilation variationnelle.** A la différence des méthodes séquentielles, pour les méthodes de type variationnel on considère une fenêtre d'assimilation entre les instants  $t_1$  et  $t_2$  sur laquelle plusieurs observations sont disponibles. Ces méthodes identifient une correction à apporter aux variables de contrôle sur une fenêtre de temps (ou fenêtre d'assimilation). Cette correction est assez généralement portée au début de la fenêtre d'assimilation et constitue ainsi une nouvelle condition initiale. Cette condition initiale est intégrée et donne une trajectoire optimale en termes de comparaison aux observations sur la fenêtre d'assimilation (voir FIG. 4). L'état analysé est identifié en minimisant une fonction coût  $f$  qui prend en compte la somme des écarts entre les observations et la trajectoire du modèle au temps d'observation (FIG. 5).

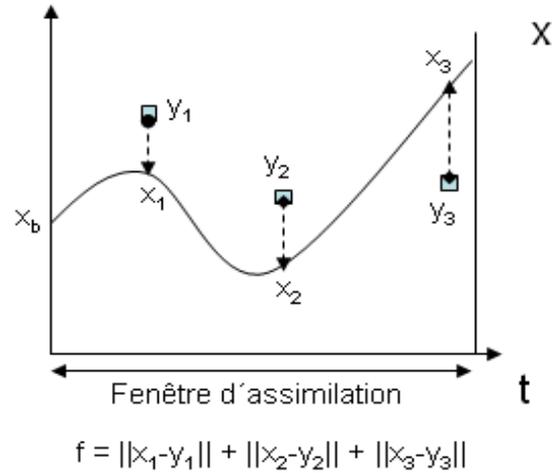


FIG. 5 – Principe de fonctionnement des méthodes variationnelles : minimisation d'une fonction coût sur une fenêtre temporelle.

Si une ébauche est disponible on considère aussi son écart à l'état vrai. Cet information permet de plus de fermer le système à résoudre généralement sous-déterminé du fait du faible nombre d'observations par rapport à la taille du vecteur d'état. Dans ce cas la fonction coût à minimiser s'écrit :

$$\mathcal{J}(\mathbf{x}) = \underbrace{\frac{1}{2} \cdot (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b)}_{\mathcal{J}_b} + \underbrace{\frac{1}{2} \cdot (\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x})}_{\mathcal{J}_o} \quad (16)$$

Il est important de noter que cette fonction est quadratique (on utilise ici l'opérateur d'observation linéarisé  $\mathbf{H}$ ), ce qui facilite la minimisation. La fonction coût  $\mathcal{J}$  est composée de deux termes :  $\mathcal{J}_b$  qui mesure l'écart entre l'état du système et l'ébauche pondéré par la matrice de covariance d'erreur d'ébauche  $\mathbf{B}$  et  $\mathcal{J}_o$  qui mesure l'écart entre l'état du système et les observations pondéré par la matrice de covariance d'erreur d'observations  $\mathbf{R}$ . La minimisation de la fonction  $\mathcal{J}$  consiste à en annuler le gradient par rapport à  $\mathbf{x}$ . L'état analysé est tel que :

$$\nabla \mathcal{J}(\mathbf{x}_a) = 0 \quad (17)$$

Sous l'hypothèse de linéarité des opérateurs et sous l'hypothèse d'erreurs gaussiennes, la minimisation de la fonction coût a pour solution [6] :

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}[\mathbf{y} - \mathbf{H}(\mathbf{x}_b)] \quad (18)$$

avec 
$$\mathbf{K} = \mathbf{B}\mathbf{H}^T[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]^{-1} \quad (19)$$

On observe que ces équations sont identiques aux équations (13) et (14). Sous ces hypothèses, l'approche séquentielle est équivalente à l'approche variationnelle, ce sur une période de temps égale. Le choix d'une méthode ou d'une autre dépendra de facteurs tels que la taille du problème physique.

Les équations de l'analyse (18) et (19) sont valides pour les systèmes linéaires. Néanmoins, les systèmes simulant les processus physiques réels sont majoritairement non-linéaires. La minimisation de la fonction coût ne peut être faite analytiquement. On choisit alors une approche itérative qui requiert l'estimation de la fonction coût et de son gradient à chaque itération par rapport aux variables du vecteur de contrôle. Ces calculs s'avèrent coûteux et complexes. Cette méthode classique de résolution numérique du problème de minimisation de la fonction  $\mathcal{J}$  est appelée *la méthode des modèles linéaire tangent et adjoint*. La description de cette méthode n'est pas l'objet de ce document.

### 1.2.3 L'exemple du pendule

**Description du problème.** Illustrons à présent les différents concepts abordés sur un exemple simple : le pendule [76]. Notre objectif : *prévoir la position du pendule à un temps donné*. Pour cela, nous disposons d'observations de la position du pendule à différentes échéances temporelles et de l'équation physique du mouvement du pendule. L'équation du mouvement du pendule amorti pour l'angle  $\alpha$  en fonction du temps  $t$  est :

$$\ddot{\alpha} + 2\lambda\dot{\alpha} + \omega_n^2 \sin \alpha = 0 \quad (20)$$

où  $\alpha$  est l'angle positif entre la verticale et le pendule (Fig. 6),  $\dot{\alpha} = d\alpha/dt$  est la vitesse angulaire

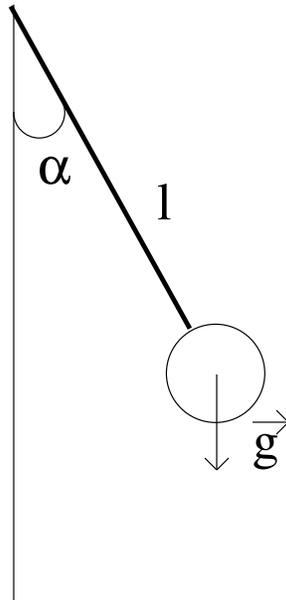


FIG. 6 – Pendule simple de longueur  $l$ . L'accélération de la pesanteur est  $g$  et  $\alpha$  est l'angle positif entre la verticale et le pendule.

du pendule et  $\ddot{\alpha} = d^2\alpha/dt^2$  est l'accélération angulaire du pendule.  $\lambda$  est le coefficient de dissipation et  $\omega_n = \sqrt{g/l}$  est la fréquence propre du pendule ( $l$  : longueur du pendule,  $g$  : accélération de la pesanteur). Les valeurs exactes de ces coefficients sont malconnues, aussi on utilise une estimation de ces paramètres pour la modélisation.

On définit le vecteur de contrôle  $\mathbf{x}$  du système par l'angle  $\alpha$  ( $\mathbf{x} = (\alpha)$ ), c.à.d, que la variable  $\alpha$  décrit totalement la position angulaire du pendule par rapport à sa position d'équilibre ( $\alpha = 0$ ).

On observe la position du pendule (mesure de l'angle  $\alpha$ ) toutes les secondes (les instants  $t_i$ ), en commettant une erreur de mesure aléatoire (Fig. 7).

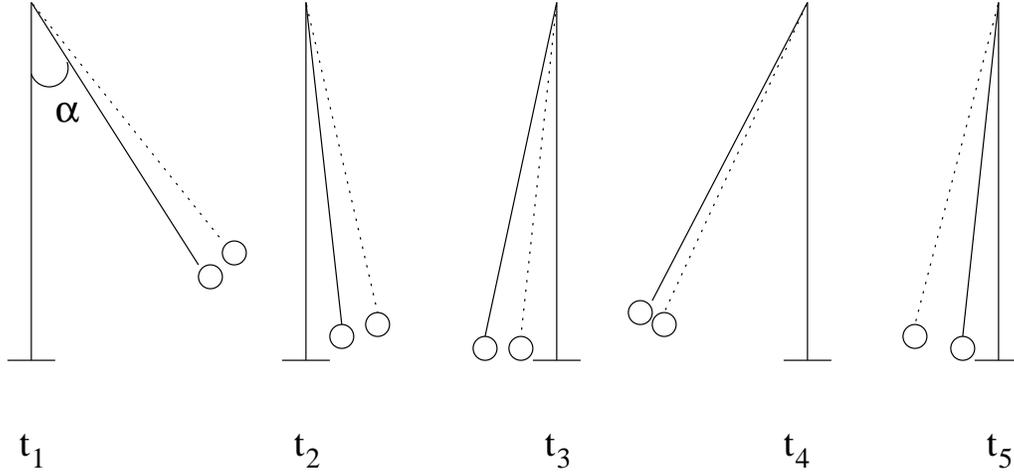


FIG. 7 – Observations : mesures bruitées de l'angle  $\alpha$  entre une position approximative du pendule (trait plein) et la verticale aux temps  $t_i$ . La position réelle du pendule à chaque instant est représentée par des pointillés.

Nos deux sources d'information sont donc les suivantes : des observations imparfaites et peu nombreuses ainsi qu'une équation du mouvement dont les paramètres sont spécifiés (éq. 20) de sorte qu'elle représente de façon approximative la position et la vitesse réelles du pendule. Il est important de noter à ce stade, que si les observations étaient parfaites et suffisamment nombreuses, nous n'aurions pas besoin de l'équation de mouvement du pendule pour connaître sa position à un instant donné. D'autre part, si l'équation physique était exacte ( $\lambda$  et  $\omega_n$  exacts) nous pourrions obtenir une solution juste en résolvant l'équation (20) analytiquement en connaissant les conditions initiales du problème, sans avoir recours aux observations. N'étant en définitive dans aucune de ces configurations, nous devons utiliser l'équation de mouvement et les observations pour connaître la trajectoire du pendule la plus proche possible de la trajectoire réelle. L'équation du mouvement approximée du pendule modélise l'évolution au cours du temps du vecteur de contrôle  $\mathbf{x}$ . Cette équation une fois discrétisée en espace et en temps constitue le *modèle numérique*. Notons que la discrétisation est en elle-même source d'une erreur numérique qui s'ajoute à l'erreur commise dans la définition des paramètres  $\lambda$  et  $\omega_n$ . Le modèle donne une description *a priori* de la position du pendule et de sa vitesse à chaque instant. Cette trajectoire est la trajectoire d'ébauche  $\alpha_b$ .

Imaginons que l'on veuille estimer la position du pendule à chaque instant d'observation  $t = t_i$ . L'intégration du modèle numérique décrit de plus la trajectoire de l'ébauche au fil du temps et donne une estimation  $\alpha_b$  de la position du pendule à  $t = t_i$ . Par exemple  $\alpha_b(t = 1) = -\frac{\pi}{18}$  rad (équivalent à 10 degrés d'arc). Cette trajectoire est une sinusoïde de fréquence  $\omega_n$  et avec un coefficient de dissipation  $\lambda$ , représentée sur la figure 8 par la courbe en trait plein. Les positions observées du pendule sont représentées par des croix sur la figure 8. L'assimilation de données consiste à faire une moyenne pondérée aux instants  $t_i$  entre les observations et la trajectoire

d'ébauche (les états d'ébauche aux instants  $t_i$  sont représentées par des cercles sur la figure 8). Les états analysés aux instants  $t_i$  sont représentées par des points sur la figure 8.

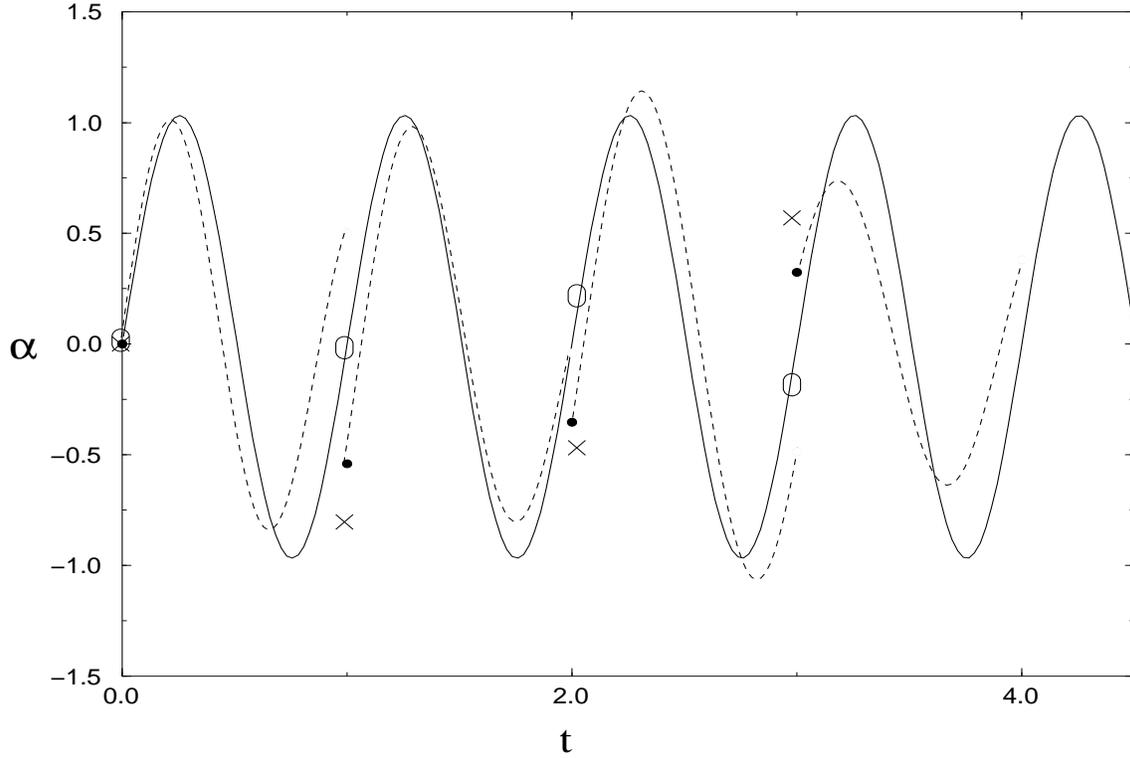


FIG. 8 – Représentation schématique de la trajectoire du pendule  $\alpha(t)$  : la courbe en trait plein représente la trajectoire de l'ébauche, les cercles représentent la valeur de l'angle  $\alpha^b$  sur l'ébauche aux temps d'observation, les croix sont les observations de la mesure de l'angle  $\alpha$  à chaque seconde et les points représentent la valeur analysée de cet angle. La trajectoire analysée pour l'angle  $\alpha$  est la courbe discontinue en pointillés.

