4D-Var/SEEK : A Consistent Hybrid Variational-Smoothing Data Assimilation Method

Fabrice Veersé, Dinh-Tuan Pham, Jacques Verron

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

4D-Var/SEEK: a consistent hybrid variational-smoothing data assimilation method

Fabrice Veersé, Dinh-Tuan Pham and Jacques Verron

No 3902

March 13, 2000

THÈME 4

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--3902--FR+ENG

4D-Var/SEEK: a consistent hybrid variational-smoothing data assimilation method

Fabrice Veersé*, Dinh-Tuan Pham† and Jacques Verron‡

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet IDOPT

Rapport de recherche n° 3902 — March 13, 2000 — 21 pages

Abstract: A mixed variational-smoothing data assimilation method is derived under the perfect-model hypothesis. An incremental 4D-Var analysis scheme is supplemented with a low-rank approximation of its equivalent Kalman smoother, under the perfect-model assumption. A consistent method results, where the analysis and forecast error covariances provided by the smoother part describe the errors performed respectively during the incremental 4D-Var analysis and the model prediction phases. This is because the whole method is built around a low-rank approximation of the forecast error covariance matrix, and different hypotheses for the computation of the analysis and that of its error covariances are avoided.

The method provides both flow-dependent analysis and forecast error covariances. In addition, most current developments for pre-operational or operational variational data assimilation systems are either already embedded within the method or may be straightforwardly included to it.

Key-words: data assimilation; meteorology; oceanography; Kalman smoother

(Résumé : tsvp)

* Fabrice.Veerse@imag.fr
† Dinh-Tuan.Pham@imag.fr
‡ Jacques.Verron@hmg.inpg.fr

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN (France)
Téléphone : 04 76 61 52 00 - International: +33 4 76 61 52 00
Télécopie : 04 76 61 52 52 - International: +33 4 76 61 52 52

4D-Var/SEEK: une méthode hybride et consistante d'assimilation de données combinant les approches variationnelle et par lissage

Résumé : On présente une méthode mixte d'assimilation combinant les approches variationnelle et par lissage, sous l'hypothèse d'un modèle parfait. Un schéma d'assimilation 4D-Var incrémental est complété par une approximation de rang faible de son lisseur de Kalman équivalent. Il en résulte une méthode consistante, où les covariances d'erreur d'analyse et de prévision fournies par la composante «lisseur» rendent compte des erreurs résultant des phases d'analyse par 4D-Var incrémental et des phases de prévision. Cette consistante provient du fait que des hypothèses différentes sont évitées pour les calculs de l'analyse et de ses covariances d'erreur, la méthode étant basée sur une approximation de rang faible de la matrice de covariances d'erreur de prévision commune aux parties lissage et 4D-Var incrémental.

La méthode fournit des covariances d'erreur d'analyse et de prévision dépendant de l'ecoulement. De plus, beaucoup de développements actuels pour les systèmes d'assimilation variationnelle de données opérationnels ou pré-opérationnels sont soit déjà inclus dans la méthode hybride ou peuvent y être facilement pris en compte.

Mots-clé : assimilation de données; météorologie; océanographie; lisseur de Kalman

Contents

1 Introduction

A key issue in variational data assimilation is the specification of the inverse forecast error covariance matrix appearing in the cost function to be minimized and the computation of the analysis error covariances, to take into account the so-called "errors of the day" (Tanguay *et al.*, 1997). Much work has been directed toward this purpose since the Second WMO Symposium on Assimilation of Observations in Meteorology and Oceanography held in Tokyo in 1995.

At the Data Assimilation Office of the NASA Goddard Space Flight Center some investigations aim at achieving this in the Physical Space Data Assimilation System either by direct specification (Riishøjgaard, 1998) or by adaptive tuning (Dee and da Silva, 1999; Dee *et al.*, 1999). At the UK Meteorological Office two approaches were recently explored for the specification of flow-dependent forecast error covariances in their 3D-Var system: a multi-layer perceptron neural network algorithm and an Error Breeding System (D. Barker's contribution to the Third WMO Symposium on Assimilation of Observations in Meteorology and Oceanography in Quebec City, Canada, 7-11 June 1999). At the Canadian Meteorological Center (CMC) the use of stationary Empirical Orthogonal Functions (EOF) for the specification of forecast error covariances is studied (Buehner *et al.*, 1999). In the future these EOF may be computed from an ensemble system, in order to provide flow-dependent forecast error covariances. At the European Center for Medium-Range Weather Forecast (ECMWF), a reduced-rank simplified Kalman filter has been developed for their incremental 4D-Var assimilation system (Rabier *et al.*, 1997; Fisher, 1999; Fisher and Andersson, 1999). Both the CMC and ECWMF approaches blend a reduced-rank background error covariance matrix with a climatological one describing model balances built from linear regressions using the so-called NCEP method (Parrish and Derber, 1992; Parrish *et al.*, 1997; Bouttier *et al.*, 1997; Gauthier *et al.*, 1998), to create a full-rank matrix.

All the approaches just described specify flow-dependent forecast error covariances by an ad-hoc computation based on simplifying assumptions to make the problem computationally tractable. Consequently the computed error statistics may not reflect the true error of the analyses and forecasts issued from the assimilation system.

In this paper, an assimilation method is derived that combines an incremental 4D-Var analysis with a rank-deficient Kalman smoother in a consistent way, under the perfect-model hypothesis. The reduced-rank smoother is analogous to the SEEK filter of Pham *et al.* (1995; 1998) and is named after it. The potential avantage of the new hybrid method over current implementations of incremental 4D-Var is clearly the use of flow-dependent forecast error covariances. Compared to the SEEK filter it constitutes a smoother extension and benefits from the refreshment of the reference trajectory in the outer loop of the incremental 4D-Var scheme, thereby reducing the impact of linearization error that may potentially lead to divergence of the extended Kalman smoother.

Section 2 presents the incremental formulation considered throughout the paper. The next section derives an equivalent fixed-interval Kalman smoother, for which incremental 4D-Var is a particular algorithm. Then some reduced-rank approximations are made in section 4 to make this equivalent smoother affordable in practice. Section 5 shows how the resulting Singular Evolutive Kalman smoother may be combined with the incremental 4D-Var analysis in a consistent way. The following section addresses some implementation issues. A short discussion concludes the paper. The notations try to follow the recommendations of Ide *et al.* (1997).

