



# Optimal Determination Of Nudging Coefficients

Arthur Vidard, François-Xavier Le Dimet, Andrea Piacentini

► **To cite this version:**

Arthur Vidard, François-Xavier Le Dimet, Andrea Piacentini. Optimal Determination Of Nudging Coefficients. *Tellus A, Co-Action Publishing*, 2003, 55 (1), pp.1–15. <inria-00325360>

**HAL Id: inria-00325360**

**<https://hal.inria.fr/inria-00325360>**

Submitted on 29 Sep 2008

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Optimal Determination Of Nudging

## Coefficients

P. A. VIDARD<sup>(1)</sup>\*

F-X. LE DIMET<sup>(2)</sup>

A. PIACENTINI<sup>(1)</sup>

September 29, 2007

(1) *Centre Européen de Recherche et de  
Formation Avancée en Calcul Scientifique  
42, avenue Coriolis  
31057 Toulouse cedex, France*

(2) *Laboratoire de Modélisation et Calcul.  
Université Joseph Fourier  
BP 53 38041 Grenoble, France*

\* corresponding author email: vidard@cerfacs.fr

## Abstract

A low-cost method for dealing with model errors in 4D variational assimilation using the innovation residual is presented and formally compared to the Kalman filter. This method is applied to two models, first a burger equation to calibrate the method and second to a more relevant Shallow water equations model. It is shown that compared to the classical 4D-Var results, this new method provides a sizeable improvement.

**Keywords :** Data assimilation, innovation vector, Adjoint methods, Nudging

## 1 Introduction

Data Assimilation is a wide class of numerical methods for estimating the state of a system by combining information from observational data with information provided by a numerical model. One of the most important applications of these methods is the estimation of the best state of the atmosphere (or the ocean) at a given time in order to improve the accuracy of the numerical forecast. In recent years, developments of both observational means (remote sensing, buoys, tomography, ...) and computing resources have permitted to improve widely Data Assimilation methods and their efficiency.

The current operational method in meteorological centres is called 4D-Var (for four dimension variational data assimilation) . It consists in assimilating all the observational and model informations from the previous time sequence. The problem is formulated as an optimal control problem where the criteria measures the misfit between the model predictions and the observations of the system state. One of the main assumption of 4D-Var is that the model describes exactly the system behavior. However, in practice the model equations do not represent the exact evolution of the system and model errors arise because of the lack of resolution, inaccurate representation of small scale physics or errors in boundary conditions, topography or forcing terms.

The next step in the 4D-Var development will be to consider that the model is not exact i.e, for example, to introduce the model errors correcting term in the control vector (see Jazwinsky, 1970; Derber, 1989 and Cohn, 1997). The principle of the complete method is to add in the control vector, a residual error correcting term which is added to the model equation at every time step. Due to the size of this new problem (the dimension of the state variable (typically  $10^6$ - $10^7$ ) times the number of time steps), this approach is unaffordable for current computational resources. In order to reduce the cost of model errors

control, one can propose to control only several and well chosen direction of the error (Vidard et al, 2000), or only the time correlated part (bias) (Nichols and Griffith, 2000 and Derber, 1989).

In this paper we will introduce a new technique for treating model errors, based on an idea of Zou, Navon and Le Dimet (1992). This method is an extension of nudging (or Newtonian relaxation) which was the first operational data assimilation suite in oceanography (Lyne et al, 1982; Krishnamurti et al, 1991 and Lorenc et al, 1991). It consists in relaxing the model state towards the observations during the assimilation period by adding a non physical 'nudging' term to the model equation. The nudging terms are defined as the difference between the observation and the model solution weighted by a nudging coefficient.

The nudging can be thought as an approximation of the Kalman filter (KF) the best nudging coefficients are those related to a KF in linear case (see Lorenc, 1986 and Lorenc et al,1991). However, the KF is very costly in practice (storage and manipulation of  $N \times N$  matrices where  $N$  is the size of the state vector) compared to the very cheap nudging. Moreover, in order to be applied, KF need simplifications that does not take into account the whole time period.

In this paper we will use optimal control methods to estimate both initial condition (like in 4D-Var) and the nudging coefficients in order to correct model errors in a much more efficient way (from the computational point of view).

First we will recall the equations of the 4D-Var and introduce Optimal Nudging, and we will show that this new method and the Kalman Filter are equivalent in a sense to be defined. Section 2 deals with the application of optimal nudging data assimilation on a very simple Burger's equation. Three forms of the nudging matrix will be tested to determine which compromise between the size of control and the quality of the results will be optimal.

And finally we will test this method on a more realistic shallow water model and compare it with classical 4D-Var data assimilation on the same model. We will show that for a little extra-cost this new method provides a quite sensitive improvement.

The notations used in this paper are the unified ones advocated by Ide et al (1997).

## 2 Optimal nudging: motivation

The aim of the 4D-Var classical approach of data assimilation is to try to reduce the misfit between the observations and the forecast state by controlling the initial condition of the analysis period.

Given a discretised model  $M$ , let  $\mathbf{x} \in \mathcal{C} \subset \mathbb{R}^N$ ,  $\mathbf{x}^b$  the background state, or first guess of the minimisation, the evolution of the state can be described as

$$\begin{cases} \mathbf{x}_0 = \mathbf{x}^b + \delta\mathbf{x}_0 \\ \mathbf{x}_{i+1} = M_{(t_i, t_{i+1})}(\mathbf{x}_i) \end{cases} \quad (1)$$

The aim of the method is to search for  $\delta\mathbf{x}_0$  that minimises the following cost function :

$$J(\delta\mathbf{x}_0) = \underbrace{\frac{1}{2} \sum_{i=0}^n \langle \mathbf{R}^{-1}(H_i(\mathbf{x}_i) - \mathbf{y}_i^o), H_i(\mathbf{x}_i) - \mathbf{y}_i^o \rangle}_{J_o} + \underbrace{\frac{1}{2} \langle \mathbf{B}^{-1} \delta\mathbf{x}_0, \delta\mathbf{x}_0 \rangle}_{J_b}$$

where  $\mathbf{R}$  is the  $m \times m$  observational error covariance matrix and  $\mathbf{B}$  the  $n \times n$  background error covariance matrix,  $\mathbf{y}_i^o \in \mathbb{R}^m$  is the observation vector at time  $t_i$  and  $H_i$  is the observation operator that computes the model equivalent quantities and interpolates them to the observation location at time  $t_i$ . The first term of the cost function (called  $J_o$ ) represents the misfit we wish to minimise in an idealised case with true observations and a model describing exactly the evolution of the system. However due to observation errors and errors induced by the model, this term can not be canceled. That is why one introduces the second term (called  $J_b$ ) that prevents the solution to be non physical (trying to fit exactly the observations) and then generating oscillations to retrieve physical equilibrium. The appropriate equilibrium between the two terms is provided by the inverse of the error covariances matrices  $\mathbf{B}^{-1}$  and  $\mathbf{R}^{-1}$  which represents the confidence we have in the background and the observations respectively.

Due to the term  $J_b$  and to the model errors, the innovation residual  $\mathbf{d}_k = H_k(\mathbf{x}_k^a) - \mathbf{y}_k^o$  remains non zero after the 4D-Var assimilation.