**Analyse à l'instant  $t = 1$ .** Supposons que l'ébauche  $\alpha_b(t = 1) = -\frac{\pi}{18}$  est entachée d'une erreur de 10%. Imaginons que la position observée du pendule à  $t = 1$  soit  $\alpha = -\frac{\pi}{6}$  (-30 degrés d'arc de circonférence). Cette observation est entachée d'une erreur de mesure de 5%.

Toute cette information nous permet d'analyser la position du pendule à l'instant  $t = 1$  en utilisant l'assimilation de données. Cette analyse on la note  $\alpha_a$ . Étant donné que la variable de contrôle est la variable observée, l'opérateur d'observation est la matrice identité. On applique les équations d'analyse optimale (13) et (14) pour le cas d'une correction séquentielle de l'ébauche :

$$\alpha_a = \alpha_b + \mathbf{K}[\alpha_{OBS} - \mathbf{H}\alpha_b] \quad (21)$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T[\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}]^{-1} \quad (22)$$

En substituant  $\alpha_b = -\frac{\pi}{18}$  et  $\alpha_{OBS} = -\frac{\pi}{6}$  dans éq. (21) :

$$\alpha_a = -\frac{\pi}{18} + \mathbf{K}\left[-\frac{\pi}{6} + \mathbf{I}\frac{\pi}{18}\right]$$

On calcule alors la matrice de gain  $\mathbf{K}$  avec  $\sigma(\alpha_b) = 0.1 \cdot \alpha_b = 0.018$ , et  $\sigma(\alpha_{OBS}) = 0.05 \cdot \alpha_{OBS} =$

0.026. Alors les matrices  $\mathbf{B}$  et  $\mathbf{R}$  sont dans ce cas des scalaires qui valent :

$$\begin{aligned}\mathbf{B} &= [\sigma(\alpha_b)]^2 = (0.1 \cdot \frac{\pi}{18})^2 \\ \mathbf{R} &= [\sigma(\alpha_{OBS})]^2 = (0.05 \cdot \frac{\pi}{6})^2\end{aligned}$$

En substituant ces valeurs dans l'éq. (22) on obtient :

$$\mathbf{K} = (0.1 \cdot \frac{\pi}{18})^2 \left[ (0.1 \cdot \frac{\pi}{18})^2 + (0.05 \cdot \frac{\pi}{6})^2 \right]^{-1} \approx 0.308$$

La valeur analysée  $\alpha_a$  est finalement :

$$\alpha_a = -\frac{\pi}{18} + 0.308 \cdot \left[ -\frac{\pi}{6} + \mathbf{I} \frac{\pi}{18} \right] \approx -\frac{\pi}{11}$$

**Erreur d'analyse à l'instant  $t = 1$ .** L'analyse optimale est la position angulaire du pendule entre l'ébauche et le modèle (approximativement -17 degrés d'arc) compte tenu des incertitudes sur ces informations (voir Fig. 8). L'incertitude sur l'état analysé selon l'éq. (15) est :

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

En substituant les valeurs de  $\mathbf{K}$ ,  $\mathbf{H}$  et  $\mathbf{B}$  dans cette équation et en faisant la racine carré pour évaluer l'erreur de l'analyse en fonction de l'écart type, on obtient que  $\sigma(\alpha_a) = 0.0145$ . L'erreur de l'état analysé a été réduite par rapport à celles de l'ébauche et de l'observation ( $\sigma(\alpha_a) < \sigma(\alpha_b)$  et  $\sigma(\alpha_{OBS})$ ).

L'analyse à  $t_i$  est utilisée comme ébauche à  $t_{i+1}$ . Cette ébauche est propagée par le modèle jusqu'à  $t_{i+1}$  pour une nouvelle analyse. Ce cyclage d'analyses décrit une trajectoire analysée de l'angle  $\alpha$  du pendule discontinue en temps comme illustré par la courbe en pointillés sur la figure (Fig. 8)

Cet exemple schématique illustre une méthode d'assimilation séquentielle qui, à chaque instant d'observation, effectue une analyse en n'utilisant que les observations à l'instant donné (une seule à chaque instant dans cet exemple).

### 1.3 La philosophie de quelques méthodes d'assimilation de données

Cette section présente les principales méthodes d'assimilation utilisées au CERFACS, ainsi que dans de nombreux centres de recherche ou opérationnels.

#### 1.3.1 Interpolation optimale (OI)

La méthode OI a été l'une des premières méthodes d'assimilation de données modernes implémentées par plusieurs centres avec l'objectif d'améliorer la simulation des champs physiques générés. Aujourd'hui l'OI est encore opérationnelle pour l'analyse des champs d'humidité du sol et de la température des surfaces continentales (primordiales comme condition limite de la basse troposphère des modèles de prédiction météorologique) de Météo-France et du CEPMMT (Centre Européen pour la Prévision Météorologique à Moyen Terme).

L'hypothèse fondamentale derrière la méthode d'OI ([47], [63]) est que pour chaque variable de contrôle en chaque point de grille, on ne prend en compte à l'instant d'analyse qu'un nombre réduit d'observations. La sélection des observations retenues autour de chaque point de grille suit un critère plus ou moins ad-hoc (voir FIG. 9).

**Avantages :** La construction de la matrice de gain  $\mathbf{K}$  se fait à un coût numérique réduit et son implémentation est relativement aisée à faire. Pour le calcul de l'analyse on utilise les équations (18) et (19), on identifie alors la solution optimale à l'instant d'analyse.

**Inconvénients :** La solution demeure sous-optimale relativement à l'ensemble des observations, et les analyses en chaque point de grille sont généralement bruitées par le choix d'un groupe réduit d'observations contribuant à l'analyse.

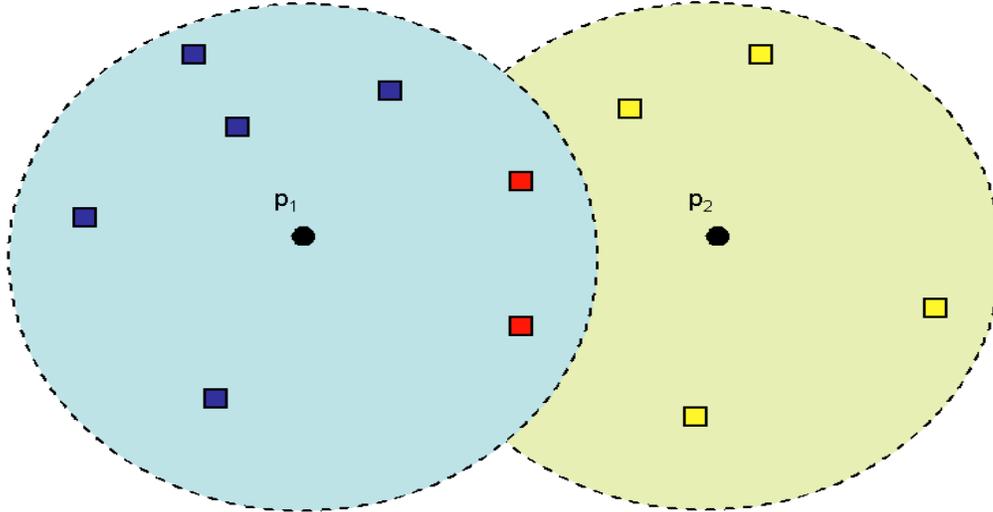


FIG. 9 – Principe de l'assimilation avec la méthode de l'OI. Les carrées représentent les observations : en bleu, celles qui contribuent à l'analyse au point  $p_1$ , en jaune celles qui contribuent à l'analyse au point  $p_2$ . Seules les observations en rouge contribuent à l'analyse aux deux points de grille.

### 1.3.2 Filtre de Kalman

Une des méthodes les plus utilisées est le Filtre de Kalman, aussi on propose une description des étapes d'un filtre de Kalman. À l'égal de l'OI, le filtre de Kalman effectue des corrections sur les variables de contrôle séquentiellement selon la disponibilité des observations (section 1.2.2).

#### 1. Initialisation du filtre de Kalman :

À l'instant  $t_i$  on dispose d'une ébauche  $\mathbf{x}_b^i$  dont les covariances sont décrites dans la matrice  $\mathbf{B}$  notée ici  $\mathbf{P}_b^i$  par convention d'écriture. Avec cette seule information, la meilleure estimation de l'état du système au temps  $i$  est l'ébauche et la covariance de l'estimation est égale à la covariance de l'ébauche :

$$\mathbf{x}_a^i = \mathbf{x}_b^0 \quad (23)$$

$$\mathbf{P}_a^i = \mathbf{P}_b^0 \quad (24)$$

#### 2. Étape de propagation :

Dans cette étape, l'état du système évolue temporellement grâce au modèle physique  $\mathcal{M}$  jusqu'au prochain temps d'observation. En première approximation on suppose  $\mathcal{M}$  linéaire (noté  $\mathbf{M}$ ) :

$$\mathbf{x}_b^{i+1} = \mathbf{M}\mathbf{x}_a^i \quad (25)$$

La matrice de covariance du système est elle aussi propagée de l'instant  $i$  à l'instant  $i + 1$  selon la loi de propagation des variances (éq. (11)) :

$$\mathbf{P}_b^{i+1} = \mathbf{M}\mathbf{P}_a^i\mathbf{M}^T \quad (26)$$

### 3. Étape d'analyse :

Dans cette étape on utilise l'observation au temps  $i+1$ . L'information de la nouvelle observation est introduite pour être combinée avec l'estimation du modèle dynamique. La recherche de l'analyse  $\mathbf{x}_a^{i+1}$  est décrite par les équations du BLUE (section 1.2.2).

$$\begin{aligned} \mathbf{x}_a^{i+1} &= \mathbf{x}_b^{i+1} + \mathbf{K}^{i+1}(\mathbf{y}^{i+1} - \mathbf{H}^{i+1} \cdot \mathbf{x}_b^{i+1}) & (27) \\ \text{avec } \mathbf{K}^{i+1} &= \mathbf{P}_b^{i+1} \mathbf{H}^{i+1} (\mathbf{H}^{i+1} \mathbf{P}_b^{i+1} (\mathbf{H}^{i+1})^T + \mathbf{R}^{i+1})^{-1} & (28) \end{aligned}$$

Finalement, pour que le filtre soit complet, une estimation de la matrice de covariance d'erreur d'analyse  $\mathbf{P}_a^{i+1}$  doit être fournie :

$$\mathbf{P}_a^{i+1} = E [(\mathbf{x}_a^{i+1} - \mathbf{x}_t^{i+1})(\mathbf{x}_a^{i+1} - \mathbf{x}_t^{i+1})^T] = (\mathbf{I} - \mathbf{K}^{i+1} \mathbf{H}^{i+1}) \mathbf{P}_b^{i+1} \quad (29)$$

Ce jeu d'équations ((23)-(29)) nous permet de corriger les variables de contrôle et leurs incertitudes à chaque fois que l'on rencontre une observation, puis de propager l'état du système et son incertitude jusqu'au prochain temps d'observation.

**Avantages :** La méthode OI est suboptimale du fait de l'utilisation d'un groupe réduit d'observations, le filtre de Kalman est lui optimal globalement pour des systèmes linéaires. Il permet aussi de combiner plusieurs types d'observations plus facilement que d'autre méthodes d'assimilation de données.

**Inconvénients :** Le filtre de Kalman est seulement applicable à des systèmes linéaires. De plus cette méthode engendre des calculs algébriques dans l'espace de l'état du modèle sur des matrices de grandes taille. Le filtre de Kalman n'est donc pas applicable pour des systèmes de grande taille tels que l'océan ou l'atmosphère. Pour de tels systèmes, on utilise des simplifications du filtre de Kalman qui consistent à travailler dans un espace de taille réduite. Ces approches ne seront pas présentées ici.