2 Incremental 4D-Var

4D-Var consists of finding the model trajectory which best fits the available observations over a given period of time. In the standard formulation of 4D-Var (Le Dimet and Talagrand, 1986) and under the perfect-model assumption, the trajectory is entirely determined by the initial conditions $\mathbf{x}_0$ of the model. The assimilation then reduces to the minimization of the following cost function, where the state of the model at the beginning of the assimilation period is used as control variable:

$$\begin{aligned} \mathcal{J}(\mathbf{x}_0) &= \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}^f(t_0)\right)^T \left[\mathbf{P}^f(t_0)\right]^{-1} \left(\mathbf{x}_0 - \mathbf{x}^f(t_0)\right) \\ &+ \frac{1}{2}\sum_{i=1}^{M} \left[H_i(\mathbf{x}_i) - \mathbf{y}_i^o\right]^T \mathbf{R}_i^{-1} \left[H_i(\mathbf{x}_i) - \mathbf{y}_i^o\right]. \end{aligned} \tag{1}$$

In the above cost function,

- $\mathbf{x}_0$ is the initial state of the model and the control variable;
- $\mathbf{x}^f(t_0)$ is the (background) forecast state at the initial time t_0 of the assimilation period. It is our best guess of the state of the atmosphere or the ocean before the observations available during the assimilation period are taken into account by the assimilation process;
- $\mathbf{P}^f(t_0)$ is the covariance matrix of forecast error describing the quality of the forecast $\mathbf{x}^f(t_0)$;
- $\mathbf{y}_i^o$ is the observation vector at time t_i;
- $\mathbf{x}_i = \mathbf{x}(t_i) = M(t_0, t_i)\mathbf{x}_0$ is the model state at time t_i corresponding to the control initial condition $\mathbf{x}_0$;
- H_i is the observation operator at time t_i. It computes the model-equivalent observation vector;
- $\mathbf{R}_i$ is the covariance matrix of observation error at time t_i, which accounts for measurements and representativeness errors (Lorenc, 1986).

An assumption of no observation at the initial time t_0 has been made. This is often the cas in meteorology since the corresponding observations are better assimilated at the end of the previous assimilation period, in order to improve the quality of ensuing forecast.

The above formulation is often not feasible in practice because of the associated computational burden and because of the non-differentiabilities of the physical parametrizations. The latter prohibite the use of efficient minimization methods requiring the computation of the cost-function gradient. For these reasons, Courtier *et al.* (1994) have proposed an incremental formulation of 4D-Var that is computationally affordable and circumvents the problem of the non-differentiabilities, *e.g.* by using adiabatic (differentiable) tangent linear and adjoint models to compute a correction to the initial condition of the diabatic (non-differentiable) model (Thépaut and Courtier, 1991). Moreover, some meteorological centres are developing differentiable physical parametrizations to be used within this incremental framework in order to improve the quality of their analyses and ensuing forecasts (Mahfouf *et al.*, 1996; Fillion and Mahfouf, 1999; Rabier *et al.*, 1997); Janisková *et al.*, 1999a; Janisková *al.*, 1999b)

In this paper, the incremental 4D-Var formulation is considered without any change of spatial resolution as it is implemented with the LODYC OPA ocean model (Weaver and Vialard, 1999). The extension to the case where two spatial resolutions are used or to the multiple-resolution incremental approach of Veersé and Thépaut (1998) should not be too difficult. Therefore the minimization of (1) is replaced by several minimizations of

$$
\begin{aligned}
\mathcal{J}(\boldsymbol{\delta}\mathbf{x}_0) &= \frac{1}{2}\,\boldsymbol{\delta}\mathbf{x}_0^T[\mathbf{P}^f(t_0)]^{-1}\boldsymbol{\delta}\mathbf{x}_0 \\
&+ \frac{1}{2}\sum_{i=1}^{M}\left[\mathbf{H}'_i(\mathbf{x}_i^g).\boldsymbol{\delta}\mathbf{x}_i - \mathbf{d}_i\right]^T \mathbf{R}_i^{-1}\left[\mathbf{H}'_i(\mathbf{x}_i^g).\boldsymbol{\delta}\mathbf{x}_i - \mathbf{d}_i\right] \\
&= \frac{1}{2}\,\boldsymbol{\delta}\mathbf{x}_0^T[\mathbf{P}^f(t_0)]^{-1}\boldsymbol{\delta}\mathbf{x}_0 \\
&+ \frac{1}{2}\sum_{i=1}^{M}\left[\mathbf{H}'_i(\mathbf{x}_i^g)\mathbf{M}'(t_0,t_i).\boldsymbol{\delta}\mathbf{x}_0 - \mathbf{d}_i\right]^T \mathbf{R}_i^{-1}\left[\mathbf{H}'_i(\mathbf{x}_i^g)\mathbf{M}'(t_0,t_i).\boldsymbol{\delta}\mathbf{x}_0 - \mathbf{d}_i\right]
\end{aligned}
\qquad (2)
$$

where

$$
\begin{aligned}
\mathbf{x}_i^g &= \mathbf{x}^g(t_i) = M(t_0,t_i)\mathbf{x}^g(t_0) \\
\mathbf{x}_i^f &= \mathbf{x}^f(t_i) = M(t_0,t_i)\mathbf{x}^f(t_0) \\
\boldsymbol{\delta}\mathbf{x}_i &= \boldsymbol{\delta}\mathbf{x}(t_i) = \mathbf{M}'(t_0,t_i).\boldsymbol{\delta}\mathbf{x}_0 \\
\mathbf{d}_i &= \mathbf{y}_i^o - H_i(\mathbf{x}_i^g) + \mathbf{H}'_i(\mathbf{x}_i^g).(\mathbf{x}_i^g - \mathbf{x}_i^f)
\end{aligned}
$$

The first-guess field $\mathbf{x}^g(t_0)$ is initially taken equal to the forecast field $\mathbf{x}^f(t_0)$. Between two successive minimizations, it is updated as

$$\mathbf{x}^g(t_0) = \mathbf{x}^f(t_0) + \boldsymbol{\delta}\mathbf{x}_0. \qquad (3)$$

As Courtier (1997) and Zhu *et al.* (1999) we introduce compact notations, using vectors and matrices to account for all the instants but the initial one :

$$\mathbf{x}^f = \begin{bmatrix} \mathbf{x}_1^f \\ \mathbf{x}_2^f \\ \vdots \\ \mathbf{x}_M^f \end{bmatrix}; \qquad \mathbf{x}^g = \begin{bmatrix} \mathbf{x}_1^g \\ \mathbf{x}_2^g \\ \vdots \\ \mathbf{x}_M^g \end{bmatrix}; \qquad \mathbf{M}' = \begin{bmatrix} \mathbf{M}'(t_0, t_1) \\ \mathbf{M}'(t_0, t_2) \\ \vdots \\ \mathbf{M}'(t_0, t_M) \end{bmatrix};$$

$$\boldsymbol{\delta}\mathbf{x} = \begin{bmatrix} \boldsymbol{\delta}\mathbf{x}_1 \\ \boldsymbol{\delta}\mathbf{x}_2 \\ \vdots \\ \boldsymbol{\delta}\mathbf{x}_M \end{bmatrix} = \mathbf{M}'.\boldsymbol{\delta}\mathbf{x}_0; \qquad \mathbf{y}^o = \begin{bmatrix} \mathbf{y}_1^o \\ \mathbf{y}_2^o \\ \vdots \\ \mathbf{y}_M^o \end{bmatrix}; \qquad \mathbf{d} = \begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \vdots \\ \mathbf{d}_M \end{bmatrix};$$

$$H(\mathbf{x}^g) = \begin{bmatrix} H_1(\mathbf{x}_1^g) \\ H_2(\mathbf{x}_2^g) \\ \vdots \\ H_M(\mathbf{x}_M^g) \end{bmatrix}; \qquad \mathbf{R} = \begin{bmatrix} \mathbf{R}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{R}_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{R}_M \end{bmatrix};$$

$$\mathbf{H}'(\mathbf{x}^g) = \begin{bmatrix} \mathbf{H}_1'(\mathbf{x}_1^g) & 0 & \cdots & 0 \\ 0 & \mathbf{H}_2'(\mathbf{x}_2^g) & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{H}_M'(\mathbf{x}_M^g) \end{bmatrix}$$