The errors sources can be written as :

- Prediction (or Forecast) Error :

$$\varepsilon_k^f = \mathbf{x}_k^f - \mathbf{x}_k^t$$

- Observation Error :

$$\varepsilon_k^o = \mathbf{y}_k^o - H_k(\mathbf{x}_k^t)$$

where  $\mathbf{x}_k^f$  represents the forecast state at time  $t_k$  and  $\mathbf{x}_k^t$  the true state.

Then we note that the innovation residual can be written as :

$$\begin{aligned} \mathbf{d}_k &= H_k(\mathbf{x}_k^f) - \mathbf{y}_k^o \\ &\approx \mathbf{H}_k \varepsilon_k^f - \varepsilon_k^o \end{aligned} \quad (2)$$

where  $\mathbf{H}_k = \frac{\partial H_k}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\mathbf{x}_k^f}$  is the tangent linear of the observation operator, and  $\varepsilon^f$  is assumed to be negligible with respect to  $\mathbf{x}^f$

Equation (??) shows that an important information about model and observations errors remains in this residual. In the following we will try to use it to improve the 4D-Var assimilation scheme.

Formally, in the case of a completely and directly observed state an intuitive approach was to modify the forecast state using a correcting term including the innovation residual. In this case, the inverse of the observation operator  $\mathbf{H}^{-1}$  exists and then we could write :

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{H}_k^{-1}(\mathbf{d}_k - \varepsilon^o)$$

Among existing data assimilation methods one can quote the Kalman filter (Kalman, 1960) which provides an analysed state using the innovation vector to correct the forecast by

$$\mathbf{x}_k^a = \mathbf{x}_k^f + \mathbf{K}_k \mathbf{d}_k$$

where  $\mathbf{K}$  is called the gain matrix and is computed as follows :

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

with  $\mathbf{P}^f$  the forecast error covariance matrix

It can be pointed out that in Kalman Filter, if observations are assumed exact ( $\mathbf{R} = 0$ ), we retrieve the gain matrix as :  $\mathbf{K}_n = \mathbf{H}_n^{-1}$  and  $\varepsilon_n^f = \mathbf{H}_n^{-1} \mathbf{d}_k$

However there are two major drawbacks in Kalman filtering : it requires the manipulation of very large matrices and it needs simplifications that do not take into account the whole time period.

That is why we try here to estimate an optimal 'Gain' matrix using variational methods (4D-Var). The following extend the method of Optimal Nudging introduced by Zou et al (1992) exploring three kind of nudging matrices and introducing the smoothed optimal nudging which avoid the shocks induced by nudging corrections in the model.

### 3 Optimal nudging: the method

In this section we introduce the main point of this paper : the Four Dimensional Optimal Nudging Data Assimilation scheme (4D-ON). The computation of the gradient of the cost function does not require more

tools than for classical 4D-Var but only a few more operations and storage.

The aim of nudging methods is to relax the model states toward the observation adding a 'nudging' term. This term is the misfit between observation and forecast  $\mathbf{y}_{i+1}^o - H_{i+1}(M_{(t_{i+1}, t_i)}(\mathbf{x}_i))$  weighted by the nudging operator  $\mathbf{G}_i$ . In this part the complete nudging case is considered that is to say that  $\mathbf{G}_i$  is a  $n \times m$  matrix where  $n$  is the dimension of the state vector and  $m$  the dimension of the observation vector. The problem is now: how to estimate  $\mathbf{G}_i$ .

Let  $M$  be a non-linear discretised model describing the ocean or atmosphere evolution. We focus on the problem :

$$\begin{cases} \mathbf{x}_0 = \mathbf{x}^b + \delta\mathbf{x}_0 \\ \mathbf{x}_{i+1} = M_{(t_{i+1}, t_i)}(\mathbf{x}_i) + \mathbf{G}_{i+1}(\mathbf{y}_{i+1}^o - H_{i+1}(M_{(t_{i+1}, t_i)}(\mathbf{x}_i))) \end{cases} \quad (3)$$

And we wish to minimise the cost function :

$$\begin{aligned} J(\delta\mathbf{x}_0, \mathbf{G}) = & \frac{1}{2} \sum_{i=0}^n \langle \mathbf{R}^{-1} H_i(\mathbf{x}_i) - \mathbf{y}_i^o, H_i(\mathbf{x}_i) - \mathbf{y}_i^o \rangle + \frac{1}{2} \langle \mathbf{B}^{-1} \delta\mathbf{x}_0, \delta\mathbf{x}_0 \rangle \\ & \underbrace{\hspace{15em}}_{J_{nudz}} \\ & + \frac{1}{2} \sum_{i=1}^n \langle \mathbf{Q}_i^{-1} \mathbf{G}_i (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))), \mathbf{G}_i (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))) \rangle \end{aligned} \quad (4)$$

Basically we will consider a 4D-Var cost function where  $\mathbf{R}$  is the observation error correlation matrix and  $\mathbf{B}$  is the background error covariance matrix, plus a regularisation term  $J_{nudz}$  that prevents  $\mathbf{G} = \mathbf{G}_1, \dots, \mathbf{G}_n$  to be too large, where  $\mathbf{Q}$  is the forecast error covariance matrix.

Actually, the regularisation term comes from the classical formulation of the cost function of the 4D-Var with imperfect model data assimilation (see Jazwinsky, 1970; Tikhonov and Arsenin, 1977; Cohn, 1997 and Alekseev and Navon, 2001]). In this entire formulation of the control of model errors the cost function is

$$\begin{aligned} J(\delta\mathbf{x}_0, \eta_1, \dots, \eta_n) = & \frac{1}{2} \sum_{i=0}^n \langle \mathbf{R}^{-1} H_i(\mathbf{x}_i) - \mathbf{y}_i^o, H_i(\mathbf{x}_i) - \mathbf{y}_i^o \rangle \\ & + \frac{1}{2} \sum_{i=1}^n \langle \mathbf{Q}_i^{-1} \eta_i, \eta_i \rangle + \frac{1}{2} \langle \mathbf{B}^{-1} \delta\mathbf{x}_0, \delta\mathbf{x}_0 \rangle \end{aligned} \quad (5)$$

where  $\eta_i$  is the correcting term at time  $t_i$ . So  $\eta_i$  just have been replaced with the nudging correcting term.