#### 1.3.3 3D-VAR

Cette méthode, largement utilisée dans les services opérationnels de météorologie et d'océanographie, s'appuie sur l'approche variationnelle. A la différence de la méthode d'OI, elle prend en compte toutes les observations disponibles sur une période temporelle et ce sur tout le domaine spatial d'étude, ce qui donne aux analyses une meilleure cohérence spatio-temporelle que l'OI. Les observations sont ramenées à un seul instant, comparées aux variables de contrôle estimées par le modèle au même instant (fréquemment au milieu de la fenêtre d'assimilation). Les variables pronostiques sont corrigées par un incrément (incrément d'analyse) et propagées par le modèle jusqu'à la fin de cette fenêtre.

**Avantages :** La méthode 3D-VAR est bien adaptée à des systèmes où le vecteur d'état est un vecteur de grande taille, tels que par exemple l'atmosphère. Elle est moins coûteuse en termes de temps de calcul que les méthodes 3D-FGAT et 4D-VAR décrites par la suite.

**Inconvénients :** Cette méthode ne prend pas en compte la dimension temporelle des observations ce qui peut biaiser l'analyse.

#### 1.3.4 3D-VAR FGAT

Le 3D-FGAT (First Guess at Appropriate Time) [50] est une version sophistiquée du 3D-VAR qui prend en compte la dynamique du système pour calculer l'écart entre l'état du modèle et les observations. Les observations sont comparées à leur équivalent modèle à l'instant exact d'observation. Le 3D-FGAT est une approche incrémentale qui vise non plus à donner une nouvelle

estimation du vecteur de contrôle mais l'estimation d'une correction à apporter à ce vecteur. Cette correction est l'incrément d'analyse. Dans l'approche 3D-FGAT on fait l'hypothèse que l'incrément est constant sur la fenêtre d'assimilation et qu'il est donc valide à tout instant de cette fenêtre.

**Avantages :** L'hypothèse émise sur l'incrément d'analyse diminue notablement le coût de calcul par rapport au 4D-VAR. De plus, elle est plus optimale que le 3D-VAR.

**Inconvénients :** Le 3D-FGAT est plus coûteux que le 3D-VAR. Contrairement au 4D-VAR, cette méthode ne prend pas en compte la dynamique complète du système.

### 1.3.5 4D-VAR

Nous appréhendons ici le 4D-VAR ([47], [30]) dans sa version incrémentale afin de poursuivre la discussion entreprise sur le 3D-FGAT. Cette méthode variationnelle prend en compte la dynamique du système dans la propagation de l'incrément d'analyse. Cet élément constitue un surcoût majeur. En effet dans le processus de minimisation, il faut évaluer la fonction coût et son gradient à chaque itération. Le calcul de la fonction coût implique de propager explicitement l'incrément par le modèle linéaire tangent. Le calcul du gradient de la fonction coût implique la propagation inverse de l'incrément par l'adjoint du modèle linéaire tangent.

En plus du surcoût de calcul, le 4D-VAR implique l'utilisation du code adjoint qui n'est pas toujours accessible. L'analyse du 4D-VAR (incrémental) identifie une correction du vecteur de contrôle généralement appliquée au début de la fenêtre d'assimilation. A partir de cette nouvelle condition initiale l'intégration du modèle de propagation fournit une nouvelle trajectoire optimale par rapport à la trajectoire de l'ébauche et aux observations. La figure 10 présente les principales différences entre la méthode 3D-VAR et le 4D-VAR.

**Avantages :** L'analyse est optimale sur la fenêtre d'assimilation et comparable à l'analyse du filtre de Kalman pour des systèmes linéaires.

**Inconvénients :** Cette méthode implique l'utilisation des codes adjoint et linéaire tangent. Le coût de calcul est élevé et comparable à celui d'un filtre de Kalman complet.

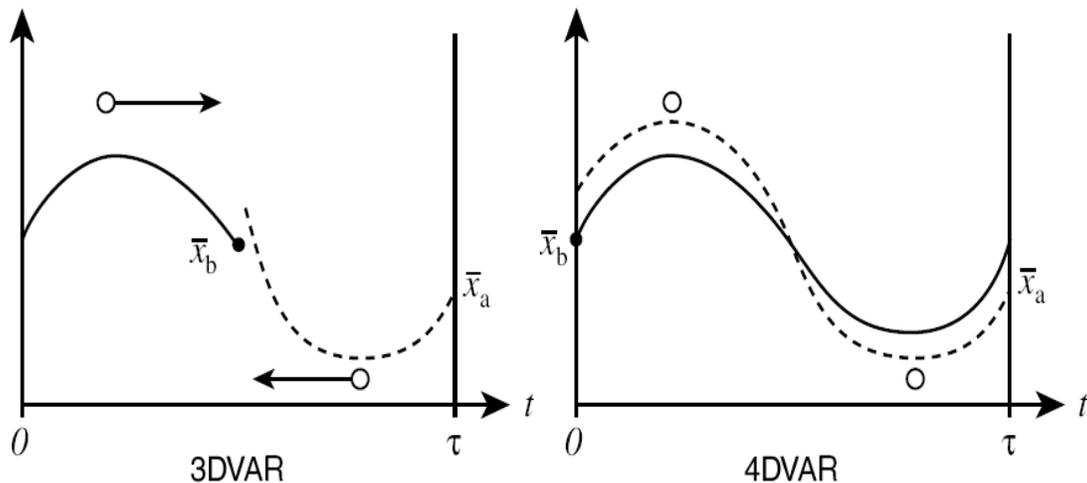


FIG. 10 – Principe d'assimilation des méthodes 3DVAR et 4DVAR. Le 3DVAR ramène les deux observations au temps d'analyse sans considérer leur dimension temporelle. Le 4DVAR prend en compte la dynamique du système et ajuste globalement la trajectoire du système aux observations disponibles.

## 2 Les outils pour l'assimilation de données

Cette section est consacrée à la description des outils pour l'assimilation utilisés régulièrement au CERFACS. Ces outils visent à une meilleure gestion et optimisation des méthodes d'assimilation de données.

### 2.1 PALM

Le logiciel PALM (Projet d'Assimilation par Logiciel Multi-méthodes) [87] est un outil de couplage dynamique de codes parallèles développé au CERFACS. Ce logiciel offre une interface graphique souple et facile à utiliser permettant le couplage de différentes unités numériques. Cette modularité facilite le développement des algorithmes d'assimilation et permet de plus de tester diverses méthodes en ne modifiant qu'un nombre limité d'unités, ce dans un environnement parallèle. PALM permet enfin l'accès à des bibliothèques d'algèbre indispensables pour l'assimilation de données.

L'environnement PALM est particulièrement adapté à l'assimilation de données. Cet outil logiciel permet de tester plusieurs méthodes d'assimilation en parallèle sans besoin de savoir programmer du calcul parallèle. Il permet d'échanger des unités correspondant à des entités des méthodes d'assimilation tel que l'opérateur d'observation, le minimiseur, les matrices de covariances d'erreurs, le tout dans un environnement graphique convivial et facile à utiliser. Il a été utilisé pour de nombreuses applications décrites sur le site web du logiciel palm :

[http://www.cerfacs.fr/globc/PALM\\_WEB/](http://www.cerfacs.fr/globc/PALM_WEB/).

### 2.2 Les algorithmes de minimisation

Aujourd'hui la plupart des services opérationnels mettent en œuvre des méthodes d'assimilation variationnelles pour l'analyse de champs physiques des variables pronostiques. En effet la minimisation de la fonction coût est généralement moins coûteuse que le calcul matriciel, notamment l'inversion de la matrice de gain. Comme l'on a déjà vu dans la section 1.2.2, les méthodes variationnelles consistent à minimiser la fonctionnelle  $\mathcal{J}$  mesurant l'écart entre le modèle et l'analyse et entre les observations et l'analyse, avec la contrainte des incertitudes sur les observations et l'ébauche (16). La minimisation de cette fonctionnelle représente un problème complexe et coûteux. Dans cette section on présente les méthodes de minimisation les plus utilisées au CERFACS.

#### 2.2.1 Méthode du gradient

La méthode du gradient recherche la direction de plus profonde descente d'une fonction  $f$ . Comme toutes les méthodes numériques de minimisation, la méthode du gradient est une méthode itérative qui applique la loi de récurrence suivante :

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \gamma_k \mathbf{s}_k \quad (30)$$

où  $\mathbf{s}_k$  est la direction de descente au temps  $k$  et  $\gamma_k$  un coefficient qui module la taille du pas de descente à l'instant  $k$ . L'algorithme est initialisé par une première approximation à la solution  $\mathbf{x} = \mathbf{x}_0$ . Ensuite, on cherche à identifier la direction de descente  $\mathbf{s}$  de pente maximum en annulant le gradient de la fonction  $\nabla f = 0$  :

$$\mathbf{x}_1 = \mathbf{x}_0 - \gamma_0 \mathbf{s}_0 \quad (31)$$

$$\mathbf{x}_2 = \mathbf{x}_1 - \gamma_1 \mathbf{s}_1$$

$$\mathbf{x}_3 = \mathbf{x}_2 - \gamma_2 \mathbf{s}_2$$

...

...

$$(32)$$

$$(33)$$

Enfin, on applique un critère qui consiste à comparer la norme du gradient de la fonction à minimiser avec une valeur très petite  $\epsilon$ .

**Avantages :** Le stockage en mémoire ainsi que le coût numérique de chaque itération sont inférieurs à celui des autres méthodes de minimisation.

**Inconvénients :** Si le pas de descente est très petit, alors la minimisation peut nécessiter de nombreuses itérations, sans pour autant que l'on ait la certitude que la minimisation converge (annexe A1).

### 2.2.2 Gradient conjugué

La méthode du gradient conjugué est une amélioration numérique de celle du gradient simple. Elle consiste à prendre de meilleures directions de descente qu'avec la méthode du gradient. Ceci se fait en construisant des directions conjuguées au moyen d'un processus orthogonal de Gram-Schmidt (d'où le nom de gradient conjugué). La direction de recherche du minimum est donc plus optimale et directe que dans la méthode du gradient. L'annexe A2 fait une présentation plus détaillée de l'algorithme du gradient conjugué.

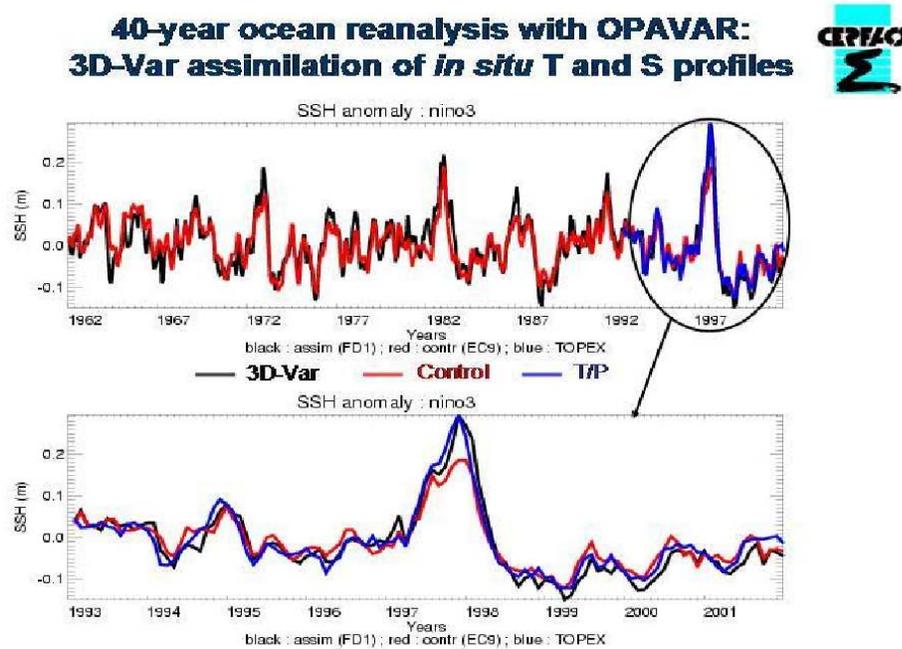


FIG. 11 – En haut, anomalie de SSH moyennée sur la boîte nino3 pour l'expérience de contrôle (sans assimilation) en rouge et avec assimilation (méthode 3DVAR-FGAT) en noir. En bas, des données indépendantes de la mission TOPEX/POSEIDON (non-assimilées) sont ajoutées en bleu.

Le projet NEMOVAR rassemble diverses améliorations tant technologiques que scientifiques :

1. Le nouveau système NEMOVAR a été conçu afin de gérer au mieux l'utilisation de la mémoire dynamique ainsi que le parallélisme de tâches. Ce code est donc parfaitement adapté à l'exploitation des calculateurs puissants dont s'équipent aujourd'hui les gros centres de calcul.
2. NEMOVAR permettra l'accès à une configuration à plus haute résolution, notamment pour des analyses à l'échelle régionale.