The incremental 4D-var cost function (2) may then be written:

$$\begin{aligned} \mathcal{J}(\boldsymbol{\delta}\mathbf{x}_0) &= \frac{1}{2}\, \boldsymbol{\delta}\mathbf{x}_o^T [\mathbf{P}^f(t_0)]^{-1} \boldsymbol{\delta}\mathbf{x}_0 \\ &+ \frac{1}{2}\, [\mathbf{H}'(x^g)\mathbf{M}'.\boldsymbol{\delta}\mathbf{x}_0 - \mathbf{d}]^T\, \mathbf{R}^{-1}\, [\mathbf{H}'(x^g)\mathbf{M}'.\boldsymbol{\delta}\mathbf{x}_0 - \mathbf{d}] \end{aligned} \tag{4}$$

Nullifying the gradient, one obtains the analysis increment:

$$\begin{aligned} \boldsymbol{\delta}\mathbf{x}_0 &= \left[[\mathbf{P}^f(t_0)]^{-1} + (\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T\mathbf{R}^{-1}\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\right]^{-1} (\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T\mathbf{R}^{-1}\mathbf{d} \\ &= \mathbf{P}^f(t_0)(\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T \left[\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{P}^f(t_0)(\mathbf{M}')^T\mathbf{H}'(x^g)^T + \mathbf{R}\right]^{-1} \\ &\times \left[\mathbf{y}^o - H(\mathbf{x}^g) + \mathbf{H}'(\mathbf{x}^g).(\mathbf{x}^g - \mathbf{x}^f)\right] \end{aligned} \tag{5}$$

The last equality follows from the Sherman-Morrison-Woddbury matrix formula.[1]

[1] $\mathbf{AB}^T\left(\mathbf{C} + \mathbf{BAB}^T\right)^{-1} = \left(\mathbf{A}^{-1} + \mathbf{B}^T\mathbf{C}^{-1}\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{C}^{-1}$ whenever the inverses exist (Wunsch, 1996, page 99).

3 Equivalent Extended Kalman Smoother

In this section, an (incremental) extended Kalman smoother equivalent to the incremental 4D-Var of the previous section is derived. The approach is somewhat different from those used by Ménard and Daley (1996), Zhu *et al.* (1999) and Li and Navon (1999).

3.1 The forecast phase

Because it will be of use in the next section, the forecast step of the 4D-Var assimilation cycle is written down explicitely. It consists in a prediction from the analysis computed at the beginning of the previous 4D-Var assimilation period, and the corresponding forecast of the error covariances. The forecast field $\mathbf{x}^f(t_0)$ which enters the present 4D-Var assimilation is given by

$$\mathbf{x}^f(t_0) = M(t_0 - T, t_0)\ \mathbf{x}^a(t_0 - T) \tag{6}$$

where T is the temporal length of the previous 4D-Var assimilation and $\mathbf{x}^a(t_0 - T)$ is the corresponding analysis at time $t_0 - T$.

As for the extended Kalman filter with a perfect model, the error covariance matrix of this forecast field is approximately given by

$$\mathbf{P}^f(t_0) = \mathbf{M}'(t_0 - T, t_0)\mathbf{P}^a(t_0 - T)\mathbf{M}'(t_0 - T, t_0)^T. \tag{7}$$

3.2 The analysis phase

The analysis increment is looked for under the form

$$\boldsymbol{\delta}\mathbf{x}_0 = \mathbf{x}_0^a - \mathbf{x}_0^f = \mathbf{K}\left[\mathbf{y}^o - H(\mathbf{x}^g) + \mathbf{H}'(x^g).(\mathbf{x}^g - \mathbf{x}^f)\right] \tag{8}$$

where $\mathbf{x}_0^f$ and $\mathbf{x}_0^a$ are respectively the background and analysis fields. In order to exhibit the equivalence with 4D-Var, the analysis $\mathbf{x}_0^a$ at the beginning of the assimilation period is determined using the *future* observations available along the assimilation period. This presentation is therefore different from the approach used for restrospective data assimilation (Cohn *et al.*, 1994) where Kalman smoothers determine the analysis at a given time using present and past observations.

One assumes that

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

where

$$\mathbf{x}^t = \begin{bmatrix} \mathbf{x}_1^t \\ \mathbf{x}_2^t \\ \vdots \\ \mathbf{x}_M^t \end{bmatrix} \qquad \text{and } \boldsymbol{\varepsilon}^o = \begin{bmatrix} \boldsymbol{\varepsilon}_1^o \\ \boldsymbol{\varepsilon}_2^o \\ \vdots \\ \boldsymbol{\varepsilon}_M^o \end{bmatrix}$$

are respectively the true state and the observation error vectors.

The gain matrix $\mathbf{K}$ is determined by minimizing the total analysis error variance, that is the trace of the analysis error covariance matrix $\mathbf{P}^a(t_0)$ given by

$$\begin{aligned}
\mathbf{P}^a(t_0) &= E\left\{(\mathbf{x}_0^a - \mathbf{x}_0^t)(\mathbf{x}_0^a - \mathbf{x}_0^t)^T\right\} \\
&= E\Big\{\left[\left(\mathbf{x}_0^f - \mathbf{x}_0^t\right) + \mathbf{K}\left(H(\mathbf{x}^t) - H(\mathbf{x}^g) + \mathbf{H}'(\mathbf{x}^g).(\mathbf{x}^g - \mathbf{x}^f) + \boldsymbol{\varepsilon}^o\right)\right] \\
&\quad \times \left[\left(\mathbf{x}_0^f - \mathbf{x}_0^t\right) + \mathbf{K}\left(H(\mathbf{x}^t) - H(\mathbf{x}^g) + \mathbf{H}'(\mathbf{x}^g).(\mathbf{x}^g - \mathbf{x}^f) + \boldsymbol{\varepsilon}^o\right)\right]^T\Big\}
\end{aligned} \tag{9}$$

where E stands for the expectation operator.