In order to compute the gradient of the cost function we will use a classical Lagrangian method under constraint of (??). Let  $L$  the Lagrangian being defined by:

$$L(\mathbf{G}, \delta \mathbf{x}_0; \mathbf{x}^*) = J(\mathbf{G}, \delta \mathbf{x}_0) + \sum_{i=1}^n \langle \mathbf{x}_i^*, \mathbf{x}_i - M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}) - \mathbf{G}_i(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))) \rangle$$

$\mathbf{x}^*$  being the Lagrangian multiplier. Computing the partial derivatives

$$\frac{\partial L}{\partial \mathbf{x}^*}(\mathbf{G}, \mathbf{v}; h_{\mathbf{x}}^*) = \mathbf{x}_i - M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}) - \mathbf{G}_i^T(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1})))$$

we can first retrieve :

$$\frac{\partial L}{\partial \mathbf{x}^*} = 0 \implies (??)$$

Moreover,

$$\begin{aligned} \frac{\partial L}{\partial \delta \mathbf{x}_0}(\mathbf{G}, h_{\delta \mathbf{x}_0}, \mathbf{x}^*) &= \sum_{i=0}^n \langle \mathbf{H}_i^T(H_i(\mathbf{x}_i) - \mathbf{y}_i^o), \hat{\mathbf{x}}_i \rangle + \langle \mathbf{B}^{-1} \delta \mathbf{x}_0, h_{\delta \mathbf{x}_0} \rangle \\ &- \sum_{i=1}^n \langle \mathbf{M}_{(t_i, t_{i-1})}^T \mathbf{H}_i^T \mathbf{G}_i^T \mathbf{Q}^{-1} \mathbf{G}_i(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))), \hat{\mathbf{x}}_{i-1} \rangle \\ &- \sum_{i=1}^n \langle \mathbf{x}_i^*, \hat{\mathbf{x}}_i \rangle + \langle \mathbf{M}_{(t_i, t_{i-1})}^T \mathbf{x}_i^*, \hat{\mathbf{x}}_{i-1} \rangle - \langle \mathbf{M}_{(t_i, t_{i-1})}^T \mathbf{H}_i^T \mathbf{G}_i^T \mathbf{x}_i^*, \hat{\mathbf{x}}_{i-1} \rangle \\ &- \langle \mathbf{x}_0^*, h_{\delta \mathbf{x}_0} \rangle \end{aligned}$$

So, if the adjoint state  $\mathbf{x}^*$  is defined by:

$$\begin{cases} \mathbf{x}_n^* = \mathbf{H}_n^T \mathbf{R}^{-1} (H_n(\mathbf{x}_n) - \mathbf{y}_n^o) \\ \mathbf{x}_i^* = (\mathbf{M}_{(t_{i+1}, t_i)} - \mathbf{G}_{i+1} \mathbf{H}_{i+1} \mathbf{M}_{(t_{i+1}, t_i)})^T \mathbf{x}_{i+1}^* + \mathbf{H}_i^T \mathbf{R}^{-1} H_i(\mathbf{x}_i) - \mathbf{y}_i^o \\ \quad - \mathbf{M}_{(t_{i+1}, t_i)}^T \mathbf{H}_{i+1}^T \mathbf{G}_{i+1}^T \mathbf{Q}^{-1} \mathbf{G}_{i+1}(\mathbf{y}_{i+1}^o - H_{i+1}(M_{(t_{i+1}, t_i)}(\mathbf{x}_i))) \end{cases} \quad (6)$$

we obtain ,  $\frac{\partial L}{\partial \delta \mathbf{x}_0}(\mathbf{G}, h_{\delta \mathbf{x}_0}, \mathbf{x}^*) = \langle \nabla_{\delta \mathbf{x}_0} J, h_{\delta \mathbf{x}_0} \rangle$  i.e :

$$\nabla_{\delta \mathbf{x}_0} J = -\mathbf{x}^*(0) + \mathbf{B}^{-1} \delta \mathbf{x}_0 \quad (7)$$

And finally :

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{G}_i}(\mathbf{v}, h_{\mathbf{G}_i}, \mathbf{x}_i^*) &= \langle \mathbf{Q}^{-1} \mathbf{G}_i(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))), h_{\mathbf{G}_i}(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))) \rangle \\ &- \langle \mathbf{x}_i^*, h_{\mathbf{G}_i}(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))) \rangle \\ &= \langle \nabla_{\mathbf{G}} J, h_{\mathbf{G}} \rangle \end{aligned} \quad (8)$$



thus the gradient of the cost function with respect to  $\mathbf{G}_i$  can be written :

$$\begin{aligned} \nabla_{\mathbf{G}_i} J &= \mathbf{Q}^{-1} \mathbf{G}_i (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))) (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1})))^T \\ &\quad - \mathbf{x}^*(t) (\mathbf{y}^o(t_i) - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1})))^T \end{aligned} \quad (9)$$

Provided the initial condition contribution to the gradient (eqn (??)) is computed (thanks to one integration of the direct model and of the adjoint model) the computation of the nudging contribution of the gradient does not need extra ‘heavy’ computations. Indeed  $\nabla J_{\mathbf{G}_i}$  is obtained by the product of vectors used in  $\nabla J \delta \mathbf{x}_0$  (i. e.  $\mathbf{G}_i (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1})))$ ,  $(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1})))^T$  and  $\mathbf{x}^*$ ).

It can be proved that this algorithm is formally equivalent to the kalman filter (see appendix A).

## 4 Basic experiment on a Burger’s equation (1D)

In this section we will apply the optimal nudging data assimilation scheme on a simple one-dimensional Burger’s equation. It can be an easy validation of the O.N. scheme and will permit to find an appropriate choice for the form of  $\mathbf{G}_i$

Let the model be

$$\begin{cases} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} = f \\ u(0, t) = u(1, t) = 0 \\ u(x, 0) = \sin \pi x \end{cases} \quad (10)$$

with  $x \in ]0, 1[$ ,  $t > 0$

Note that a known analytic exact solution is associated to fixed forcing term ( $f$ ) . For this experiment we impose  $f$  to be  $f(x, t) = e^{-t}(\pi u \cos \pi x + (-1 + \nu \pi^2) \sin \pi x)$ , and then  $u = e^{-t} \sin \pi x$  is an exact solution of (??).

The 4D-ON is defined by adding the nudging term into the model equations :

$$\begin{cases} \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} = f + g(u_{obs} - h(u)) \\ u(0, t) = u(1, t) = 0 \\ u(x, 0) = \sin \pi x \end{cases} \quad (11)$$

We will focus on a sequence  $[0,1]$  split into  $N + 1$  time steps. The same holds for spatial discretisation of  $[0,1]$  in  $E$  points. In this experiment  $N = 5000$  and  $E = 20$ .

During this time sequence, we have  $M$  observations located at  $(x_m, t_m)$   $m = 1..M$ .

Therefore we can rewrite our model as :

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} = f + \sum_{m=1}^M g_m (u_m^{obs} - u_m) \delta(x - x_m) \delta(t - t_m)$$

In a discrete formulation :

$$\frac{d\mathbf{u}}{dt}(t) + M(\mathbf{u})(t) = F(t) + \sum_{k=1}^K \mathbf{G}_k (\mathbf{u}_k^{obs} - H(\mathbf{u}_k)) \delta(t - t_k)$$

where  $H$  is the observation operator. In this experiment the minimisation is performed by a home made conjugate gradient method using the Hessian of  $J$  as a preconditioner. The Hessian-vector product is computed using the second order adjoint of the model (see Wang et al, 1992 and Le Dimet et al, 1998)

We will apply two types of model error simultaneously on the forcing term  $f$  : a five percent white noise and a ten percent correlated noise (bias). The observations are obtained from the model free of error, only one point over five in space and one time step over fifty will be retained as observation and we add a five percent white noise on these observations. Moreover, we add 20% noise to the initial condition in order to simulate background errors, and we setting all the control vector (increment of initial condition, and nudging coefficients) to 0.