3. Un apport conséquent a été réalisé sur le plan de l'algorithme de minimisation du gradient conjugué avec l'utilisation de préconditionneurs, pour assurer une convergence suffisamment rapide du processus de minimisation. Dans OPAVAR deux routines basées sur les méthodes du gradient conjugué (annexe A.2) et Newton (paragraphe ??) sont utilisées pour obtenir la solution de problèmes quadratiques à la suite d'une formulation incrémentale de 3D-VAR et 4D-VAR. L'objectif dans NEMOVAR sera de rassembler les avantages de ces deux routines pour créer une seule et plus puissante méthode de minimisation.
4. La réalisation d'un ensemble d'analyses, ainsi que la mise en place d'un opérateur de diffusion généralisé permettront de proposer une meilleure représentation de la matrice de corrélation des erreurs d'ébauche (matrice **B**).

### 2.2.3 Perspectives

Dans un premier temps, on s'attachera à tester la performance des algorithmes de minimisation et des différents préconditionneurs, ainsi que les développements effectués sur la matrice de covariance d'erreur d'ébauche sur un nombre réduit de cycles d'analyse (expérience courte afin de réduire les coûts de calcul). Dans un deuxième temps, la réalisation d'expériences longues permettra la réalisation des prévisions saisonnières et d'effectuer des réanalyses de l'océan.

## 2.3 L'assimilation de données pour le suivi d'espèces chimiques atmosphériques au CERFACS

### 2.3.1 Introduction

Au cours des dernières décennies, l'effet de serre, le changement climatique ainsi que la qualité de l'air sont devenus des préoccupations importantes qui ont donné une plus grande importance à l'étude de l'évolution et de la propagation des espèces chimiques au sein de l'atmosphère qui contribuent à ces effets. Au CERFACS, l'intérêt a été centré spécifiquement sur la surveillance de l'ozone atmosphérique. L'ozone est un composant vital de la basse stratosphère à cause de son filtrage des radiations UV nocives pour la vie sur les surfaces continentales. De plus, il joue un rôle très important dans la troposphère sur l'effet de serre et sur la photochimie atmosphérique. Cependant, les observations classiques (instruments à la surface, capteurs à bords des avions, etc.) n'offrent pas une bonne couverture spatial-temporelle pour sa surveillance. Le lancement de satellites avec des instruments capables de fournir un signal physique proportionnel à la quantité de gaz trace résidant dans l'atmosphère, permet l'utilisation de schémas d'assimilation modernes pour une meilleure représentativité de l'évolution des espèces chimiques.

### 2.3.2 Le schéma d'assimilation

Le CERFACS développe un système d'assimilation de données chimiques atmosphériques flexible et performant. Pour cela, il est construit autour du modèle de chimie-transport 3D MOCAGE (Modèle Chimie Atmosphérique à Grande Échelle) du CNRM (Centre National de Recherches Météorologiques). Pour assurer la flexibilité de l'assimilation, ce modèle propose des configurations variées qui permettent d'effectuer aussi bien des analyses rapides que des analyses précises. Le domaine est le globe avec une discrétisation spatiale base résolution ( $2^\circ \times 2^\circ$ ) ou haute résolution ( $0.5^\circ \times 0.5^\circ$ ). Différents schémas chimiques sont disponibles avec des versions linéaires (peu coûteuses en temps de calculs) pour quelques espèces (ozone, monoxyde de carbone,...), des versions plus dédiées à la stratosphère et d'autres plus globales (troposphère et stratosphère). La discrétisation verticale débute au niveau du sol et s'étend au choix jusqu'à environ 35 ou 60 km. Les modèles de chimie-transport utilisent des forçages dynamiques externes. Ceux de MOCAGE proviennent soit des analyses de Météo-France, soit des analyses du CEPMMT (Centre Européen pour les Prévisions Météorologiques à Moyen Terme).

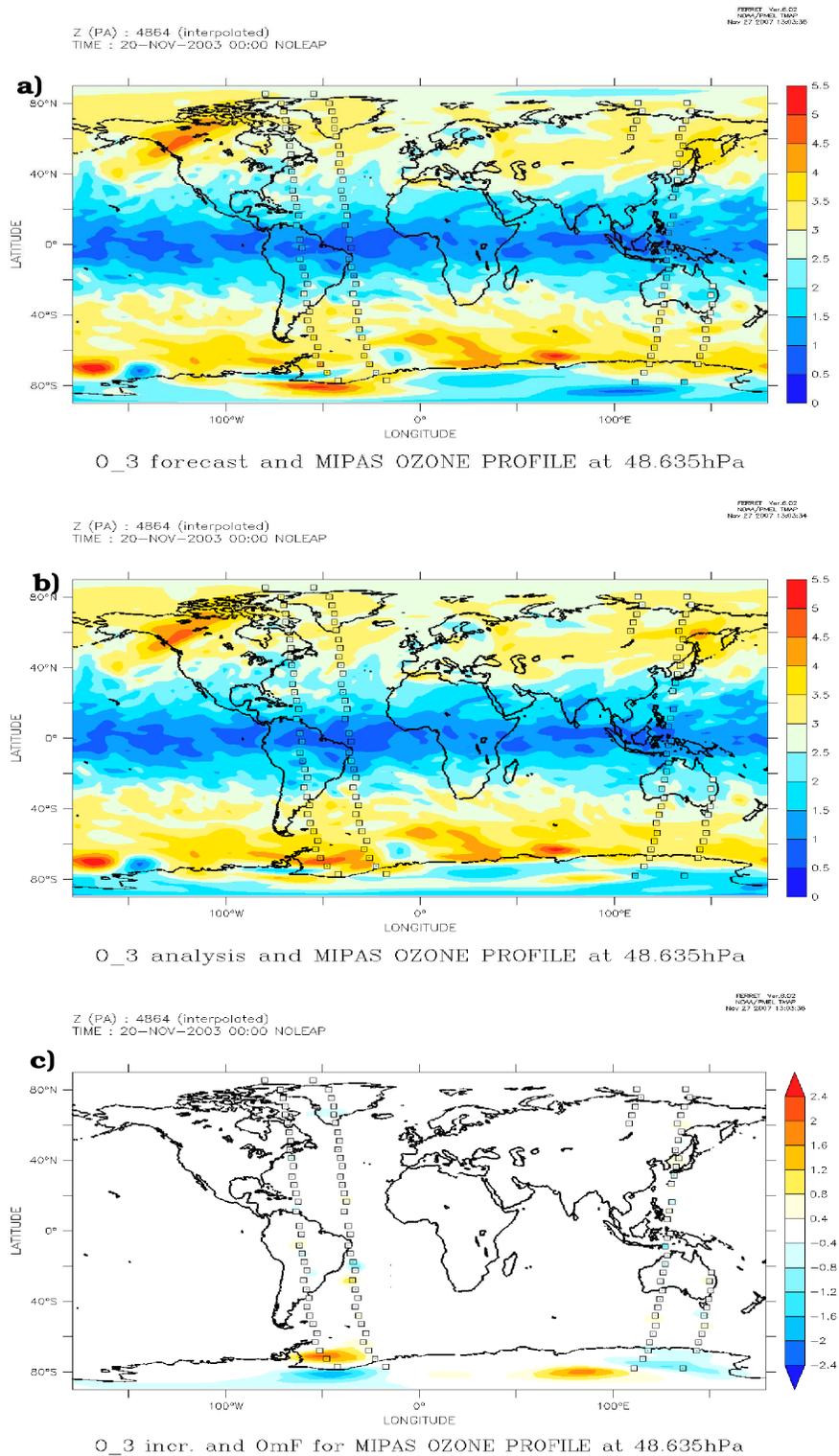


FIG. 12 – Distribution de l’ozone atmosphérique en ppm le 20 Novembre 2003. a) Simulation de MOCAGE, b) Analyses après l’assimilation des observations du capteur MIPAS, c) différence entre la simulation de MOCAGE et les analyses. Les carrés représentent les points géographiques des observations.

Afin de piloter l'ensemble de ces configurations qu'offre MOCAGE, le système d'assimilation nommé MOCAGE-PALM a été réalisé sous l'environnement PALM (section 2.1). La méthode d'assimilation utilisée est le 3D-VAR FGAT (voir section 1.3.4). La formulation incrémentale de cette méthode facilitera à terme la transition du système vers un 4D-VAR incrémental. Pour cela, il sera d'abord nécessaire d'améliorer les ressources de mémoire dans la minimisation de la fonctionnelle, notamment au moyen des algorithmes de minimisation et d'utilisation de préconditionneurs.

Enfin, la dernière possibilité de flexibilité provient du choix des observations que l'on souhaite analyser. En effet, la variété des données chimiques est importante avec des mesures *in situ* provenant de sondages ou d'instruments embarqués à bord d'avions et des mesures spatiales provenant de satellites géostationnaires ou défilants, fournissant des données ponctuelles ou intégrales (intégrées sur une couche partielle ou complète de l'atmosphère). MOCAGE-PALM a été conçu afin de pouvoir assimiler l'ensemble de ces observations.

Ce système d'assimilation a été utilisé lors du projet européen ASSET afin d'analyser les profils d'ozone de l'instrument MIPAS (embarqué à bord du satellite ENVISAT) sur une période de 5 mois. Les analyses d'ozone qui en étaient issues, étaient de qualité comparable aux analyses des systèmes d'assimilation des autres participants. À titre d'illustration, la figure 12 montre la distribution de l'ozone dans la stratosphère à 48 hPa, le 20 de novembre de 2003. La figure 12a correspond à une simulation de MOCAGE de référence sans assimilation. Les petits carrés dans les trois figures représentent les observations effectuées par le capteur MIPAS le 20 Novembre le long de sa trajectoire autour de la Terre. Ces observations sont intégrées dans le schéma d'assimilation, et les analyses qui en résultent sont montrées à la figure 12b. Cette figure montre que les changements sont situés seulement autour des points de mesures. Enfin, la figure 12c montre la différence entre analyses et observations. Pour la majorité des latitudes où les observations étaient disponibles, la concentration d'ozone a très peu changé, sauf pour le pôle sud où de fortes disparités sont observées. Or, c'est justement dans ces latitudes que la couche d'ozone atmosphérique est la plus fine et la plus sensible aux activités humaines. Cela justifie l'approche de l'assimilation de données qui est capable de corriger les estimations de MOCAGE.

### 2.3.3 Perspectives

Actuellement, le système d'assimilation n'a été utilisé que pour effectuer des analyses d'épisodes passés. L'objectif à court terme consiste à adapter le système d'assimilation afin d'analyser des épisodes présents, afin de réaliser en quasi temps réel des prévisions des concentrations des gaz trace de l'atmosphère. Ces prévisions pourront servir par exemple à anticiper la formation du trou d'ozone ou pour l'analyse de la qualité de l'air. Afin d'améliorer ces futures analyses d'espèces chimiques, il sera souhaitable d'utiliser le plus grand nombre d'observations possibles, d'autant plus qu'elles présentent toutes une spécificité différente (résolution verticale et temporelle). La concordance entre les différentes mesures devra être vérifiée.

## 2.4 Assimilation de données pour la neutronique au CERFACS

### 2.4.1 Introduction

Des domaines peu exploités jusqu'à aujourd'hui, tels que la neutronique, traitée dans ce paragraphe, bénéficient également de l'assimilation de données. L'assimilation de mesure de diagnostique du cœur a déjà démontré une amélioration par rapport à l'estimation brute par le modèle COCCINELLE d'EDF R&D de la nappe de flux neutronique 3D d'un réacteur nucléaire. Précédemment, le code d'interpolation CAMARET exploitait les mesures à l'intérieur du cœur du réacteur et les interpolait pour créer une carte optimisée d'activités neutroniques, en prenant en compte les écarts par rapport à la simulation de COCCINELLE. Cependant, toute l'information issue de la mesure ne pouvait être traitée et aucune optimisation des paramètres du code COCCINELLE n'était envisageable. Dans ce contexte, le CERFACS a développé deux maquettes d'assimilation qui complètent CAMARET. L'objectif final est la meilleure connaissance de l'état

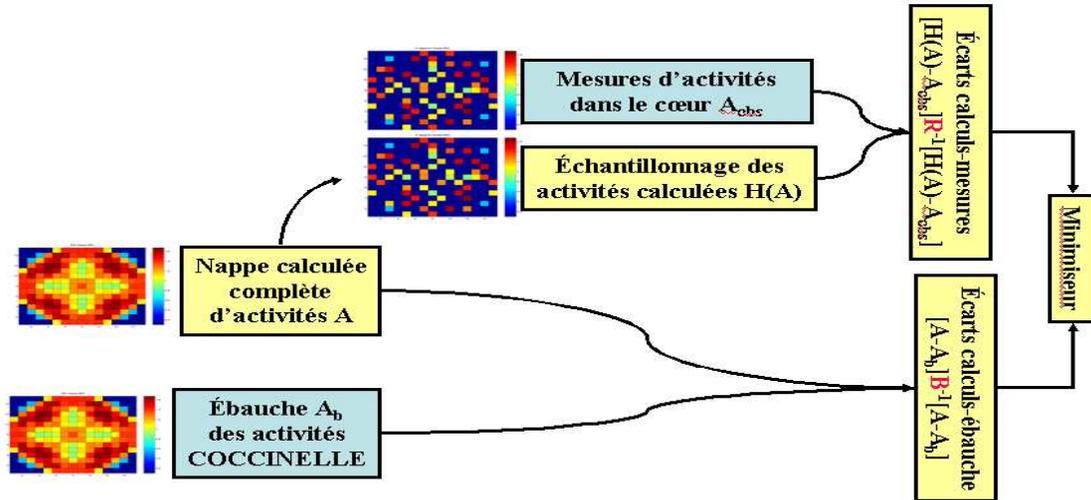


FIG. 13 – Schéma de l'assimilation des observations neutroniques.

neutronique du cœur et une optimisation des modèles de simulation.