The following additional assumptions are made :

- the observations are unbiased: $E\{\boldsymbol{\varepsilon}^o\} = 0$, and their error covariance matrix is $E\left\{\boldsymbol{\varepsilon}^o(\boldsymbol{\varepsilon}^o)^T\right\} = \mathbf{R}$;
- the background initial field is unbiased: $E\{\mathbf{x}_0^f\} = E\{\mathbf{x}_0^t\}$, and its error covariance matrix is $E\left\{\left(\mathbf{x}_0^f - \mathbf{x}_0^t\right)\left(\mathbf{x}_0^f - \mathbf{x}_0^t\right)^T\right\} = \mathbf{P}^f(t_0)$;
- the background and observation errors are uncorrelated: $E\left\{\left(\mathbf{x}_0^f - \mathbf{x}_0^t\right)(\varepsilon_i^o)^T\right\} = 0$ for $i = 1, \cdots, M$;
- the observation operators are quasi-linear:
$$H(\mathbf{x}^t) - H(\mathbf{x}^g) \approx \mathbf{H}'(\mathbf{x}^g).(\mathbf{x}^t - \mathbf{x}^g);$$
- $\mathbf{x}^t - \mathbf{x}^f \approx \mathbf{M}'.(\mathbf{x}_0^t - \mathbf{x}_0^f)$ (tangent-linear hypothesis).

Then, the analysis error covariance matrix is approximately given by

$$\begin{aligned}
\mathbf{P}^a(t_0) &\approx E\Big\{\left[\left(\mathbf{x}_0^f - \mathbf{x}_0^t\right) + \mathbf{K}\left(\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'.\left(\mathbf{x}_0^t - \mathbf{x}_0^f\right) + \boldsymbol{\varepsilon}_0\right)\right] \\
&\quad \times \left[\left(\mathbf{x}_0^f - \mathbf{x}_0^t\right) + \mathbf{K}\left(\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'.\left(\mathbf{x}_0^t - \mathbf{x}_0^f\right) + \boldsymbol{\varepsilon}_0\right)\right]^T\Big\} \\
&= \mathbf{K}\left[\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{P}^f(t_0)(\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T + \mathbf{R}\right]\mathbf{K}^T \\
&- \mathbf{K}\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{P}^f(t_0) - \mathbf{P}^f(t_0)(\mathbf{M}')^T\mathbf{H}'(x^g)^T\mathbf{K}^T + \mathbf{P}^f(t_0) \\
&= \left(\mathbf{K} - \mathbf{P}^f(t_0)(\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T\mathbf{G}^{-1}\right)\mathbf{G}\left(\mathbf{K} - \mathbf{P}^f(t_0)(\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T\mathbf{G}^{-1}\right)^T \\
&- \mathbf{P}^f(t_0)(\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T\mathbf{G}^{-1}\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{P}^f(t_0) + \mathbf{P}^f(t_0)
\end{aligned} \tag{10}$$

where $\mathbf{G} = \mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{P}^f(t_0)(\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T + \mathbf{R}$ and a matrix formula[2] has been used.

The only term depending on $\mathbf{K}$ in the last expression is positive. Therefore the gain matrix that minimizes the total error variance, *i.e.* the trace of $\mathbf{P}^a(t_0)$, is given by:

$$\begin{aligned} \mathbf{K} &= \mathbf{P}^f(t_0)(\mathbf{M}')^T\mathbf{H}'(x^g)^T\mathbf{G}^{-1} \\ &= \mathbf{P}^f(t_0)(\mathbf{M}')^T\mathbf{H}'(x^g)^T\left[\mathbf{H}'(x^g)\mathbf{M}'\mathbf{P}^f(t_0)(\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T + \mathbf{R}\right]^{-1} \end{aligned} \tag{11}$$

Substituting this expression of the gain matrix into Eq. (8), the 4D-Var analysis increment given in (5) is recovered, which shows the equivalence of the two approaches.

From (10) one has :

$$\begin{aligned} \mathbf{P}^a(t_0) &= \mathbf{P}^f(t_0) - \mathbf{P}^f(t_0)(\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T\mathbf{G}^{-1}\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{P}^f(t_0) \\ &= \left(\mathbf{I} - \mathbf{K}\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\right)\mathbf{P}^f(t_0) \\ &= \left[[\mathbf{P}^f(t_0)]^{-1} + (\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T\mathbf{R}^{-1}\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\right]^{-1} \end{aligned} \tag{12}$$

The second equality makes use of the expression of gain matrix $\mathbf{K}$ and the last one derives from the matrix-inversion lemma.[3]

Using the second equality together with the expression (11) of $\mathbf{K}$, it is easy to show that the gain matrix is also given by:

$$\mathbf{K} = \mathbf{P}^a(t_0)(\mathbf{M}')^T\mathbf{H}'(x^g)^T\mathbf{R}^{-1}. \tag{13}$$

4 A Singular Evolutive Extended Kalman (SEEK) smoother

In this section, following the lines of Pham *et al.* (1998), an approximation of the equivalent Kalman smoother of the previous section is derived, that uses low-rank forecast and analysis error covariance matrices. Such an approximation was also used by Cohn and Todling (1996) and Todling *et al.* (1998). The resulting scheme is named SEEK smoother after the denomination SEEK already used for the filter of Pham *et al.* (1998): Singular because the error covariance matrices are singular; Evolutive because the error covariances are propagating in time; Extended because the algorithm is extended to nonlinear systems; and Kalman smoother because the scheme is based on the fixed-interval Kalman smoother.

[2] $\mathbf{ABA}^T - \mathbf{CA}^T - \mathbf{AC}^T = \left(\mathbf{A} - \mathbf{CB}^{-1}\right)\mathbf{B}\left(\mathbf{A} - \mathbf{CB}^{-1}\right)^T - \mathbf{CB}^{-1}\mathbf{C}^T$ where the inverses are assumed to exist (Wunsh, 1996, page 100).

[3] $\left(\mathbf{A}^T\mathbf{B}^{-1}\mathbf{A} + \mathbf{C}^{-1}\right)^{-1} = \mathbf{C} - \mathbf{CA}^T\left(\mathbf{ACA}^T + \mathbf{B}\right)^{-1}\mathbf{AC}$ where it is assumed the inverses exist (Wunsh, 1996, page 99).

Substituting the expression (7) of $\mathbf{P}^f(t_0)$ into Eq. (12), one obtains:

$$\begin{aligned}\mathbf{P}^a(t_0) = \mathbf{M}'_- \big(\mathbf{P}^a_- \quad &- \quad \mathbf{P}^a(\mathbf{M}'_-)^T(\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T \\ &\times \quad \left[\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{M}'_-\mathbf{P}^a_-(\mathbf{M}'_-)^T(\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T + \mathbf{R}\right]^{-1} \\ &\times \quad \mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{M}'_-\mathbf{P}^a_-\big)\,(\mathbf{M}'_-)^T\end{aligned} \tag{14}$$

where $\mathbf{M}'_-$ and $\mathbf{P}^a_-$ stand for $\mathbf{M}'(t_0 - T, t_0)$ and $\mathbf{P}^a(t_0 - T)$ respectively.