Now we will analyse the results of three types of nudging coefficient matrices:

1.  $\mathbf{G}_k$  is a full rank  $E \times M$  matrix, i.e. at each point, the correcting term is a linear combination of all forecast/observation misfit at the same time. This is the “dream” case, because each component of the state vector will be corrected and not only observed . Even if only a very small size problem allows this form of  $\mathbf{G}_k$ , this experiment allows to compare the efficiency of this ideal case with simpler forms of the nudging matrix and to emphasise the fact that the most is not necessarily the best.
2.  $\mathbf{G}_k$  is a diagonal matrix. In this case the model equations are only corrected at observation locations and the correction coefficient is different for each observation.
3.  $\mathbf{G}_k$  is a scalar coefficient (just time varying). It is the same case as previously but for a fixed time  $t_k$  the correction coefficient is the same at every observation location.

Using one of these three forms of  $\mathbf{G}_k$  increases the size of the control vector (in relation to the classical 4D-Var which control size is  $E$ ) respectively by  $E \times M \times N$ ,  $M \times N$  and  $N$ .

In table (??) we compare the level of error RMS and the computational cost for these three forms of 4D-ON. The first column shows the “raw” results of the assimilation and the second one is obtained using a linear time interpolation of correcting terms as temporal smoothing and , for  $\mathbf{G}$  diagonal matrix and scalar coefficients, a convolution product with a gaussian function  $\lambda$  as spatial smoothing (eqn [??]).

$$\mathbf{u}(t) = \mathbf{u}(t) + \mathbf{G}_k(\mathbf{u}_k^{obs} - H(\mathbf{u}(t))) \text{ if } t = t_k$$

The level of rms of error is computed comparing the results of the assimilation to the known analytic solution.

$$\mathbf{u}(t) = \mathbf{u}(t) + \left( \frac{t-t_k}{t_{k+1}-t_k} \mathbf{G}_{k+1} + \frac{t_{k+1}-t}{t_{k+1}-t_k} \lambda \star (\mathbf{G}_k) \right) \left( \frac{t-t_k}{t_{k+1}-t_k} \mathbf{u}_{k+1}^{obs} + \frac{t_{k+1}-t}{t_{k+1}-t_k} \mathbf{u}_k^{obs} - H(\mathbf{u}(t)) \right) \text{ if } t_k < t < t_{k+1} \quad (12)$$

As a fair conclusion of this first experiment it can be noticed that despite all the 4D-ON schemes provide better results than classical 4D-Var, the complete 4D Optimal Nudging ( $\mathbf{G}_k$  full matrix, fig.??) does not really provide better results than other alternatives. Even if in the raw method case the full matrix is a little bit more efficient compared to simpler forms of  $\mathbf{G}_k$ , the difference does not justify the huge computational over-cost induced (the full matrix control is more than 60 times more expensive than the scalar control !). Moreover in the results showed for interpolated methods the diagonal matrix and the scalar coefficient provide better results. This can be explained by the too many degrees of freedom that the complete method contains. The simpler the control vector, the better conditioned the optimisation problem, which leads to a much more efficient minimisation. Moreover we can notice that both diagonal matrix and scalar coefficient give almost the same results (different but indistinguishable on the graphs) which could indicate that the misfit vector contains spatial informations i.e the misfits vector is itself well equilibrated on the whole grid (the optimal nudging coefficient is nearly the same for all the observed points at a given time).

## 5 Numerical experiment with a non-linear shallow water model

### 5.1 The Model

For this testing experiment, we use a non-linear one-layer shallow-water model on a square basin with flat bottom (this model is classically used, see for example Adcroft and Marshall, 1998).

Even if this model is not a very realistic one, the non-linearities included in the equation and the size of the state vector (more than  $10^4$ ) make its complexity sufficient to be a relevant test case.

Actually, the framework of this experiment being a twin experiment one. That is to say that the observations do not come from the reality but from the model (with light difference from the the model used in assimilation). In this way two different models were used. The first one is used to represent the reality, it provides a simulated ‘true’ state evolution and then using the observation operator (and possibly a white noise), the synthetic observations can be obtained. The second one represents the model (called forecast model in the following by opposition to the true model or the truth evolution). In this way the presence of model errors is simulated and the evolution of the (pseudo) reality is known.

Both evolutions of  $\mathbf{x}_{true/forecast} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{h} \end{bmatrix}$  are described by :

$$\begin{cases} \partial_t \mathbf{u} - (f + \zeta) \mathbf{v} + \partial_x B = \frac{\tau}{\rho_0 \mathbf{h}} - r \mathbf{u} + \nu \Delta \mathbf{u} \\ \partial_t \mathbf{v} - (f + \zeta) \mathbf{u} + \partial_y B = \frac{\tau}{\rho_0 \mathbf{h}} - r \mathbf{v} + \nu \Delta \mathbf{v} \\ \partial_t \mathbf{h} + \partial_x(\mathbf{h} \mathbf{u}) + \partial_y(\mathbf{h} \mathbf{v}) = 0 \end{cases} \quad (13)$$

where  $(\mathbf{u}, \mathbf{v})$  represents the current velocity,  $\mathbf{h}$  is the height of the layer,  $\zeta = \partial_x \mathbf{v} - \partial_y \mathbf{u}$  is the relative vorticity,  $B = g^* \mathbf{h} + \frac{1}{2}(\mathbf{u}^2 + \mathbf{v}^2)$  is the Bernoulli potential,  $g^*$  is the reduced gravity,  $r$  is the linear friction coefficient and  $\nu$  is the viscosity coefficient.

The forcing terms of the models are :

- wind :  $\tau_{forecast} = \tau_0 \frac{\sin(2\pi(y - \frac{L}{4}))}{L} L \vec{i}$ , and  
 $\tau_{true} = \tau_{forecast} \times [1 + 0.8 \times \sin(\frac{2\pi t}{\Delta t \times 480})]$  where  $L$  is the basin length
- Coriolis factor :  $f = f_0 + \beta y$ .

boundary conditions :

- $\mathbf{v}=0$  on north/south boundaries and  $\mathbf{u}=0$  on east-west boundaries
- non-slippery boundary conditions.

In this experiment the numerical values are:

$L = 2000 \text{ km}$ ,  $f_0 = 0.7 \times 10^{-4} \text{ s}^{-1}$ ,  $\beta = 2 \times 10^{-11} \text{ m}^{-1} \text{ s}^{-1}$ ,  $\nu = 15 \text{ m}^2 \text{ s}^{-1}$  (forecast),  $\nu = 0,9 \times 15 \text{ m}^2 \text{ s}^{-1}$  (true),  $r = 10^{-7} \text{ s}^{-1}$  (forecast),  $r = 0,9 \times 10^{-7} \text{ s}^{-1}$  (true),  $\rho_0 = 10^3 \text{ kg m}^{-3}$ ,  $g^* = 0.02 \text{ m s}^{-2}$ , et  $\tau_0 = 0.015 \text{ N m}^{-2}$ .

For the spatial discretisation a second order centered scheme is used on an Arakawa C-grid with  $\Delta x = \Delta y = 25 \text{ km}$  and for the time discretisation, a leap-frog scheme with a time step size  $\Delta t = 30 \text{ mn}$  and an Asselin time filter is added in the equations.