La figure 13 montre de manière très simplifiée un schéma du procès d'assimilation pour la neutronique au CERFACS. On observe clairement les deux termes de la fonction coût (eq. (16)) ; d'un côté le système calcul l'écart entre un état de l'activité neutronique du système ( $\mathbf{A}$ ) et l'ébauche par le modèle COCCINELLE ( $\mathbf{A}_b$ ) et de l'autre l'écart entre les observations ( $\mathbf{A}_{obs}$ ) et les estimations du modèle dans l'espace des observations ( $\mathcal{H}(A_{obs})$ ). Ces deux termes sont minimisés, la solution est une valeur optimisée des activités neutroniques.

Dans le cadre des activités menées au sein du CERFACS, deux maquettes d'assimilation pour la neutronique ont été développées : MANARA et KAFEINE.

#### 2.4.2 La maquette MANARA

L'objectif de la maquette MANARA (Modular Application for Neutronic Activity Reconstruction and Assimilation) [72] est l'amélioration de l'estimation de la carte de puissance 3D du réacteur nucléaire. Le modèle COCCINELLE utilisé par le CERFACS est un modèle qui enchaîne différentes étapes de la physique des cœurs : l'étape neutronique, l'étape thermo-hydraulique, la thermique et les contre-réactions qui servent à itérer jusqu'à convergence du calcul. Les prédictions de COCCINELLE ne sont pas parfaites à cause d'une incertitude sur les paramètres d'entrée ainsi que sur la physique nucléaire modélisée dans les équations du modèle. Ces prévisions peuvent être améliorées en utilisant des systèmes d'observation insérés à l'intérieur du réacteur nucléaire. Ces systèmes sont très précieux, car ils permettent de faire un diagnostic du fonctionnement du réacteur. Plusieurs systèmes d'observation de principes physiques différents sont implantés à l'intérieur et à l'extérieur du cœur du réacteur. Ils mesurent par exemple l'activité neutronique ou la température du fluide caloporteur échauffé par les réactions nucléaires.

La méthode d'assimilation de la maquette MANARA est le 3D-VAR. Étant donné la taille relativement faible de ce système, la minimisation de la fonction coût est obtenue de manière directe. Dans cette maquette deux points principaux ont été abordés :

- un travail important a été mené sur la modélisation des valeurs de la matrice de covariance d'ébauche  $\mathbf{B}$ .
- le deuxième point abordé est l'assimilation d'observations provenant de plusieurs instruments. Jusqu'à présent seulement les mesures issues d'un seul type d'instrument avaient été assimilées. Néanmoins, selon la théorie d'estimation optimale la précision totale de l'analyse est plus grande quand le nombre d'observations augmente, indépendamment de leur

précision, selon :

$$\frac{1}{\sigma^2} = \sum_{i=1}^n \frac{1}{\sigma_i^2} \quad (34)$$

avec  $n$  le nombre d'observations. Les travaux sur la maquette MANARA ont montré que l'utilisation observations peu précises est négligeable si d'autres observations plus précises sont disponibles. En revanche, avec peu d'instruments disponibles, l'assimilation d'observations peu précises peut contribuer fortement à la qualité des analyses.

Dans les perspectives de cette étude, il est envisagé de raffiner la modélisation des différents termes qui contribuent à la correction de l'ébauche (opérateur d'observation, matrices d'incertitudes, etc.). La validation croisée pourra être également développée avec l'optimisation de l'algorithme d'assimilation. Enfin, le processus d'analyse permettra de mettre en lumière les points où le modèle COCCINELLE peut être amélioré.

### 2.4.3 La maquette KAFEINE

La maquette d'assimilation KAFEINE a pour objectif d'ajuster de manière simultanée plusieurs paramètres neutroniques qui s'ajustent mieux aux observations du cœur du réacteur. Le problème d'estimation de paramètres est aussi un problème très commun en assimilation de données et c'est une autre manière d'améliorer les simulations d'un modèle, notamment quand les paramètres d'entrée ont une forte incertitude (voir FIG. 2). La méthode retenue dans cette maquette est le Filtre de Kalman Étendue (EKF, [33]). Cette méthode est spécialement adaptée pour des systèmes dont le nombre de paramètres de contrôle est réduit, le modèle mathématique est supposé linéaire et les statistiques d'erreur sont supposées gaussiennes. Ces conditions sont en grande partie réalisées dans ce problème d'assimilation pour la neutronique. Les résultats de KAFEINE ont démontré les avantages de cette méthodologie pour améliorer la procédure classique de calage de paramètres. Cette maquette exploite l'exécution optimale et en parallèle de tâches numériques dans l'environnement PALM.

## 2.5 Une approche pour l'assimilation appliquée à l'hydrologie et à l'hydraulique fluviale

### 2.5.1 Introduction

L'assimilation de données dans le domaine de l'hydrologie est relativement nouveau par rapport à son usage dans les domaines de la météorologie ou de l'océanographie. Plusieurs études ont déjà montré que l'utilisation des systèmes variationnels, tels que le 3D-VAR, ou des systèmes basés sur les filtres de Kalman sont capables d'améliorer le suivi de l'état de la surface, notamment par le biais de son contenu en eau. L'humidité de la surface joue un rôle clé dans le système surface-atmosphère, car elle influence fortement les échanges de flux de chaleur entre la surface et la moyenne troposphère (responsables de phénomènes climatiques à échelle locale comme la formation des nuages) et contrôle la répartition des précipitations entre infiltration et ruissellement, ce qui est vital pour l'équilibre du débit des rivières. Le suivi spatio-temporel de l'humidité du sol est, en conséquence, de première importance pour les modèles de prévision numérique, d'hydraulique fluviale ou les modèles d'agriculture entre autres.

D'autre part, les phénomènes météorologiques extrêmes plus fréquents en raison du changement climatique ont amené à une préoccupation croissante pour la prévision des crues extrêmes. Par exemple, les inondations catastrophiques du 8 et 9 Septembre 2002 dans les départements du Gard, du Vaucluse et de l'Hérault ont causé d'énormes dégâts matériels et pertes des vies humaines. Pour faire face aux risques associés aux inondations, aujourd'hui des services d'annonce de crue fonctionnent de façon opérationnelle et sont repartis en France. Ces services ainsi que d'autres organismes ont développé des modèles hydrologiques et hydrauliques adaptés aux caractéristiques

physiques des différents bassins versants. Cependant, la complexité des processus physiques impliqués lors d'une crue font que le fonctionnement de ces modèles n'est pas encore totalement satisfaisant. L'utilisation de l'assimilation de données dans ce contexte est encore marginale et ses potentialités ne sont pas encore exploitées.

### 2.5.2 Initialisation de l'humidité du sol pour les modèles de prévision numérique

Actuellement à Météo France, l'humidité du sol dans les modèles de prévision numérique opérationnels ARPEGE (modèle global) et ALADIN (modèle à aire limitée) est initialisée en utilisant la méthode d'OI (paragraphe 1.3.1) pour l'assimilation des températures et de l'humidité du sol à 2 m de hauteur du réseau SYNOP. Les dernières recherches au CNRM (Centre National de Recherches Météorologiques) portent sur l'implémentation d'un prototype d'un schéma d'assimilation 2D-variationnel simplifié [2] (basé sur la linéarisation de l'opérateur d'observation au moyen d'une approximation linéaire par différences finies) qui permettra d'améliorer les analyses d'OI. Ce système permet l'analyse de l'humidité du sol ainsi que d'autres variables toutes aussi importantes pour caractériser l'état de la surface, telles que la biomasse de la végétation ou la température du sol, sont facilement intégrables dans le système global. L'assimilation d'observations de l'humidité de la surface et du LAI (Leaf Area Index) à échelle locale avec la version 1D a été testée sur le site de SMOSREX [24] au Sud-Ouest de Toulouse pour l'analyse du réservoir d'eau utile du sol et de la biomasse de la végétation [53]. Ce résultat est présenté sur la figure 14. Les trois graphiques représentent l'évolution temporelle de l'humidité de la zone racinaire ( $w_2$ ), de la biomasse de la végétation et du LAI sur la période 2001-2004 sur le site de SMOSREX. Chaque graphique compare la simulation du modèle de surface ISBA-A- $g_s$  (simulation de contrôle), les observations et les analyses. On constate que le modèle n'est pas capable de bien reproduire les épisodes de sécheresse des années 2003 et 2004. Ce problème est dû à l'incapacité du modèle à représenter les valeurs de l'humidité du sol inférieures au point de flétrissement. En été des années 2003 et 2004, l'humidité du sol a été observée au dessous du point de flétrissement. L'assimilation améliore la simulation brute de l'humidité et des variables de végétation du modèle ISBA-A- $g_s$ . La valeur ajoutée de l'assimilation est clairement observée sur les périodes de sécheresse. Même si les analyses sont encore loin des observations, elles sont capables cependant de fournir des états de l'humidité initiale plus réalistes au dessous du point de flétrissement "théorique" du sol.

L'extension de ce schéma à deux dimensions est en cours de validation sur le bassin Adour-Garonne, qui inclut la région Atlantique-Méditerranée au Sud-Ouest de la France et qui offre une grande diversité des conditions bio-climatiques. Les observations du LAI des satellites MODIS (Moderate Resolution Imaging Spectroradiometer) et SPOT/VGT, ainsi que les observations de l'humidité du sol du réseau SMOSMANIA de Météo-France serviront comme produits de validation des observations de télédétection, spécialement celles qui seront fournies par le futur satellite SMOS (Soil Moisture and Ocean Salinity) de l'ESA (European Space Agency) en bande L.

### 2.5.3 Caractérisation de l'état des rivières

Une application importante de l'assimilation de données pour l'hydrologie est la caractérisation de l'état des rivières au moyen de variables telles que le débit ou la hauteur du niveau de l'eau à l'exutoire d'un bassin versant. Le CERFACS a déjà engagé quelques études préliminaires sur la création d'une maquette d'assimilation pour l'hydrologie. L'objectif fondamental de la maquette est l'amélioration du débit total simulé du bassin de l'Ardèche à Vogüe et du bassin du Gardon à Anduze. Pour cela, une version simplifiée du modèle TOPModel a été utilisée. Les résultats préliminaires ont montré la réduction des écarts [débit analysé-débit observé] des hydrogrammes par rapport aux écarts [débit simulé-débit observé]. Cela montre l'intérêt d'intégrer une structure complète d'assimilation de données dans la modélisation hydrologique. L'objectif est double, il s'agit non seulement de l'amélioration de la capacité de prévision directement sur les variables pronostiques de ces modèles (débit total de rivières, hauteur de la zone racinaire, etc.), mais aussi du recalage optimal de leurs paramètres d'entrée et/ou des conditions limites afin de les ajuster à la réalité des observations. Dans ce contexte, le CERFACS en collaboration avec le SCHAPI, a