Assume now that $\mathbf{P}^a_-$ is a low-rank matrix factorized into $\mathbf{L}_-\mathbf{U}_-\mathbf{L}_-^T$ where $\mathbf{U}_-$ is a symmetric positive define matrix with dimension equal to the rank of $\mathbf{P}^a_-$. Then:

$$\mathbf{P}^a(t_0) = \mathbf{L}_0\mathbf{U}_0\mathbf{L}_0^T \tag{15}$$

where

$$\begin{aligned}\mathbf{L}_0 &= \mathbf{M}'_-\mathbf{L}_- \\ \mathbf{U}_0 &= \mathbf{U}_- - \mathbf{U}_-\mathbf{L}_0^T(\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T\left[\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{L}_0\mathbf{U}_-\mathbf{L}_0^T(\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T + \mathbf{R}\right]^{-1} \\ &\qquad \times \mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{L}_0\mathbf{U}_-\end{aligned} \tag{16}$$

Assuming there is no perfect observation, $\mathbf{R}$ is invertible and, using the matrix-inversion lemma, the latter equation may be written

$$\mathbf{U}_0^{-1} = \mathbf{U}_-^{-1} + \mathbf{L}_0^T(\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T\mathbf{R}^{-1}\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{L}_0 \tag{17}$$

showing that $\mathbf{U}_0$ is also symmetric definite positive (hence invertible) and that $\mathbf{P}^a$ and $\mathbf{P}^a_-$ have same rank provided the rank of $\mathbf{L}_0$ equals that of $\mathbf{L}_-$.

5 Combining incremental 4D-Var with the SEEK smoother

In this section, the computation of the analysis increment by 4D-Var is combined with the computation of the analysis and forecast error covariance matrices of the SEEK smoother in a consistent way. The resulting incremental 4D-Var assimilations use dynamical forecast error covariances (the so-called "errors of the day"), and the computations of the analysis increment and of the analysis error covariance matrix are based on the same low-rank approximation.

Evolving the low-rank analysis error covariance matrix $\mathbf{P}^a_- = \mathbf{L}_-\mathbf{U}_-\mathbf{L}_-^T$ according to Eq. (7), one obtains the forecast error covariance matrix:

$$\mathbf{P}^f(t_0) = \mathbf{M}'_-\mathbf{L}_-\mathbf{U}_-\mathbf{L}_-^T(\mathbf{M}'_-)^T = \mathbf{L}_0\mathbf{U}_-\mathbf{L}_0^T \tag{18}$$

The matrix $\mathbf{U}_-$ has dimension equal to the rank of $\mathbf{P}^a_-$, in practice small enough for one to afford a Cholesky factorization:

$$\mathbf{U}_- = \mathbf{W}_- \mathbf{W}_-^T. \tag{19}$$

The usual preconditioning of the minimization may then be performed by a change of control variable, rewriting the incremental cost-function (4) as

$$\mathcal{J}(\boldsymbol{\chi}) = \frac{1}{2}\, \boldsymbol{\chi}^T \boldsymbol{\chi} + \frac{1}{2}\, \left[\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'.\boldsymbol{\delta}\mathbf{x}_0 - \mathbf{d}\right]^T \mathbf{R}^{-1} \left[\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'.\boldsymbol{\delta}\mathbf{x}_0 - \mathbf{d}\right] \tag{20}$$

where

$$\begin{aligned} \mathbf{d} &= \mathbf{y}^o - H(\mathbf{x}^g) + \mathbf{H}'(\mathbf{x}^g).(\mathbf{x}^g - \mathbf{x}^f) \\ \boldsymbol{\delta}\mathbf{x}_0 &= (\mathbf{L}_0 \mathbf{W}_-)\, \boldsymbol{\chi} \end{aligned} \tag{21}$$

This cost function is exactely that used by Blayo *et al.* (1999) for reducing the dimension of the control space, if $\mathbf{L}_0$ is taken to be the leading model EOFs and $\mathbf{W}_-$ is their Cholesky-like factor of the background error covariance matrix.

The analysis error covariance matrix is obtained as $\mathbf{P}^a(t_0) = \mathbf{L}_0\mathbf{U}_0\mathbf{L}_0^T$ where

$$\begin{aligned} \mathbf{U}_0^{-1} &= \mathbf{U}_-^{-1} + \mathbf{L}_0^T(\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T\mathbf{R}^{-1}\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{L}_0 \\ &= \mathbf{U}_-^{-1} + \sum_{i=1}^{M} \mathbf{L}_0^T\mathbf{M}'(t_0,t_i)^T\mathbf{H}_i'(\mathbf{x}_i^g)^T\mathbf{R}_i^{-1}\mathbf{H}_i'(\mathbf{x}_i^g)\mathbf{M}'(t_0,t_i)\mathbf{L}_0 \end{aligned} \tag{22}$$

Then a Cholesky factorization is performed :

$$\mathbf{U}_0^{-1} = \mathbf{W}_0^{-T}\mathbf{W}_0^{-1} \tag{23}$$

and the forecast error covariance matrix to be used in the next incremental 4D-Var assimilation is computed by :

$$\begin{aligned} \mathbf{L}_+ &= \mathbf{M}'(t_0,t_f)\mathbf{L}_0 \\ \mathbf{P}^f(t_f) &= \mathbf{L}_+\mathbf{U}_0\mathbf{L}_+^T \\ &= \mathbf{L}_+\mathbf{W}_0\mathbf{W}_0^T\mathbf{L}_+^T \end{aligned} \tag{24}$$

where t_f corresponds to the end of the current assimilation period and the beginning of the next one.

The whole algorithm is a combined incremental 4D-Var/singular evolutive Kalman smoother with:

- **an analysis step:** an incremental 4D-Var analysis where several cost functions (20) are successively minimized. Between two consecutive minimizations the first-guess field is updated:

$$\mathbf{x}^g(t_0) = \mathbf{x}^f(t_0) + \boldsymbol{\delta}\mathbf{x}_0 = \mathbf{x}^f(t_0) + (\mathbf{L}_0\mathbf{W}_-)\ \boldsymbol{\chi}.$$

After the last minimization, the low-rank analysis error covariance matrix is determined by computing $\mathbf{U}_0^{-1}$ according to (22). Its Cholesky factorization (23) is then computed. Alternatively, when the matrices $\mathbf{R}_i$ are diagonal, the Cholesky factors of $\mathbf{U}_-^{-1}$ may be updated directly using techniques described in Gill *et al.* (1974) or Davis and Hager (1999).

- **a forecast step:** the forecast field at time t_f is computed as

$$\mathbf{x}^f(t_f) = M(t_0, t_f)\mathbf{x}^a(t_0)$$

and the corresponding forecast error covariance matrix is determined according to (24), their flow-dependency resulting from the tangent-linear model integrations (first equation in (24).

6 Implementation details

In this section some aspects having a strong impact on the practical implemention of the 4D-Var/SEEK method are described.