The initial condition (first guess) of our data assimilation window is provided using a 6 years of the forecast model whereas true initial state and observations are computed with a true model spin-up.

## 5.2 The Assimilation

The cost function is defined by (??) but in order to improve the minimisation efficiency we operate a change of variable,

$$\mathbf{v} = \mathbf{B}^{-\frac{1}{2}} \delta \mathbf{x}_0 \quad (14)$$

where  $\mathbf{B}^{-\frac{1}{2}}$  is defined as  $A = A^{\frac{1}{2}} A^{\frac{T}{2}}$ . This preconditioning has two advantages:

- $\mathbf{B}^{-\frac{1}{2}}$  is the square root of the Hessian of  $J_b$  (and then a very good preconditioner for the minimisation) (Thepaut and Moll, 1990; Yang et al, 1996; Courtier, 1997 and Derber and Bouttier, 1999)
- it implies the manipulation of  $\mathbf{B}^{\frac{1}{2}}$  instead of its inverse (actually the inverse of ??  $\delta \mathbf{x}_0 = \mathbf{B}^{\frac{1}{2}} \mathbf{v}$  is performed at the beginning of each iteration in order to retrieve a convenient initial condition).

Even in this simplified case (and consequently for more realistic ones) the entire  $\mathbf{Q}$  and  $\mathbf{B}$  matrices can neither be estimated nor stored explicitly because of their size ( $\mathbf{B}$  is an  $N \times N$  matrix where  $N$  is the size

of state vector). We are therefore forced to see  $\mathbf{Q}$  and  $\mathbf{B}$  as operators. Roughly (for more explanation see Weaver and Courtier, 2001) following Derber and Bouttier (1999) we can write for instance  $\mathbf{B}$  :

$$\mathbf{B} = \mathbf{K}\mathbf{B}_u\mathbf{K}^T$$

where  $\mathbf{K}$  is a balance operator that relies one state variable to the other (here  $\mathbf{K}$  uses the geostrophic balance to compute the balanced part of  $\mathbf{u}$  and  $\mathbf{v}$  from  $\mathbf{h}$ ) and  $\mathbf{B}_u$  is the error covariance matrix for the unbalanced part of variables and assumed to be block diagonal (i.e the cross covariances of the unbalanced part of variables are negligible)

$$\mathbf{B}_u = \Sigma_{\mathbf{B}}\mathbf{C}\Sigma_{\mathbf{B}}$$

where  $\Sigma_{\mathbf{B}}$  is the diagonal matrix of background error standard deviation and represents the symmetric matrix of background error correlations for the unbalanced part of the state variables. The  $\mathbf{C}$  operator is modeled using a diffusion equation (see Weaver and Courtier, 2001) Therefore the inverse of change of variable can be written

$$\delta\mathbf{x}_0 = \mathbf{K}\Sigma_{\mathbf{B}}\mathbf{C}^{\frac{1}{2}}\mathbf{v}$$

The same way is used to build  $\mathbf{Q}$ , modifying standard error deviations ( $\Sigma_{\mathbf{Q}}$ ) and possibly the parameters of the diffusion equation ( $\mathbf{C}$ ).

In the same way, for this more realistic problem we can not really control the entire  $\mathbf{G}$  matrices but only a few coefficients. But this implies that the corrections are only applied on observed locations. In order to correct this drawback we can pre multiply the correcting term by  $\mathbf{Q}^{\frac{1}{2}}$  ; this will smooth the correction and simplify the computation of the cost function.

Now we can rewrite equation (??)

$$\begin{cases} \mathbf{x}_0 = \mathbf{x}^b + \mathbf{B}^{\frac{1}{2}}\mathbf{v} \\ \mathbf{x}_{i+1} = M_{(t_{i+1}, t_i)}(\mathbf{x}_i) + \mathbf{Q}^{\frac{1}{2}}\mathbf{G}_{i+1}(\mathbf{y}_{i+1}^o - H_{i+1}(M_{(t_{i+1}, t_i)}(\mathbf{x}_i))) \end{cases} \quad (15)$$

And the cost function

$$\begin{aligned}
J(\mathbf{v}, \mathbf{G}) &= \frac{1}{2} \sum_{i=0}^n \langle \mathbf{R}^{-1} H_i(\mathbf{x}_i) - \mathbf{y}_i^o, H_i(\mathbf{x}_i) - \mathbf{y}_i^o \rangle \\
&\quad + \frac{1}{2} \sum_{i=1}^n \langle \mathbf{G}_i(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))), \mathbf{G}_i(\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))) \rangle \\
&\quad + \frac{1}{2} \langle \mathbf{v}, \mathbf{v} \rangle
\end{aligned} \tag{16}$$

This change of variable modifies the adjoint equations

$$\left\{ \begin{array}{l}
\mathbf{x}_n^* = \mathbf{H}_n^T \mathbf{R}^{-1} (H_n(\mathbf{x}_n) - \mathbf{y}_n^o) \\
\mathbf{x}_i^* = (\mathbf{M}_{(t_{i+1}, t_i)} - \mathbf{Q}^{\frac{1}{2}} \mathbf{G}_{i+1} \mathbf{H}_{i+1} \mathbf{M}_{(t_{i+1}, t_i)})^T \mathbf{x}_{i+1}^* + \mathbf{H}_i^T \mathbf{R}^{-1} H_i(\mathbf{x}_i) - \mathbf{y}_i^o \\
\quad - \mathbf{M}_{(t_{i+1}, t_i)}^T \mathbf{H}_{i+1}^T \mathbf{G}_{i+1}^T \mathbf{G}_{i+1} (\mathbf{y}_{i+1}^o - H_{i+1}(M_{(t_{i+1}, t_i)}(\mathbf{x}_i)))
\end{array} \right. \tag{17}$$

The gradients become

$$\left\{ \begin{array}{l}
\nabla_{\mathbf{v}} J = \mathbf{B}^{\frac{T}{2}} \mathbf{x}_0^* + \mathbf{v} \\
\nabla_{\mathbf{G}_i} J = \mathbf{Q}^{\frac{T}{2}} \mathbf{x}_i^* (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1})))^T \\
\quad + \mathbf{G}_i (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1}))) (\mathbf{y}_i^o - H_i(M_{(t_i, t_{i-1})}(\mathbf{x}_{i-1})))^T
\end{array} \right. \tag{18}$$

Note that this second change of variable cannot be considered as a good preconditioner because  $\mathbf{Q}^{\frac{1}{2}}$  is not the square root of the inverse of the Hessian of  $J_{nudg}$ . Thus the minimization efficiency can be improved by considering  $J_{nudg} = \frac{1}{2} \sum_{i=1}^n \langle \tilde{\mathbf{Q}}^{-1} \mathbf{G}_i, \mathbf{G}_i \rangle$  (before the change of variable) where  $\mathbf{G}_i$  is a state dimension vector.

Although the optimization problem is modified and the Kalman equivalence (see appendix A) is lost, the assimilation provides roughly the same quality of results compared with preceding  $J_{nudg}$  and the number of needed iterations is highly reduced (about 1/10).

In the following, in both  $\mathbf{B}$  and  $\mathbf{Q}$  operators the correlation scales are about 250 km.