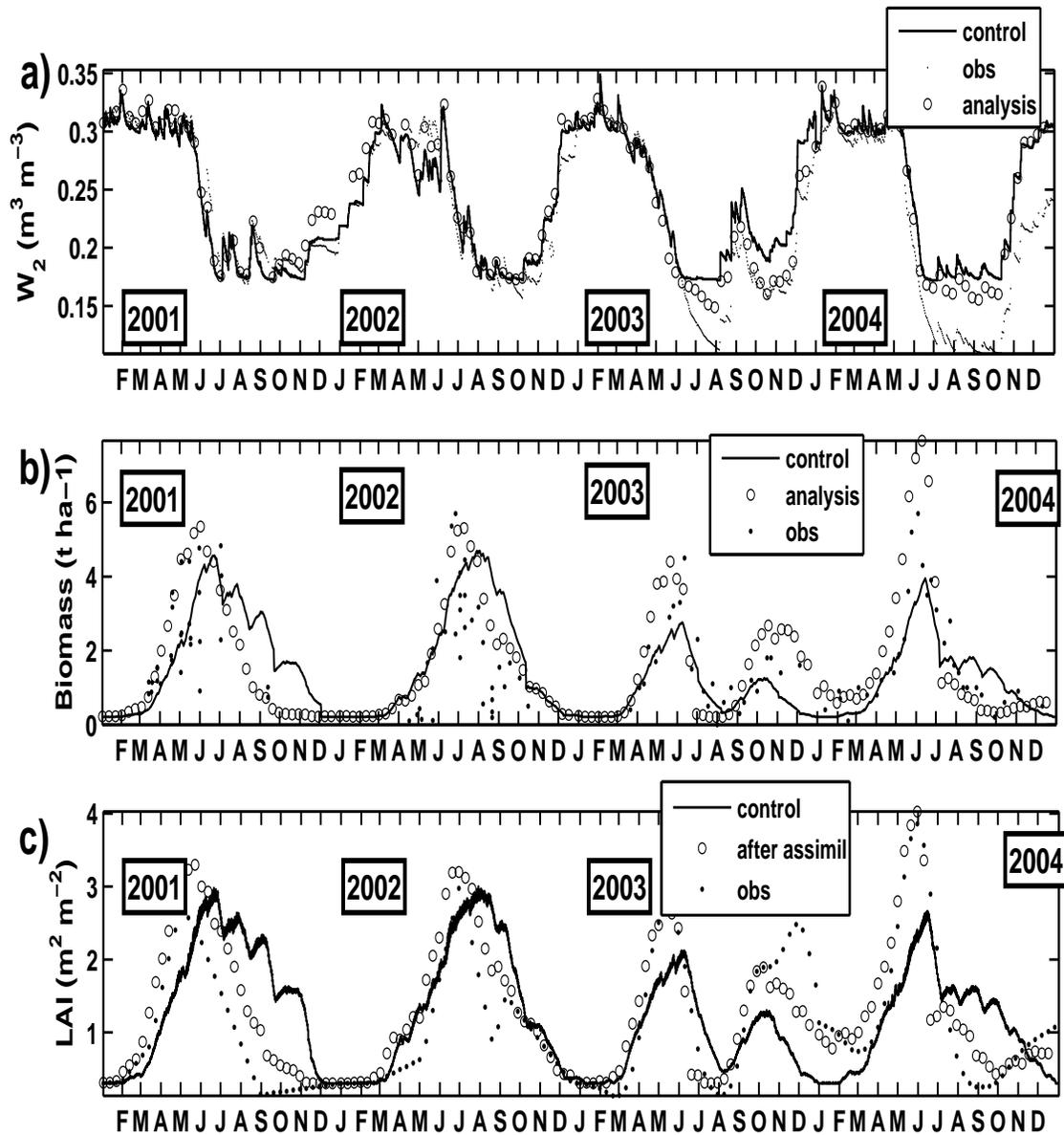


FIG. 14 – Analyses de a) l'humidité de la zone racinaire (cercles) et b) la biomasse de la végétation, avec la méthode 1D-VAR simplifié pour la période 2001-2004 sur le site de SMOSREX. c) LAI avant et après l'assimilation des observations simultanées de l'humidité de la surface et de LAI. Par motifs de comparaison, les valeurs analysées sont superposées aux observations (points) et la simulation du modèle (ligne solide).

souhaité tester l'intégration de plusieurs modèles dans une structure d'assimilation de données :

- La plateforme de modélisation ATHYS décrit la propagation de crue d'un bassin-versant. Cette plateforme est déjà utilisée au niveau pré-opérationnel sur plusieurs bassins français. Elle réunit un ensemble de modèles associés à des traitements de données hydro-climatiques et géographiques. Son outil de modélisation spatiale MERCEDES permet de simuler différents épisodes de crues avec des fonctions de production pluie/débit et des fonctions de transfert à l'exutoire du bassin.
- La plateforme MASCARET décrit la propagation de crue d'estuaire. Cette plateforme est

aussi utilisée sur plusieurs bassins français avec une composante de type maritime. Il s'agit d'un code qui résout les équations de Saint-Venant, contrairement à l'approche des modèles hydrologiques de type conceptuel utilisés dans d'ATHYS.

- La chaîne SIM (SAFRAN-ISBA-MODCOU) opérationnelle à Météo-France. Elle analyse en temps réel l'eau du sol et le débit des fleuves pour la France. Ce système utilise les forçages atmosphériques SAFRAN (Système d'Analyse Fournissant des Renseignements Atmosphériques pour la Nivologie) pour estimer l'humidité de trois réservoirs du sol interpolés aux points de grille du modèle de surface ISBA. Les estimations d'excès d'eau superficielle (ruissellement) et d'excès d'eau profonde (drainage) d'ISBA servent ensuite à alimenter le modèle hydrologique MODCOU, qui simule le débit des rivières.
- Le modèle récemment développé DASSFLOW. Il s'agit d'un modèle de type Saint-Venant en 1D ou 2D, spécialement développé pour l'assimilation de données. Ce code a déjà testé avec succès l'assimilation d'observations de type eulériennes (observations faites à un point fixe du domaine d'étude) et de type lagrangiennes (observations effectuées au fil de l'eau de quantités transportées par le courant), comme par exemple des observations de trajectoires ([36]).

Dans un premier temps on étudiera la possibilité d'implémenter une méthode simple d'assimilation de données pour l'assimilation des observations systématiques du débit du réseau hydrographique français sur un bassin versant, telle qu'un filtre de Kalman ou une méthode variationnelle simplifiée [2]. L'objectif premier est d'améliorer la précision des simulations des hydrogrammes, en particulier pour estimer le temps d'occurrence et l'intensité du pic de l'onde de crue. Après validation de cette méthode, on envisage d'implémenter des méthodes plus sophistiquées afin de prendre en compte la dynamique du modèle (section 1.3). Nous appliquerons cette approche à la propagation d'une crue dans un bassin-versant ainsi que pour la propagation d'une crue dans un estuaire. Ces méthodes devront être capables d'intégrer divers types d'observations, telles que les observations ponctuelles des vitesses de l'eau ou de mesures provenant des systèmes de télédétection.

Le cadre décrit ci-dessus permet tout un éventail de possibilités pour assimiler différents types d'observations hydrologiques avec pour objectif majeur d'améliorer la capacité de prévision de l'évolution des crues suite à un épisode de pluie important.

### 3 L'assimilation de données dans les Services de Prédiction de Crues

Cette section détaille la situation actuelle de l'utilisation de l'assimilation de données dans les différents SPC de France. Le tableau 2 résume ces activités. La dernière colonne du tableau classe les SPC en trois catégories en fonction du degré de connaissance et d'utilisation de l'assimilation de données, selon :

1. classe 1 : certaines structures d'assimilation complexes existent en mode opérationnel ;
2. classe 2 : utilisation de techniques d'analyses simples ;
3. classe 3 : aucune utilisation actuelle de l'assimilation de données.

D'un point de vue général, l'assimilation de données est peu utilisée par les SPC, au sens où toute l'information disponible (ébauche des modèles hydrologiques, observations, qualité de l'information dont on dispose) n'est pas utilisée pour développer des systèmes optimaux de prédiction de crue. Cependant, certaines observations sont utilisées pour analyser et corriger partiellement la valeur des variables de contrôle (débit, hauteur du niveau de l'eau, remplissage de réservoirs, etc.). Une bonne partie de ces techniques consistent en la recalibration (en temps-réel ou non) des paramètres d'entrée de modules de production pluie/débit et des modules de transfert débit/débit pour ajuster les simulations à la réalité des observations. Parfois, le calage de ces paramètres est fait "à la main" en prenant comme référence des événements de pluie et en trouvant un jeu de paramètres qui ajuste au mieux les simulations aux observations. Dans d'autres cas, les paramètres sont fixés à partir du tirage d'un nombre élevé de combinaisons possibles (méthodes de Monte-Carlo). Néanmoins, ces méthodes n'assurent pas l'optimalité du jeu de paramètres obtenu pour la prédiction d'événements de crues ultérieures.

On peut s'interroger sur les points suivants : l'approche multi-modèle pure donne-t-elle vraiment le meilleur estimateur d'une variable de contrôle? Est-il légitime d'appliquer des règles simples d'ajustement de la prédiction à l'observation, telles que l'ajustement brutal de la prédiction à la dernière valeur observée, l'ajustement lissé de l'hydrogramme par rapport aux dernières valeurs observées, etc. ?

Malgré les techniques d'analyse et les modèles numériques mis en place par les SPC, il reste encore des améliorations pour une prédiction précise des phénomènes de crue possibles. Étant donné l'utilisation relativement faible par les SPC de systèmes d'assimilation de données basés sur une théorie d'estimation statistique ou de contrôle optimal, il est probable que leur implémentation pourrait apporter un bénéfice aux prévisions de crue des rivières dans les bassins versants.

# Carte des territoires des SPC



MEDD/DE/SCHAPI  
05/07/2006

		<b>assimilation</b>			
	<b>SPC</b>	<b>actuel</b>	<b>envisagé</b>	<b>modèles-techniques utilisées</b>	<b>classe</b>
1	Artois-Picardie	O		utilisation de l'autorégression avec modèles Sophie, Gardénia, réseaux de neurones, etc.	1
2	Meuse-Moselle	O		modèles Moïse avec BCEOM, et GR en Stream	3
3	Rhin-Sarre	N			3
4	Normandie	N			3
5	Oise-Aisne	O		assimilation avec les modèles Hydra et Topmodel	2
6	Seine amont-Marne	N	O	modèle Mascaret	3
7	Seine-Yonne	O		assimilation avec les modèles GR3P et Cassandre	1
8	Vilaine-Côtiers Bretons	N	O	projet d'utilisation du modèle MIKE	3
9	Maine-Loire aval	O		assimilation avec le modèle Moïse, mais pas transparent	2
10	Loire-Cher-Indre	N	O	projet d'assimilation avec le modèle Hydra	3
11	Allier	O		utilisation du logiciel PACHA	1
12	Vienne-Thouet	O		modification de relations Q-Q en terme de coefficients et de temps de propagation	2
13	Littoral Atlantique	O		ajustements simples de prévisions aux observations, recalage de variables intermédiaires et ajustement de paramètres du modèle	2
14	Dordogne	N			2
15	Garonne	N			2
16	Adour (voir SPC 23)				2
17	Tam-Lot	O		utilisation du modèle Hydra	2
18	Rhône amont-Saône	N			3
19	Alpes du Nord	N	O	projet d'assimilation avec l'adaptation du modèle hydraulique Carima	3
20	Grand Delta	O		minimisation de l'écart quadratique par l'algorithme du gradient réduit généralisé	1
21	Méditerranée Ouest	O		utilisation des modèles GR3 et Moïse avec BCEOM, Stream et Isis, Mascaret	2
22	Méditerranée Est				2
23	SCHAPI	O		utilisation des modèles GR en Stream avec Moïse (BCEOM en CS)	2

TAB. 2 – utilisation de l'assimilation de données par les SPC de France.

# A Algorithmes de minimisation

## A.1 Méthode du gradient

Premièrement, on va considérer la forme quadratique suivante :

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - b^T \mathbf{x} + c \quad (35)$$

avec  $\mathbf{A}$  une matrice de dimension  $n \times n$ , symétrique ( $\mathbf{A}^T = \mathbf{A}$ ), définie positive (c.à.d.,  $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$  pour tout vecteur non-nul  $\mathbf{x} \in \mathbb{R}^n$ ). Cette fonction quadratique est équivalente à la fonction coût de la méthode variationnelle (équation (16)). Pour obtenir cette forme il est nécessaire d'effectuer une série de transformations qui ne sont pas détaillées dans ce document. Pour minimiser cette fonction on résout :

$$\begin{aligned} \nabla f &= 0 \\ &= \mathbf{A} \mathbf{x} - b \end{aligned} \quad (36)$$

c'est à dire, que le minimum de  $f$  est la solution du système linéaire  $\mathbf{A} \mathbf{x} = b$ . L'objectif de la méthode du gradient est de rechercher la direction  $\mathbf{s}$  de descente de plus grande pente d'une fonction  $f$  de plusieurs variables. Cette direction de descente est proportionnelle au gradient de la fonction  $f$  ( $\mathbf{s} \propto \nabla f$ ). Pour identifier un point plus proche du minimum de la fonction  $f$  on effectue un déplacement (pas) vers la direction de l'antigradient au moyen de la loi de récurrence suivante :

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \gamma_k \mathbf{s}_k \quad (37)$$

Un déplacement de longueur  $\gamma > 0$  correctement choisi assure la décroissance de  $f$ . Pour la fonction quadratique (35) le paramètre  $\gamma_{k+1}$  est calculé comme l'unique minimiseur de la fonction  $\Phi(\gamma) = f(\mathbf{x}_k - \gamma \mathbf{s}_k)$  dépendant seulement de la variable réelle  $\gamma$ . La solution de l'équation  $\Phi'(\gamma) = 0$  est :

$$\gamma_k = \frac{\mathbf{s}_k^T \mathbf{s}_k}{\mathbf{s}_k^T \mathbf{A} \mathbf{s}_k} \quad (38)$$

Ainsi, la loi de récurrence (37) devient :

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{\mathbf{s}_k^T \mathbf{s}_k}{\mathbf{s}_k^T \mathbf{A} \mathbf{s}_k} \mathbf{s}_k \quad (39)$$

À partir d'un point de recherche initial  $\mathbf{x}_0$ , ce processus est répété itérativement jusqu'à atteindre un critère de convergence. Généralement ce critère compare la norme du gradient à une valeur négligeable. Le désavantage de cette méthode est que si le pas de recherche est très petit, alors la minimisation est très lente et nécessite des nombreuses itérations (et pourtant de minimisations peu-coûteuses à faire). En revanche si le pas est très grand, la méthode risque d'osciller autour du minimum de la fonction à minimiser.