6.1 Model trajectory computations

The preconditioned incremental 4D-Var cost function (20) may be written

$$\begin{aligned}\mathcal{J}(\boldsymbol{\chi}) &= \tfrac{1}{2}\,\boldsymbol{\chi}^T\boldsymbol{\chi} \\ &+ \tfrac{1}{2}\,[\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{L}_0\mathbf{W}_-\ \boldsymbol{\chi} - \mathbf{d}]^T\,\mathbf{R}^{-1}\,[\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{L}_0\mathbf{W}_-\ \boldsymbol{\chi} - \mathbf{d}]\end{aligned} \tag{25}$$

Clearly, the quantity $\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{L}_0$ appears both in the cost function and in the computation of the analysis error covariance matrix (see Eq. (22)). In order to minimize the computational work it can be computed once for all before each minimization, at the cost of a number of tangent-linear model integrations equal to the rank of the forecast error covariance matrix (*i.e.* the rank of $\mathbf{U}_-$). The cost of each minimization itself then becomes negligible.

Moreover, it is possible to implement the computation of the cost function and its gradient and that of the analysis error covariance in such a way as to need only one column of $\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\mathbf{L}_0$ at a time in core memory. This ensures both the optimality in terms of computational cost and the feasibility with operational numerical weather prediction or oceanic circulation models.

6.2 Basis orthonormalization

To initialize the 4D-Var/SEEK assimilation system, the matrix $\mathbf{P}^a_- = \mathbf{L}_-\mathbf{U}_-\mathbf{L}_-^T$ must be specified at the very first time. Following the experience of Pham *et al.* (1998) with the SEEK filter applied to oceanic circulation models, a sensible choice is to use a leading model EOF eigen-decomposition. An alternate approach would be to build $\mathbf{P}^a_-$ from a principal component analysis of the sample covariance matrix of the so-called NMC method, used in most meteorological centers for the specification of climatologic forecast error covariances (Buehner *et al.*, 1999).

Even if one initializes the system with one of the two methods suggested, for which the columns of $\mathbf{L}$ are orthonormal, the forecast equation in (24) indicates that they will not remain so in general and that $\mathbf{L}$ may well become rank-deficient in the future. Pham *et al.* (1998) indicate that in the SEEK filter, and similarly in Eq. (22) of the 4D-Var/SEEK smoother, one can change $\mathbf{L}_0$ to $\mathbf{L}_0\mathbf{W}_0$ and $\mathbf{U}_0^{-1}$ to the identity matrix without changing the algorithm. This may be necessary to prevent the matrix $\mathbf{U}$ from becoming ill-conditioned.

But for 4D-Var/SEEK one may more conveniently transform $\mathbf{L}_0$ into $\tilde{\mathbf{L}}_0 = \mathbf{L_0}\mathbf{W}_-$ and $\mathbf{U}_0^{-1}$ into $\tilde{\mathbf{U}}_0^{-1} = \mathbf{W}_-^T\mathbf{U}_0^{-1}\mathbf{W}_-$; Eq. (22) is then rewritten

$$\tilde{\mathbf{U}}_0^{-1} = \mathbf{I} + \tilde{\mathbf{L}}_0^T(\mathbf{M}')^T\mathbf{H}'(\mathbf{x}^g)^T\mathbf{R}^{-1}\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'\tilde{\mathbf{L}}_0, \tag{26}$$

and the definition of the increment in Eq. (21) is now

$$\boldsymbol{\delta}\mathbf{x}_0 = \tilde{\mathbf{L}}_0\ \boldsymbol{\chi}. \tag{27}$$

As noted above $\tilde{\mathbf{L}}_0$ may become rank deficient, essentially because of the tangent-linear model integrations in Eq. (24). Indeed the collection of the columns of $\tilde{\mathbf{L}}_0$ may be viewed as a basis for an approximate local tangent hyperplane to the attractor of the model. If the restriction of the tangent-linear model to this local hyperplane is rank deficient, the information projecting on one or more directions of this hyperplane will be lost. This situation may be easily diagnosed. For that purpose, it suffices to note that Eq. (27) specifies the increment $\boldsymbol{\delta}\mathbf{x}_0$ as a linear combination of the directions contained in $\tilde{\mathbf{L}}_0$ and can be rewritten

$$\boldsymbol{\delta}\mathbf{x}_0 = \overline{\mathbf{L}}_0\ \overline{\boldsymbol{\chi}} \tag{28}$$

where the columns of $\overline{\mathbf{L}}_0$ are obtained by orthonormalizing those of $\tilde{\mathbf{L}}_0$. During the course of such an orthonormalization a rank deficiency can be easily diagnosed. In such a case, the initial rank may be recovered by enlarging the orthonormal basis $\overline{\mathbf{L}_0}$ with the first leading model EOF that do not project essentially on this basis or some directions deduced from the innovation vector (Brasseur *et al.*, 1999). These new directions are orthonormalized with respect to the kept basis vectors to form a new basis $\overline{\mathbf{L}_0}$. The minimization is then performed in terms of $\overline{\boldsymbol{\chi}}$ using the incremental cost function

$$\begin{aligned}\mathcal{J}(\overline{\boldsymbol{\chi}}) &= \tfrac{1}{2}\,\overline{\boldsymbol{\chi}}^T\left(\overline{\mathbf{L}}_0^T\tilde{\mathbf{L}}_0\tilde{\mathbf{L}}_0^T\overline{\mathbf{L}}_0\right)\overline{\boldsymbol{\chi}} \\ &+ \tfrac{1}{2}\,\left[\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'.\boldsymbol{\delta}\mathbf{x}_0 - \mathbf{d}\right]^T\mathbf{R}^{-1}\left[\mathbf{H}'(\mathbf{x}^g)\mathbf{M}'.\boldsymbol{\delta}\mathbf{x}_0 - \mathbf{d}\right]\end{aligned} \tag{29}$$

where $\boldsymbol{\delta}\mathbf{x}_0$ is given by Eq. (28). As the first term of the sum is no more ideally conditioned, this should be done only when an enlargement of the basis is needed. Otherwise, the original cost function (20) should be minimized using the definition (21) or (27) of the increment $\boldsymbol{\delta}\mathbf{x}_0$.

Finally two variants of 4D-Var/SEEK can be proposed. The first one is an asymptotic smoother where the forecast step in Eq. (24) is replaced by $\mathbf{L}_+ = \mathbf{L}_0$ and the covariance matrix of forecast error is kept constant and equal to $\mathbf{P}^f(t_f) = \mathbf{P}^f(t_0) = \mathbf{L}_0\mathbf{U}_-\mathbf{L}_0^T$. The analysis error covariance matrix is still computed according to Eq. (22), but for diagnostic purpose only. The resulting $\mathbf{U}$ matrix should not be used to update the forecast error covariance matrix in the asymptotic smoother, otherwise the forecast field may be given more and more weight and observations may be rejected by the assimilation system. This clearly happens in the experiments of Ballabrera *et al.* (1999).
The other variant is a nonlinear extension which uses the direct (nonlinear) model to forecast the basis vectors (*i.e.* the colums of $\mathbf{L}$) according to

$$(\mathbf{L}_+)_i = M(t_0, t_f)\left(\mathbf{x}^f(t_0) + \alpha_i\ (\mathbf{L}_0)_i\right) - M(t_0, t_f)\mathbf{x}^f(t_0)$$

where the α_i are suitably chosen scaling factors.