The minimisation method used in the following experiment is M1QN3 Quasi-Newton with limited memory developed at INRIA by Gilbert and Lemaréchal (1989). Both classical 4D-Var and 4D-ON methods have been implemented with the PALM modular data assimilation system. This software package allows a full modularity by splitting data assimilation algorithms into elementary units (see appendix B and Piacentini et al, 2000), therefore it has been of a great help for the construction and the handling of this two methods.

### 5.3 Numerical experiment

In this section we will compare the 4D-ON results with classical 4D-Var ones. In both case, the control vectors is set to 0 at the beginning of the assimilation. The assimilation is carried out for 30 days. This time window represent a relevant time scale of ocean model, furthermore the tangent linear hypothesis<sup>1</sup> which ensure the validity of the adjoint computation of the gradient, is still valid. Following the results of the burger’s experiment we only focus on nudging coefficient as time-dependant scalar coefficients. The observations are simulated as satellite tracks (see Fig.??) with one track every 3 days. This is not a favorable case for nudging because of the lack of information provided by the observations; indeed only 500 grid points are observed whereas the state vector size is about 20000). In Fig.?? value of the different terms of the cost function and the initial and the final values of the gradient are shown. The norme of the gradient decrease of 6 orders of magnitude during the minimization. We note a rapid decrease in the  $J_o$  term of the cost function during the first few iterations. After six iterations the algorithm has already converge. Fig.?? shows the corresponding evolution of the nudging coefficient  $\mathbf{G}_h$  according to time at several steps of the minimization and confirms the convergence after six iterations.

Due to the undersampling of the observation, comparing the different level of RMS (computed thanks to the knowledge of the “true” trajectory), it ca be noticed that, with no care, the former version of 4D-ON (long-dotted line) gives slightly better results compared to 4D-Var (plain line) and bad results compared to smoothed 4D-ON (with the change of variable) (short dotted line, see Fig.??). The non-smoothed 4D-ON corrects the model only on observed locations therefore it is not consistent with the other coordinates and then gravity waves can be created to return to a more coherent state (nasty fluctuations on Fig.??). Moreover the  $J_o$  term is only computed by the corrected terms, so the cost function does not really reflect the gap between reality and model state. Indeed the corrected state can be a smooth field with some Dirac on observed locations and the  $J_o$  term will be very small.

The smoothed version of 4D-ON allows the under-sampling of the observation, replacing a Dirac correction by a Gaussian correction (with the help from  $\mathbf{Q}^{\frac{1}{2}}$ ). In addition to spatial smoothing, a temporal smoothing is performed multiplying the correcting term by a time decreasing coefficient. The time smoothing distance is six hours, this parameter has been tuned by hand, and a more generalistic estimation may be explored as a further development of this method.

---

<sup>1</sup> $M_{(t_0, t_N)}(\mathbf{x}^b + \delta\mathbf{x}_0) = M_{(t_0, t_N)}(\mathbf{x}^b) + \mathbf{M}_{(t_0, t_N)}\delta\mathbf{x}_0$



All these developments permit 4D-ON to give noticeable improvement to classical 4D-Var for the same number of iterations. On the observed component (sea surface elevation) the improvement compared to 4D-Var is very significant (RMS of the error is up to 50% less, see Fig.??). On the other components (current velocity Fig.??) thanks to  $\mathbf{K}$ , the balance operator included in  $\mathbf{Q}$ , the gain of RMS error remain about 10%.

The introduction of smoothing implies, for this Shallow-water model, that the computational cost of one iteration of 4D-ON is about 10 percent more expensive than a 4D-Var one and may be smaller for bigger models. Compared to the gain provided by the 4D-ON this over-head is not a limitation.

Due to shocks induced by nudging corrections, we want to check if the analysed state obtained at the end of the 4D-ON assimilation window will not be consistent with the dynamic and then will provide a quality forecast. Figures ?? and ?? show that it is not the case: after the assimilation on the left part of the graph (under the grid) a one month forecast is performed starting from analysed state at the end of the assimilation window. Due to model errors the RMS of error become more and more important, but the deviation is roughly the same for 4D-Var than for smoothed 4D-ON i.e. the analysed state obtained by optimal nudging is as consistent as by classical 4D-Var (but closer to the true state).

## 6 Conclusion

In this paper is presented a new data assimilation method, starting from 4D-Var and nudging. We have shown that, provided the change of variable introduced in section ?? is done, the 4D-ON gives encouraging results compared to 4D-Var. Indeed owing to spatial and temporal smoothing the 4D-ON gives better results than 4D-Var. Moreover it is very easy to implement (low algorithm modification - all new terms are already calculated) even if the theory is quite complicated.

The experiment on Burger's equation (section ??) shows that the control of entire  $\mathbf{G}$  matrix is not necessary and even harmful. Therefore the control can be restricted to a few scalar coefficients without lost quality.

Thanks to this simplification and due to the current growth of computational means, the extra cost induced by the control of nudging terms (roughly +10%) is not really a limitation to practical applications and the 4D-ON could be successfully implemented in the near future on a variety of meteorological or oceanographic problems.

In parallel to its implementation in realistic models, further development can be carried out: In the considered case  $\mathbf{C}_Q$  the diffusion operator included in operator  $\mathbf{Q}$  is the same as the background diffusion operator  $\mathbf{C}_B$  but there is no reasons to that, the model error and the background error do not have necessarily the same correlation structures. Finding the adequate balance could be a natural future development of this method.

## A Links between Kalman Filter and Optimal Nudging

In this section we will point out that in a linear case, the Kalman filter and the Optimal Nudging method introduced in previous section, will provide the same results. This equivalence is based on a demonstration provided by Li and Navon (2001) for a more classical 4D-Var.

Let now consider Optimal nudging for a discrete linear model where the  $\mathbf{G}_i$  are considered to be  $n \times m$  matrices

$$\begin{cases} \mathbf{x}_0 = \mathbf{x}^b + \delta \mathbf{x}_0 \\ \mathbf{x}_{i+1} = \mathbf{M}_{(t_i, t_{i+1})} \mathbf{x}_i + \mathbf{G}_{i+1} (\mathbf{y}_{i+1}^o - \mathbf{H}_{i+1} (\mathbf{M}_{(t_i, t_{i+1})} \mathbf{x}_i)) \end{cases} \quad (19)$$

For the sake of simplicity, we note  $\mathbf{d}_i = \mathbf{y}_i^o - \mathbf{H}_i (\mathbf{M}_{(t_i, t_i)} \mathbf{x}_{i-1})$  and  $\mathbf{c} = (\delta \mathbf{x}_0, \mathbf{G}_1, \dots, \mathbf{G}_n)$ . We would like to minimize the cost function :

$$J(\mathbf{c}) = \frac{1}{2} \sum_{i=0}^n \langle \mathbf{R}^{-1} (\mathbf{H}_i (\mathbf{x}_i) - \mathbf{y}_i^o), \mathbf{H}_i (\mathbf{x}_i) - \mathbf{y}_i^o \rangle + \frac{1}{2} \sum_{i=0}^n \langle \mathbf{Q}_i^{-1} \mathbf{G}_i \mathbf{d}_i, \mathbf{G}_i \mathbf{d}_i \rangle + \langle \mathbf{B}^{-1} \delta \mathbf{x}_0, \delta \mathbf{x}_0 \rangle$$

where  $\langle \cdot, \cdot \rangle$  denotes the ‘‘classical’’ Euclidean inner product, and  $(\cdot, \cdot)$  the Frobenius inner product between two matrices :  $(A, B) = \sum_j \sum_i a_{ij} b_{ij}$ .