## A.2 Gradient conjugué

Contrairement à la méthode du gradient, la méthode du gradient conjugué recherche plusieurs directions de descente. Pour autant ceci est fait sans le surcoût du calcul des dérivées secondes impliqué dans la méthode de Newton (section ??). C'est au moyen d'un procès orthogonal de Gram-Schmidt que l'on parvient à construire les directions conjuguées  $\mathbf{d}_k$  du gradient. On choisit alors la direction la plus proche du gradient  $\mathbf{s}_k$  sous la contrainte du conjugué. On définit le résidu à l'itération  $k$  par :

$$\mathbf{r}_k = \mathbf{A}\mathbf{x}_k - b \quad (40)$$

À partir de cette définition et d'une première approximation  $\mathbf{x}_0$  de la solution du système  $\mathbf{A}\mathbf{x} = b$ , on présente l'algorithme du gradient conjugué :

1. calcul du résidu initial pour la première direction de descente :  $\mathbf{r}_0 = \mathbf{A}\mathbf{x}_0 - b$ ;  $\mathbf{s}_0 = -\mathbf{r}_0$ ,
2. loi de récurrence pour le calcul du pas de descente :  $\gamma_k = \frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{s}_k^T \mathbf{A}\mathbf{s}_k}$ ,
3. loi de récurrence pour la recherche de la solution :  $\mathbf{x}_{k+1} = \mathbf{x}_k + \gamma_k \mathbf{s}_k$ ,
4. loi de récurrence pour le calcul du résidu :  $\mathbf{r}_{k+1} = \mathbf{r}_k + \gamma_k \mathbf{A}\mathbf{s}_k$ ,
5. calcul du coefficient  $\beta$  :  $\beta_{k+1} = \frac{\mathbf{r}_k^T \mathbf{r}_k}{\mathbf{r}_{k+1}^T \mathbf{A}\mathbf{r}_{k+1}}$
6. loi de récurrence pour la direction de recherche :  $\mathbf{s}_{k+1} = -\mathbf{r}_{k+1} + \beta_{k+1} \mathbf{s}_k$

On itère les étapes 2 à 6 jusqu'à convergence. Cette méthode n'utilise pas beaucoup de mémoire quand la matrice  $\mathbf{A}$  est symétrique définie positive et est beaucoup plus efficace que la méthode du gradient. En conséquence, cette méthode a besoin d'un nombre inférieur d'itérations pour converger (voir FIG. ??). En revanche, si le point initial est loin de la solution, elle risque de ne pas converger, et le temps de calcul par itération est plus grand. Finalement, il est important de bien conditionner la matrice  $\mathbf{A}$  (section ??) afin de faciliter la convergence.

## B Tableau-résumé de l'assimilation au CERFACS

Voir dernière page de ce document.

## Références

- [1] A.T.Weaver, J.Vialard, and D.L.T. Anderson. Three and four-dimensional variational assimilation with a general circulation model of the tropical pacific ocean.Part I : formulation, internal diagnostics and consistency checks. *Monthly Weather Review*, 131 :1360–1378, 2003.
- [2] G. Balsamo, F. Bouyssel, and J. Noilhan. A simplified bi-dimensional variational analysis of soil moisture from screen-level observations in a mesoscale numerical weather-prediction model. *Q.J.R.Meteorological Society*, 130 :895–915, 2004.
- [3] F. Baret, C. Bacour, M. Weiss, K. Pavageau, D. Béal, V. Bruniquel, P. Regner, J. Moreno, C. Gonzalez, , and J. Chen. Canopy biophysical variables estimation from MERIS observations based on neural networks and radiative transfer modelling : principles and validation. In *35th COSPAR Scientific Assembly*, page 1677, 2004.
- [4] F. Baret, C. Bacour, M. Weiss, K. Pavageau, D. Béal, V. Bruniquel, P. Regner, J. Moreno, C. Gonzalez, , and J. Chen. Canopy biophysical variables estimation from meris observations based on neural networks and radiative transfer modelling : principles and validation. *Q.J.R.Meteorological Society*, 130 :895–915, 2004.
- [5] B.C.Gao. NDWI : A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, 58 :257–266, 1996.
- [6] F. Bouttier and P. Courtier. Data assimilation concepts and methods. Technical report, European Centre for Medium-Range Weather Forecasts, march 1999.
- [7] A.M. Bratseth. Statistical interpolation by means of successive corrections. *Tellus*, 38(A) :439–447, 1986.
- [8] R.G. Brown and P.Y.C. Hwang. *Introduction to Random Signals and Applied Kalman Filtering*. John Wiley & Sons, New York, NY, 3rd ed., 1997.
- [9] G. Burgers, P.J. Van Leeuwen, and G. Evensen. Analyses scheme in the ensemble kalman filter. *Monthly Weather Review*, 126 :1719–1724, 1998.
- [10] J.-C. Calvet. Investigating soil and atmospheric plant water stress using physiological and micrometeorological data. *Agricultural and Forest Meteorology*, 103 :229–247, 2000.
- [11] J.-C. Calvet, P. Bessemoulin, J. Noilhan, C. Berne, I. Braud, D. Courault, N. Fritz, E. Gonzalez-Sosa, J. P. Goutorbe, R. Haverkamp, G. Jaubert, L. Kergoat, G. Lachaud, J. P. Laurent, P. Mordelet, A. Olioso, P. Péris, J. L. Roujean, J.-L. Thony, C. Tosca, M. Vauclin, and D. Vignes. Murex : a land-surface field experiment to study the annual cycle of the energy and water budgets. *Annales Geophysicae.*, 17 :838–854, 1999.
- [12] J.-C. Calvet and J. Noilhan. From near-surface to root-zone soil moisture using year-round data. *Journal of Hidrometeorology*, 1(5) :393–411, 2000.
- [13] J.-C. Calvet, J. Noilhan, and P. Bessemoulin. Retrieving the root-zone soil moisture from surface soil moisture or temperature estimates : A feasibility study based on field measurements. *Journal of Applied Meteorology*, 37(4) :371–386, 1998.
- [14] J.-C. Calvet, J. Noilhan, J.L. Roujean, P. Bessemoulin, M. Cabelguenne, A. Olioso, and J.P. Wigneron. An interactive vegetation SVAT model tested against data from six contrasting sites. *Agricultural and Forest Meteorology*, 92 :73–95, 1998.
- [15] J.-C. Calvet, V. Rivalland, C. Picon-Cochard, and J.-M. Guehl. Modelling forest transpiration and CO<sub>2</sub> fluxes - response to soil moisture stress. *Agricultural and Forest Meteorology*, 124 :143–156, 2004.
- [16] J.-C. Calvet and J.F. Soussana. Modelling CO<sub>2</sub>-enrichment effects using an interactive vegetation SVAT scheme. *Agricultural and Forest Meteorology*, 108 :129–152, 2001.
- [17] J.C. Calvet, P.Viterbo, P. Ciais, B. Van-Den-Hurk, E. Moors, A. Kaptein, M. Leroy, and J. Mu noz Sabater. Assimilation of remote sensing data to monitor the terrestrial carbon cycle : the carbon observatory of geoland. 2004.

- [18] P. Cayrol, A. Chehbouni, L. Kergoat, G. Dedieu, P. Mordelet, and Y. Nouvellon. Grassland modeling and monitoring with spot-4 vegetation instrument during the 1997-1999 salsa experiment. *Agricultural and Forest Meteorology*, 105 :91–115, 2000.
- [19] S.E. Cohn. An introduction to estimation theory. *Journal of the Meteorological Society of Japan*, 75(1B), 1997.
- [20] G.P. Cressman. An operational objective analysis system. *Monthly Weather Review*, 87(10) :367–374, 1959.
- [21] J.W. Deardorff. A parameterization of ground-surface moisture content for use in atmospheric prediction models. *Journal of Applied Meteorology*, 16 :1182,1185, 1977.
- [22] J.W. Deardorff. Efficient prediction of ground temperature and moisture with inclusion of a layer of vegetation. *Journal of Geophysical Research*, 83 :1889–1903, 1978.
- [23] D.W. Deering. *Rangeland reflectance characteristics measured by aircraft and spacecraft sensors*. PhD thesis, Texas A&M University, 338 pp., 1978.
- [24] P. DeRosnay, J.C. Calvet, Y. Kerr, J.P. Wigneron, F. Lemaitre, M.J. Escorihuela, J. Muñoz Sabater, and et al. SMOSREX : A long term field campaign experiment for soil moisture and land surface processes remote sensing. *Remote Sensing of Environment*, 102 :377–389, 2006.
- [25] G. Desroziers. Optimal estimation in meteorology. Technical report, ALATNET seminar, june 2001.
- [26] D.Giard and E.Bazile. Implementation of a new assimilation scheme for soil and surface variables in a global nwp model. Technical report, Météo-France / CNRM.
- [27] D. Entekhabi, H. Nakamura, and E. Njoku. *Passive Microwave Remote Sensing of Land-Atmosphere Interactions*, pages 485–498. VPS publisher, 1995.
- [28] G. Evensen. Sequential data assimilation with a non linear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99 :10143–10162, 1994.
- [29] G. Evensen. The ensemble kalman filter : theoretical formulation and practical implementation. *Ocean Dynamics*, 53 :343–367, 2003.
- [30] F.Rabier and P.Courtier. Four-dimensional assimilation in the presence of baroclinic instability. *Q.J.R.Meteorological Society*, 118 :649–672, 1992.
- [31] C. Francois, A. Quesney, and C. Ottlé. Sequential assimilation of ERS-1 and ERS-2 data into a coupled land surface-hydrological model using an extended kalman filter. *Journal of Hydrometeorology*, 4 :473–486, 2003.
- [32] P. Viterbo et al. G. Seuffert, H. Wilker. Soil moisture analysis combining screen-level parameters and microwave brightness temperature : A test with field data. *Geophysical Research Letters*, 30(10) :1498–1501, 2003.
- [33] A. Gelb. *Applied Optimal Estimation*, page 382 pp. Analytic Science Corporation, 1974.
- [34] D. Giard and E. Bazile. Implementation of a new assimilation scheme for soil and surface variables in a global nwp model. Technical report, Météo-France / CNRM.
- [35] A.-L. Gibelin, J.-C. Calvet, J.-L. Roujean, L. Jarlan, and S.O. Los. Ability of the land surface model ISBA-A-g<sub>s</sub> to simulate leaf area index at the global scale : Comparison with satellites products. *Journal of Geophysical Research*, 111(D18102, doi :10.1029/2005JD006691), 2006.
- [36] M. Honnorat. *Assimilation de données lagrangiennes pour la simulation numérique en hydraulique fluviale*. PhD thesis, Institut National Polytechnique de Grenoble, 2007.
- [37] E. Isaacson and H. Bishop Keller. *Analysis of numerical methods*. Courier Dover Publications, 1994.
- [38] C.M.J. Jacobs, B.J.J.M. Van-Den-Hurk, and H.A.R. de Bruin. Stomatal behaviour and photosynthetic rate of unstressed grapevines in semi-arid conditions. *Agricultural and Forest Meteorology*, 80 :111–134, 1996.