6.3 Temporal weighting

Although our incremental 4D-Var/SEEK method assumes the model is perfect, one knows that in reality this is not the case. It is therefore tempting to introduce a temporal weighting to emphasize the impact of the most recent observations, as proposed by Courtier and Talagrand (1990). A new incremental cost function is then defined where the observation error covariance matrices $\mathbf{R}_i$ are now multiplied by a scalar ρ_i.

Following Pham *et al.* (1995) this scalar is called "forgetting factor" and may be taken as

$$\rho_i = \rho^{\frac{t_f - t_0}{t_i - t_0}} \tag{30}$$

where $\rho = 1/(1+\alpha)$, α being a positive constant. The scalar α may be adaptively tuned using techniques described in Dee (1991, 1995), Gong *et al.* (1998) and Evensen *et al.* (1998).

Alternate formulations of the forgetting factor, such a Gaussian ponderation in time centered at the end of the assimilation period (*i.e.* at time t_f), may be used.

This ad-hoc method to account for model error has moreover the advantage of preserving the rank of the 4D-Var/SEEK smoother (Pham *et al.*, 1995 and 1998).

7 Discussion

An assimilation method has been proposed that combines an incremental 4D-Var analysis computation with a Singular Evolutive Extended Kalman (SEEK) smoother, in a consistent way. The computation of the analysis by the incremental 4D-Var method and the computation of the forecast and analysis error covariances in the SEEK smoother part of the method are based on a common low-rank approximation of the forecast error covariance matrix. The computed error covariances thus reflect the performance of the assimilation system in the subspace where the analysis computation is done.

Interestingly, many of the current developments of pre-operational and operational variational data assimilation systems in meteorology can be embedded within this method. A natural equivalence between the reduction of the dimension of the control space proposed by Blayo *et al.* (1998) and the specification of a low-rank approximate forecast error covariance matrix based on an EOF decomposition (Buehner *et al.*, 1999) has emerged during the derivation of the 4D-Var/SEEK method, that comprises both approaches. The quasi-continuous technique (Luong, 1995; Järvinen *et al.*, 1996; Pires *al.*, 1996; Luong *et al.*, 1998) may be incorporated to the method without any change needed. The 4D-Var/SEEK method could also be adapted to accomodate 4D-Var formulations involving a change of spatial resolution, such as the incremental strategy of Courtier *et al.* (1994) and the multiple-truncation incremental approach of Veersé and Thépaut (1998). A more interesting incremental approach using multiple control spaces of increasing dimensions, corresponding to increasing atmospheric or oceanic variability, may be straightforwardly achieved with 4D-Var/SEEK.

Finally, because it comprises a simplified Kalman smoother and a decomposition on a low-rank basis of the forecast error covariances, the method can be thought of as a consistent blend of the CMC and ECMWF approaches for the specification of flow-dependent forecast error covariances, with the additional advantage of providing also analysis error covariances.

A current major limitation of this hybrid method comes from the perfect-model assumption, although the temporal weighting introduced in the last subsection adresses this problem to some extent; the treatment of model error within 4D-Var/SEEK requires further study.

The approach used to derive the scheme is quite general :

- derive the equivalent smoother of a generic incremental 4D-Var scheme;
- make some simplifications based on a particular approximation of the error covariances;
- and finally write the specific incremental 4D-Var scheme based on this approximation.

Proceeding this way, the smoother part is always full-rank with respect to the dimension of the space where the analysis increment is computed by the incremental 4D-Var part of the scheme. And a number of hybrid methods may be designed this way. For instance, for strongly nonlinear dynamics one could think of using a 4D-Var/SEIK smoother where a reduced-rank ensemble smoother analogous to the SEIK filter of Pham (1996) is used to compute the error covariances.

Acknowledgments

The work of F. Veersé is supported by a post-doctoral research grant from the Institut National de la Recherche en Informatique et en Automatique (INRIA), Montbonnot Saint-Martin, France. IDOPT is a joint project of INRIA, Centre National de la Recherche

Scientifique (CNRS), Université Joseph Fourier, and Institut National Polytechnique de Grenoble.

References

Ballabreara, J., Brasseur, P. and Verron, J., 1999: Reduced order evolutive Kalman filter for altimeter data assimilation into an ocean primitive equation model, submitted to *Quart. J. Roy. Meteor. Soc.*

Blayo, E., Blum, J. and Verron, J., 1998: Assimilation variationnelle de données en océanographie et réduction de la dimension de l'espace de contrôle. (Translation: Variational data assimilation in oceanography and reduction of the control-space dimension.) *Equations aux Dérivées Partielles et Applications*, Gauthiers-Villars, 199-219.

Blayo, E., Blum, J., Durbiano, S. and Verron, J., 1999: Reduction of the dimension of the control space for variational data assimilation. Proceedings, *Third WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography*, 7-11 June 1999, Quebec City, Canada, World Meteorological Organization.

Bouttier, F., Derber, J. and Fisher, M., 1997: The 1997 revision of the Jb term in 3D/4D-Var. *ECMWF Tech. Memo.* **238**, 54 pp., ECMWF, Reading, UK.

Brasseur, P., Ballabrera-Poy, J. and Verron, J., 1999: Assimilation of altimetric observations in a primitive equation model of the Gulf Strean using a Singular Evolutive Extended Kalman filter, *J. Mar. Syst.*, **22,** 269-294.

Buehner, M., Brunet, G. and Gauthier, P., 1999: Empirical Orthogonal Functions for modeling 3D-Var forecast error statistics. Proceedings, *Third WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography*, 7-11 June 1999, Quebec City, Canada, World Meteorological Organization.

Cohn, S.E., Sivakumaran, N.S. and Todling, R., 1994: A fixed-lag Kalman smoother for restrospective data assimilation. *Mon. Wea. Rev.*, **122,** 2838-2867.

Cohn, S.E. and Todling, R., 1996: Approximate data assimilation schemes for stable and unstable dynamics. *J. Meteor. Soc. Japan*, **74-1,** 63-75.

Courtier, P. and Talagrand, O., 1990: Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. *Tellus*, **42A,** 531-549.

Courtier, P., Thépaut, J.-N. and Hollingsworth, A., 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120,** 1367-1387.

Davis, T.A. and Hager, W.W., 1999: Modifying a sparse Cholesky factorization. *SIAM J. Matrix Anal. Appl.*, **20,** 606-627.

Dee, D.P., 1991: Simplification of the Kalman filter for meteorological data assimilation. *Quart. J. Roy. Meteor. Soc.*, **117,** 365-384.

Dee, D.P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123,** 1128-1145.

Dee, D.P. and da Silva, A.M., 1999: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Mon. Wea. Rev.*, **124,** 1822-1834.

Dee, D.P., Gaspari, G., Redder, C., Rukhovets, L. and da Silva, A.M., 1999: Maximum-likelihood estimation of forecast and observation error covariance parameters. Part II: Applications. *Mon. Wea. Rev.*, **124,** 1835-1849.