First of all, we can note that there exists a relationship between those two inner products. If  $\mathbf{X}$  is a  $n$  vector,  $\mathbf{Y}$  a  $m$  vector and  $B$  a  $n \times m$  matrix then

$$\langle \mathbf{X}, B\mathbf{Y} \rangle = (\mathbf{X}, \mathbf{Y}^T, B)$$

If the model errors are assumed to be Gaussian, Li and Navon (2001) have shown the additive properties of a 4D-Var, i.e. that solving the optimisation problem on the whole data assimilation period provide the same results as splitting our data assimilation period in two subspaces  $1..n_1$  and  $n_1..n$  and defining two costs functions as :

$$J_1(\mathbf{c}) = \frac{1}{2} \langle \mathbf{B}^{-1} \delta \mathbf{x}_0, \delta \mathbf{x}_0 \rangle + \frac{1}{2} \sum_{i=0}^{n_1} (\dots)$$

and

$$J_2(\mathbf{c}) = \frac{1}{2} \left\langle (\mathbf{P}_{\mathbf{x}_{n_1}}^a)^{-1} (\mathbf{x}_{n_1} - \hat{\mathbf{x}}_{n_1}), \mathbf{x}_{n_1} - \hat{\mathbf{x}}_{n_1} \right\rangle + \frac{1}{2} \sum_{i=n_1+1}^n (\dots)$$

where  $\mathbf{P}_{\mathbf{x}_{n_1}}^a$  is the analysis error covariance matrix of first data assimilation subsequence (which can be obtained by the computation of the Hessian matrix of  $J_1$  (Rabier and Courtier, 1992) and  $\hat{\mathbf{x}}_{n_1}$  is the analysed state (the result of this assimilation).

In order to compare optimal nudging with Kalman Filter, we assume that we have already performed the data assimilation on a former subspace of observations  $\{\mathbf{y}_1, \dots, \mathbf{y}_{k-1}\}$ , and so we know  $\mathbf{P}_{\mathbf{x}_{n_1}}^a$  and  $\hat{\mathbf{x}}_{n_1}$ .

We wish to minimize the following cost function :

$$\begin{aligned} J_k(\mathbf{c}_k) &= \frac{1}{2} \left\langle (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} (\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}), \mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1} \right\rangle \\ &+ \frac{1}{2} \left\langle \mathbf{R}^{-1} (\mathbf{H}_k(\mathbf{x}_k) - \mathbf{y}_k^o), \mathbf{H}_k(\mathbf{x}_k) - \mathbf{y}_k^o \right\rangle \\ &+ \left\langle \mathbf{Q}_k^{-1} \mathbf{G}_k \mathbf{d}_k, \mathbf{G}_k \mathbf{d}_k \right\rangle \end{aligned} \quad (20)$$

Using

$$\mathbf{x}_{k-1} = \mathbf{M}_{k,k-1}^{-1} (\mathbf{x}_k - \mathbf{G}_k \mathbf{d}_k)$$

The cost function could be rewritten as :

$$\begin{aligned} J_k &= \frac{1}{2} \left\langle (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} (\mathbf{M}_{k,k-1}^{-1} (\mathbf{x}_k - \mathbf{G}_k \mathbf{d}_k) - \hat{\mathbf{x}}_{k-1}), \mathbf{M}_{k,k-1}^{-1} (\mathbf{x}_k - \mathbf{G}_k \mathbf{d}_k) - \hat{\mathbf{x}}_{k-1} \right\rangle \\ &+ \frac{1}{2} \left\langle \mathbf{R}^{-1} \mathbf{H}_k(\mathbf{x}_k) - \mathbf{y}_k^o, \mathbf{H}_k(\mathbf{x}_k) - \mathbf{y}_k^o \right\rangle \\ &+ \frac{1}{2} \left\langle \mathbf{Q}_k^{-1} \mathbf{G}_k \mathbf{d}_k, \mathbf{G}_k \mathbf{d}_k \right\rangle \end{aligned} \quad (21)$$

Then our problem is to find  $(\hat{\mathbf{x}}_k, \hat{\mathbf{G}}_k)$  as

$$\begin{cases} \frac{\partial J_k}{\partial \mathbf{x}_k} = 0 \\ \frac{\partial J_k}{\partial \mathbf{G}_k} = 0 \end{cases} \quad (22)$$

Using (??), (??) becomes :

$$\begin{cases} 0 = \mathbf{H}_k^T \mathbf{R}^{-1} (\mathbf{H}_k(\hat{\mathbf{x}}_k) - \mathbf{y}_k^o) + \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} [\mathbf{M}_{k,k-1}^{-1} (\hat{\mathbf{x}}_k - \hat{\mathbf{G}}_k \mathbf{d}_k) - \hat{\mathbf{x}}_{k-1}] \\ 0 = \mathbf{Q}_k^{-1} \hat{\mathbf{G}}_k \mathbf{d}_k \mathbf{d}_k^T + \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} \mathbf{M}_{k,k-1}^{-1} \hat{\mathbf{G}}_k \mathbf{d}_k \mathbf{d}_k^T \\ + \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} [\hat{\mathbf{x}}_{k-1} - \mathbf{M}_{k,k-1}^{-1} \hat{\mathbf{x}}_k] \end{cases} \quad (23)$$

according to (??) we obtain

$$\hat{\mathbf{G}}_k \mathbf{d}_k = \mathcal{H}_{\mathbf{Q}_k}^{-1} \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} [\hat{\mathbf{x}}_{k-1} - \mathbf{M}_{k,k-1}^{-1} \hat{\mathbf{x}}_k] \quad (24)$$

where  $\mathcal{H}_{\mathbf{Q}_k} = \mathbf{Q}_k^{-1} + \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} \mathbf{M}_{k,k-1}^{-1}$

Using (??) into (??) leads to

$$\begin{aligned} 0 &= \mathbf{H}_k^T \mathbf{R}^{-1} (\mathbf{y}_k^o - \mathbf{H}_k (\hat{\mathbf{x}}_k)) \\ &\quad + \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} (I + \mathbf{M}_{k,k-1}^{-1} \mathcal{H}_{\mathbf{Q}_k}^{-1} \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1}) [\hat{\mathbf{x}}_{k-1} - \mathbf{M}_{k,k-1}^{-1} \hat{\mathbf{x}}_k] \end{aligned}$$