- [39] P.G. Jarvis. The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. *Philosophical Transactions of the Royal Society of London*, 273(927) :593–610, 1976.
- [40] C.F. Jordan. Derivation of leaf area index from quality of light on the forest floor. *Ecology*, 50 :663–666, 1969.
- [41] A.G. Journel. Kriging in terms of projections. *Mathematical Geology*, 9 :563–586, 1977.
- [42] S.J. Julier and J.K. Uhlmann. Unscented filtering and nonlinear estimation. In *IEEE Review 92(3)*, pages 401–422, 2004.
- [43] Y. Kerr, P. Waldteufel, J.-P. Wigneron, J.-M. Martinuzzi, J. Font, and M. Berger. Soil moisture retrieval from space : The soil moisture and ocean salinity (SMOS) mission. *IEEE Trans. Geosc. Remote Sens.*, 39(8) :1729–1735, 2001.
- [44] Y.H. Kerr, P. Waldteufel, J.-P. Wigneron, J. Font, and M. Berger. The soil moisture and ocean salinity mission. In *International Geoscience and Remote Sensing Symposium, Toulouse, France, 21–25 July 2003*, 2003.
- [45] F. Lemaitre, J.C. Poussière, Y. Kerr, M. Dejus, R. Durbe, and P. De Rosnay et al. Design and test of the ground based L-band radiometer for estimating water in soils (LEWIS). *IEEE Trans. Geosc. Remote Sens.*, 42(8) :1666–1676, 2004.
- [46] F.L. Lewis. *Optimal Estimation, with an Introduction to Stochastic Control Theory*, page 400 pp. John Wiley & Sons, 1986.
- [47] A. Lorenc. Analysis methods for numerical weather prediction. *Q.J.R. Meteorological Society*, 112 :1177–1194, 1986.
- [48] J.-F. Mahfouf. Analysis of soil moisture from near-surface parameters : a feasibility study. *Journal of Applied Meteorology*, 30 :1534–1547, 1991.
- [49] R.T. McNider, A.J. Song, D.M. Casey, P.J. Wetzel, W.L. Crosson, and R. Rabin. Toward a dynamic-thermodynamic assimilation fo satellite surface temperature in numerical atmospheric models. *Monthly Weather Review*, 122 :2784–2803, 1994.
- [50] M.Fischer and E.Anderson. Developments in 4D-VAR and kalman filtering. TR 347. Technical report, European Centre for Medium-Range Weather Forecast, Reading, UK, 2001.
- [51] M.Jing and A.Leetmaa. Impact of data assimilation on ocean initialization and El Niño prediction. *Monthly Weather Review*, 125(5) :742–753, 1997.
- [52] J. Muñoz-Sabater, J.C. Calvet, and P. De Rosnay. On the assimilation of multispectral remote sensing data in a svat model. 2004.
- [53] J. Muñoz-Sabater, L.Jarlan, J.C. Calvet, F. Bouyssel, and P. De Rosnay. From near surface to root-zone soil moisture using different assimilation techniques. *Journal of Hidrometeorology*, 8(2) :194–206, 2007.
- [54] M.Weiss, D.Trouffleau, F.Baret, H.Chauki, L.Prevot, A.Oliosio, N.Bruguier, and N.Brisson. Coupling canopy functioning and radiative transfer models for remote sensing data assimilation. *Agricultural and Forest Meteorology*, 108 :113–128, 2001.
- [55] NASA. *Studying Earth’s Environment From Space*, 2000.
- [56] R.R. Nemani and S.W. Running. Estimation of regional surface resistance to evapotranspiration from NDVI and thermal-IR AVHRR data. *Journal of Applied Meteorology*, 28 :276–284, 1989.
- [57] W. Ni-Meister, P.R. Houser, and J.P. Walker. Soil moisture initialization for climate prediction : Assimilation of scanning multifrequency microwave radiometer soil moisture data into a land surface model. *Journal of Geophysical Research*, 111(D20102) :doi:10.1029/2006JD007190, 2006.
- [58] J. Noilhan and J.F. Mahfouf. The ISBA land surface parameterisation scheme. *Global and Planetary Change*, 13 :145–159, 1996.

- [59] J. Noilhan and S. Planton. A simple parameterization of land surface processes for meteorological models. *Monthly Weather Review*, 117 :536–549, 1989.
- [60] V.J. Ostdiek and D.J. Bord. *Inquiry into physics*, chapter 6. West Publishing Company, 2005.
- [61] M. Pardé, J.-P. Wigneron, P. Waldteufel, Y. Kerr, Y.H. Chanzy, and S.S. Sobjaerg. N-parameter retrievals from l-band microwave observations acquired over a variety of crop fields. *IEEE Transactions on Geoscience and Remote Sensing*, 42(6) :1168–1178, 2004.
- [62] T. Pellarin, J.-P. Wigneron, J.-C. Calvet, and P. Waldteufel. Global soil moisture retrieval from a synthetic l-band brightness temperature data set. *Journal of Geophysical Research*, 108(D12) :ACL :9.1–9.15, 2003.
- [63] R.Daley. *Atmospheric data analysis*, page 457 pp. Cambridge University, 1991.
- [64] B. Reed, J. Brown, D. Van der Zee, T. Loveland, and J. Merchant. Vegetation effects on the microwave emission of soils. *Journal of Vegetation Science*, 5 :703–714, 1994.
- [65] R.H. Reichle, D.B. McLaughlin, and D.Entekhabi. Variational data assimilation of microwave radiobrightness observations for land surface hydrologic applications. *IEEE Transactions on Geoscience and Remote Sensing*, 39(8) :1708–1718, 2001.
- [66] J.L. Roujean and R. Lacaze. Global mapping of vegetation parameters from POLDER multiangular measurements for studies of surface-atmosphere interactions : A pragmatic method and its validation. *Journal of Geophysical Research*, 107(D12) :4150, doi :10.1029/2001JD000751, 2002.
- [67] J.L. Roujean, M. Leroy, and P.Y. Deschamps. A bidirectional reflectance model of the earth’s surface for the correction of remote sensing data. *Journal of Geophysical Research*, 97 :20455–20468, 1992.
- [68] J. Muñoz Sabater. *Assimilation de données de télédétection pour le suivi des surfaces continentales : mise en œuvre sur un site expérimental*. PhD thesis, Université Toulouse III, 163 pp., 2007.
- [69] K. Saleh, J.-P. Wigneron, P. De Rosnay, J.-C. Calvet, M. J. Escorihuela, Y. Kerr, and P. Waldteufel. Impact of rain interception by vegetation and mulch on the L-band emission of natural grass. *Remote Sensing of Environment*, 101(1) :127–139, 2006.
- [70] K. Saleh, J.-P. Wigneron, P. De Rosnay, J.-C. Calvet, and Y. Kerr. Semi-empirical regressions at L-band applied to surface soil moisture retrievals over grass. *Remote Sensing of Environment*, 101(3) :415–426, 2006.
- [71] O. Samain. *Fusion multi-capteurs de données satellitaires optiques pour la restitution de variables biophysiques de surface*. PhD thesis, Université Toulouse III, 190 pp., 2006.
- [72] S.Buis and S.Massart. Implémentation d’une maquette PALM de système d’assimilation d’activités neutroniques avec le modèle COCCINELLE. Technical report, Technical Report TR/CMGC/04/67, CERFACS, Toulouse, France 2004.
- [73] T.J. Schmugge. Remote sensing of soil moisture : Recent advances. *IEEE Trans. Geosc. Remote Sens.*, GE21 :334–344, 1983.
- [74] S.Jacquemoud and F.Baret. PROSPECT : A model of leaf optical properties spectra. *Remote Sensing of Environment*, 34 :75–91, 1990.
- [75] S.Massart. *Méthodologies de l’assimilation de données. Application à la chimie atmosphérique et à la mécanique des structures*. PhD thesis, Doctorat de L’I.N.P.T, 270 pp, 2003.
- [76] S.Ricci. *Assimilation variationnelle océanique : modélisation multivariée de la matrice de covariance d’erreur d’ébache*. PhD thesis, Université Toulouse III, 275 pp., 2004.
- [77] D.O. Staley and G.M. Jurica. Effective atmospheric emissivity under clear skies. *Journal of Applied Meteorology*, 11 :349–356, 1972.
- [78] J.-C. Calvet T. Pellarin and J.-P. Wigneron. Surface soil moisture retrieval from l-band radiometry : A global regression study. *IEEE Trans. Geosc. Remote Sens.*, 41(9) :2037–2051, 2003.

- [79] J.-C. Calvet et al. T. Pellarin, J.-P. Wigneron. Two-year global simulation of l-band brightness temperatures over land. *IEEE Trans. Geosc. Remote Sens.*, 41(9) :2135–2139, 2003.
- [80] O. Talagrand. Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan*, 75(1B) :191–209, 1997.
- [81] A. Tarantola and B. Valette. Generalized nonlinear inverse problems solved using the least squares criterion. *Reviews of Geophysics and Space Physics*, 20(2) :219–232, 1982.
- [82] A. Tarantola and B. Valette. Inverse problems = quest for information. *Journal of Geophysical Research*, 50 :159–170, 1982.
- [83] P.J.G. Teunissen. *Adjustment Theory*, chapter 1, page 193 pp. Delft University Press, 1999.
- [84] A.H. Thiessen. Precipitation averages for large areas. *Monthly Weather Review*, 39 :1082–1089, 1911.
- [85] Y. Tian, C.E. Woodcock, Y. Wang, J.L. Privette, N.V. Shabanov, L. Zhou, W. Buermann, J. Dong, B. Veikkanen, T. Hame, M. Ozdogan, Y. Knyazikhin, and R.B. Myneni. Multiscale analysis and validation of the MODIS LAI product I. Uncertainty assessment. *Remote Sensing of Environment*, 83 :414–430, 2002.
- [86] Y. Tian, C.E. Woodcock, Y. Wang, J.L. Privette, N.V. Shabanov, L. Zhou, W. Buermann, J. Dong, B. Veikkanen, T. Hame, M. Ozdogan, Y. Knyazikhin, and R.B. Myneni. Multiscale analysis and validation of the MODIS LAI product II. Sampling strategy. *Remote Sensing of Environment*, 83 :431–441, 2002.
- [87] T.Lagarde, A.Piacentini, and O.Thual. A new representation of data assimilation methods : the PALM flow charting approach. *Q.J.R.Meteorological Society*, 127 :189–207, 2001.
- [88] W. Tobler. Smooth pycnophylactic interpolation for geographic regions. *Journal of the American Statistical Association*, 74(367) :519–536, 1979.
- [89] B. Van-Den-Hurk and H. The. Assimilation of satellite derived surface heating rates in a numerical weather prediction model. Technical report, KNMI, The Netherlands, 2002.
- [90] W. Verhoef. Light scattering by leaf layers with application to canopy reflectance modeling : The sail model. *Remote Sensing of Environment*, 141 :16–125, 1984.
- [91] N. Viovy, C. Francois, A. Bondeau, G. Krinner, J. Polcher, L. Kergoat, G. Dedieu, N. De Noblet, P. Ciais, and P. Friedlingstein. Assimilation of remote sensing measurements into the ORCHIDEE/STOMATE DGVM biosphere model. In *8<sup>th</sup> International Symposium on physical measurements and signatures in remote sensing*, pages 713–718, CNES/ISPRS, Aussois, France, 2001. M.Leroy.
- [92] P. Viterbo and P. Courtier. The importance of soil water for medium-range weather forecasting. implications for data assimilation. In *WMO workshop on imbalances of slowly varying components of predictable atmospheric motions, Beijing, China, March 1995*, pages 121–130, 1995.
- [93] W. Wagner, K. Scipal, C. Pathe, D. Gerten, W. Lucht, and B. Rudolf. Evaluation of the agreement between the first global remotely sensed soil moisture data with model and precipitation data. *Journal of Geophysical Research*, 108 :4611–4626, 2003.
- [94] J.P. Walker. *Estimating soil moisture profile dynamics from near surface soil moisture measurements and standard meteorological data*. PhD thesis, Dept. of Civil, Surveying and Environmental Engineering, Univ. of Newcastle, Callaghan, Australia, 1999.
- [95] J.P. Walker and P.R. Houser. A methodology for initializing soil moisture in a global climate model : Assimilation of near-surface soil moisture. *Journal of Geophysical Research*, 106(D11) :11761–11774, 2001.
- [96] P.J. Wetzel, D. Atlas, and R.H. Woodward. Determining soil moisture from geosynchronous satellite infrared data : A feasibility study. *Climate Applied Meteorology*, 108(23) :375–391, 1984.

- [97] J.-P. Wigneron, M. Pardé, P. Waldteufel, A. Chanzy, Y. Kerr, and S.S. Schmidl. Characterizing the dependence of vegetation model parameters on crop structure, view angle and polarization at L-band. *IEEE Transactions on Geoscience and Remote Sensing*, 42(2) :416–425, 2004.
- [98] J.-P. Wigneron, T. Pellarin, J.-C. Calvet, P. de Rosnay, K. Saleh, and Y. Kerr. *Radiative Transfer Models for Microwave Radiometry (L-MEB : A simple model at L-band for the continental areas. Application to the simulation of a half-degree resolution and global scale data set)*. Stevenage, UK, 2005.

application	taille de l'application	méthodes d'assimilation	observations		
			type	rés. spatiale	rés. temporelle
<b>océanographie</b>	meso-échelle et globale	3D-FGAT et 4D-VAR incrémental	bouées bateaux télédétection	ponctuelle locale, régionale globale	continue semi-continue de quelques jours à quelques semaines
<b>neutronique</b>	locale (taille du réacteur)	3D-VAR avec OI  EKF			
<b>chimie</b>	meso-échelle et globale	3D-FGAT et 4D-VAR incrémental	télédétection	800 m jusqu'à plusieurs km	de quelques jours à quelques semaines
<b>hydrologie</b>	bassin-versant, régionale	OI, 2D-VAR simplifié, techniques d'ensembles	humidité du sol superficielle  débits, hauteurs d'eau vitesses de l'eau télédétection	ponctuelle  ponctuelle globale	toutes les 30 min  horaire, journalière  3 jours - semaines

TAB. 3 – *Tableau-résumé de l'assimilation au CERFACS.*