Evensen, G., Dee, D. and Schroter, J., 1998: Parameter estimation in dynamical models. *NATO ASI, Ocean Modeling and Parametrizations*, E.P. Chassignet and J. Verron (eds), Kluwer Academic Publishers, 373-398.

Fillion, L. and Mahfouf, J.-F., 1999: Coupling of moist convective and stratiform precipitation processes for variational data assimilation. To appear in *Mon. Wea. Rev.*

Fisher, M., 1999: Evaluation of a reduced-rank Kalman filter based on Hessian singular vectors. Proceedings, *Third WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography*, 7-11 June 1999, Quebec City, Canada, World Meteorological Organization.

Fisher, M. and Andersson, E. 1999: Calculation of flow-dependent background errors using a randomization Kalman filter technique. Proceedings, *Third WMO International Symposium on Assimilation of Observations in Meteorology and Oceanography*, 7-11 June 1999, Quebec City, Canada, World Meteorological Organization.

Gauthier, P., Buehner, M. and Fillion, L., 1998: Background-error statistics modeling in a 3D variational data assimilation scheme: estimation and impact on the analyses. Proceedings, *ECMWF Workshop on diagnosis of data assimilation systems*, 2-4 November 1998, Reading, UK.

Gill, P.E., Golub, G.H., Murray, W. and Saunders, M.A., 1974: Methods for modifying matrix factorizations. *Math. Comput.*, **28,** 505-535.

Gong, J., Wahba, G., Johnson, D.R. and Tribbia, J., 1998: Adaptive tuning of numerical weather prediction models: simultaneous estimation of weighting, smoothing and physical parameters. *Mon. Wea. Rev.*, **125,** 210-231.

Ide, K., Courtier, P., Ghil, M. and Lorenc ,A.C., 1997: Unified notation for data assimilation: operational, sequential and variational. *J. Meteor. Soc. Japan*, **75-1B,** 181-189.

Janisková, M., Thépaut, J.-N. and Geleyn, J.-F., 1999a: Simplified and regular physical parametrizations for incremental four-dimensional variational data assimilation. *Mon. Wea. rev.*, **127,** 26-45.

Janisková, M., Veersé, F., Thépaut, J.-N., Desroziers, G. and Pouponneau, B., 1999b: Impact of a simplified physical package in 4-D Var analyses of FASTEX situations. *Quart. J. Roy. Meteor. Soc.*, **125,** 2465-2486.

Järvinen, H., Thépaut, J.-N. and Courtier, P., 1996: Quasi-Continuous variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **122,** 515-534.

Li, Z. and Navon, M., 1999: Optimality of 4D-Var and its relationship with the Kalman filter and Kalman smoother. Submitted to *Quart. J. Roy. Meteor. Soc.*

Lorenc, A.C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112,** 1177-1194.

Luong, B., 1995: *Techniques de contrôle optimal pour un modèle quasi-géostrophique de circulation océanique. Application à l'assimilation variationnelle des données altimétriques satellitaires. (Translation: Optimal control techniques for a quasi-geostrophic oceanic circulation model. Application to the variational assimilation of satellite altimetric data.)* Thèse de doctorat de l'Université Joseph Fourier, Grenoble, France. [Available from ftp://ftp.imag.fr/pub/IDOPT/THESES/these_luong.ps.gz]

Luong, B., Blum, J. and Verron, J., 1998: A variational method for the resolution of a data assimilation problem in oceanography. *Inverse Problems*, **14,** 979-997.

Mahfouf, J.-F., Buizza, R. and Errico, R., 1996: Strategy for including physical processes in the ECMWF variational data assimilation system. Proceedings, *ECMWF workshop on non-linear aspects of data assimilation*, Shinfield Park, Reading, RG2 9AX, 9-11 September 1996.

Ménard, R. and Daley, R., 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus*, **48A,** 221-237.

Parrish, D.F. and Derber, J.C., 1992: The National Meteorological Centre's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120,** 1747-1763.

Parrish, D.F., Derber, J.C., Purser, R.J., Wu, W.-S. and Pu, Z.-X., 1997: The NCEP Global Analysis System: Recent Improvements and Future Plans. *J. Meteor. Soc. Japan*, **75-1B,** 359-365.

Pham, D.T., 1996: A Singular Evolutive Interpolated Kalman filter for data assimilation in oceanography, *Tech. Report, IDOPT project INRIA-CNRS-UJF-INPG*, RT **163**, 22 pp.

Pham, D.T., Verron, J. and Roubaud, M.-C., 1995: Assimilation of oceanic altimeter data using a quasi-geostrophic model and the extended Kalman filter. *Tech. Report, IDOPT project INRIA-CNRS-UJF-INPG*, RT **141**, 20 pp.

Pham, D.T., Verron, J. and Roubaud, M.-C., 1998: A singular evolutive extended Kalman filter for data assimilation in oceanography. *J. Mar. Syst.*, **16,** 323-340.

Pires, C., Vautard, R. and Talagrand, O., 1996: On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus*, **48A,** 96-121.

Rabier, F., Mahfouf, J.-F., Fisher, M., Järvinen, H., Simmons, A., Andersson, E., Bouttier, F., Courtier, P., Hamrud, M., Haseler, J., Hollingsworth, A., Isaksen, L., Klinker, E., Saarinen, S., Temperton, C., Thépaut, J.-N., Undén, P. and Vasiljević, 1997: Recent experimentation on 4D-Var and first results from a simplified Kalman filter. *ECMWF Tech. Memo.* **240**, ECMWF, Reading, UK, 42 pp.

Riishøjgaard, L.P., 1998: A direct way of specifying flow-dependent background error correlations for meteorological analysis systems. *Tellus*, **50A,** 42-57.

Tanguay, M., Anderson, D.L.T, Bennett, A.F., Busalacchi, A.J., Cohn, S.E., Courtier, P., Derber, J., Lorenc, A.C., Parrish, D., Purser, J., Sato, N. and Schlatter, T., 1997: Data Assimilation in the Ocean and in the Atmosphere: What Should be Next? *J. Meteor. Soc. Japan*, **75-1B,** 489-496.

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

Todling, R., Cohn, S.E. and Sivakumaran, N.S., 1998: Suboptimal schemes for restrospective data assimilation based on the fixed-lag Kalman smoother. *Mon. Wea. Rev.*, **126,** 2274-2286.

Veersé, F. and Thépaut, J.-N., 1998: Multiple-truncation incremental approach for four-dimensional variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **124,** 1889-1908.

Weaver A. and Vialard, J., 1999: Development of an ocean incremental 4D-Var scheme for seasonal prediction. Proceedings, *Third WMO international Symposium on Assimilation of Observations in Meteorology and Oceanography*, 7-11 June 1999, Quebec City, Canada, World Meteorological Organization.

Wunsch, C., 1996: *The Ocean Circulation Inverse Problem.* Cambridge University Press, 442 pp.

Zhu, Y., Todling, R. and Cohn, S.E., 1999: Technical remarks on smoother algorithms. *DAO Office Note* **99-02**, Data Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland. [available from http://dao.gsfc.nasa.gov/subpages/office-notes.html]

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399