We can rewrite it as

$$\begin{aligned} &[\mathbf{H}_k^T \mathbf{R}^{-1} \mathbf{H}_k + \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} \mathbf{M}_{k,k-1}^{-1} \\ &\quad + \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} \mathbf{M}_{k,k-1}^{-1} \mathcal{H}_{\mathbf{Q}_k}^{-1} \mathbf{M}_{k,k-1}^{-T} (\mathbf{P}_{\mathbf{x}_{k-1}}^a)^{-1} \mathbf{M}_{k,k-1}^{-1}] [\mathbf{M}_{k,k-1} \hat{\mathbf{x}}_{k-1} - \hat{\mathbf{x}}_k] \\ &= \mathbf{H}_k^T \mathbf{R}^{-1} [\mathbf{M}_{k,k-1} \hat{\mathbf{x}}_{k-1} - \mathbf{y}_k^o] \end{aligned} \quad (25)$$

If we pose

$$\mathbf{P}_k^f = \mathbf{M}_{k,k-1} \mathbf{P}_{\mathbf{x}_{k-1}}^a \mathbf{M}_{k,k-1}^T + \mathbf{Q}_k \quad (26)$$

Using a matrix inversion lemma <sup>2</sup> :

$$\begin{aligned} \mathcal{H}_{\mathbf{Q}_k}^{-1} &= \mathbf{Q}_k - \mathbf{Q}_k (\mathbf{P}_k^f)^{-1} \mathbf{Q}_k \\ &= \mathbf{M}_{k,k-1} \mathbf{P}_{\mathbf{x}_{k-1}}^a \mathbf{M}_{k,k-1}^T (\mathbf{P}_k^f)^{-1} \mathbf{Q}_k \end{aligned} \quad (27)$$

Introducing (??) in (??) and (??)  $\times (\mathbf{P}_k^f)^{-1}$

$$[\mathbf{H}_k^T \mathbf{R}^{-1} \mathbf{H}_k + (\mathbf{P}_k^f)^{-1}] [\mathbf{M}_{k,k-1} \hat{\mathbf{x}}_{k-1} - \hat{\mathbf{x}}_k] = \mathbf{H}_k^T \mathbf{R}^{-1} [\mathbf{M}_{k,k-1} \hat{\mathbf{x}}_{k-1} - \mathbf{y}_k^o] \quad (28)$$

we can retrieve, using another matrix formula<sup>3</sup> :

$$\hat{\mathbf{x}}_k = \mathbf{M}_{k,k-1} \hat{\mathbf{x}}_{k-1} + \mathbf{P}_k^f \mathbf{H}_k^T [\mathbf{H}_k \mathbf{P}_k^f \mathbf{H}_k^T + \mathbf{R}]^{-1} [\mathbf{y}_k^o - \mathbf{M}_{k,k-1} \hat{\mathbf{x}}_{k-1}]$$

It is exactly the expression of  $\hat{\mathbf{x}}_k$  in the Kalman Filter. It shows that, in theory and when the model

is linear, the optimal nudging leads to the same results as the Kalman filter which is optimal.

<sup>2</sup>  $(\mathbf{A}^T \mathbf{B}^{-1} \mathbf{A} + \mathbf{C}^{-1})^{-1} = \mathbf{C} - \mathbf{C} \mathbf{A}^T (\mathbf{A} \mathbf{C} \mathbf{A}^T + \mathbf{B})^{-1} \mathbf{A} \mathbf{C}$

<sup>3</sup>  $(\mathbf{A}^{-1} + \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B})^{-1} \mathbf{B}^T \mathbf{C}^{-1} = \mathbf{A} \mathbf{B}^T (\mathbf{C} + \mathbf{B} \mathbf{A} \mathbf{B}^T)^{-1}$  whenever the inverses exist (see Wunsch , 1996 pp 99)

## B The PALM assimilation coupler

### References

- [1] Alekseev, K. Q. and Navon, I. M. 2001. The Analysis of an Ill Posed Problem Using Multiscale Resolution and Second Order Adjoint Techniques. *Computer methods on Applied Mechanics and Engineering*, Vol 190, 15-17, 1937-1953.
- [2] Adcroft, A. and Marshall, D. 1998. How slippery are piecewise-constant coastlines in numerical ocean models. *Tellus* 50A, 95-108.
- [3] Cohn, S.E. 1997. An Introduction to Estimation Theory *J. Meteor. Soc. Japan* 75, 1B, 257-288
- [4] Le Dimet, F.-X. and Charpentier, I. 1998. Méthodes du second ordre en assimilation de données in *Equations aux Dérivées Partielles et Applications*, Gauthier-Vilars, Paris, 107-125.
- [5] Courtier, P. 1997. Dual formulation of four dimensional variational assimilation. *Q. J. Roy. Meteorol. Soc.* 123, 2449-2462.
- [6] Derber, J. 1989. A variational continuous assimilation technique. *Mon. Wea. Rev.*, 117, 2437-2446.
- [7] Derber, J. and Bouttier, F. 1999. A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus* 51A, 195-221 .
- [8] Ide, K., Courtier, P., Ghill, M. and Lorenc, A. C. 1997. Unified notation for data assimilation: operational, sequential and variational. *J. Meteorol. Soc. Japan* 75-B, 181-189
- [9] Kalman, R. E. 1960. A New Approach to Linear Filtering and Prediction Problems,” Transaction of the *ASME–Journal of Basic Engineering*, 35-45 (March 1960).
- [10] Krishnamurti, T. N. , Jishan, X., Bedi, H. S., Ingles, K. and Oosterhof, D. 1991. Physical initialization for numerical weather prediction over the tropics. *Tellus* 43A, 53-81.
- [11] Gilbert, J.-C. and Lemarechal, C. 1989. Some numerical experiment with variable storage quasi-Newton algorithms. *Math. Prog.*, B25, 407-435.
- [12] Lellouche, J.-M., Devenon, J.-L. and Dekeyser, I. 1994. Boundary control of Burgers’ Equation-A Numerical Approach *Computers Math Applic.* 28, 33-44.

- [13] Li, Z. and Navon, I. M. 2001. Optimality of variational data assimilation and its relationship with the Kalman filter and smoother. *Q. J. Roy. Meteorol. Soc.*, 127 Part B, 661-684.
- [14] Lorenc, A. C. 1986. Analysis methods for numerical weather prediction. *Q. J. Roy. Meteorol. Soc.* 112, 1177-1194.
- [15] Lorenc, A. C., Bell, R. S. and Macpherson, B. 1991. The Meteorological Office analysis correction data assimilation scheme. *Q. J. Roy. Meteorol. Soc.* 117, 59-89.
- [16] Lyne, W. H., Swinbank, R. and Birch, N. T. 1982. A data assimilation experiment and the global circulation during the FGGE special observing periods. *Q. J. Roy. Meteorol. Soc.* 108, 575-594.
- [17] Jazwinski, A. H. 1970. *Stochastic processes and filtering theory*. Academic Press, New York.
- [18] Ménard, R. 1993. Kalman Filtering of Burger's Equation and its Application to Atmospheric Data Assimilation. PhD thesis, McGill University, Montréal, Canada.
- [19] Nichols, N. K. and Griffith, A. K. 2000: Adjoint Methods in Data Assimilation for estimating Model Error. *Journal of Flow, turbulence and combustion*, 65,
- [20] Piacentini, A. and the PALM group, 1999 PALM : A modular data assimilation system. *Proceedings of the third WMO symposium on data assimilation, 7-11 june 1999, Québec City, Canada*, WMO/TD-986, 321-323.
- [21] Rabier, F. and Courtier, P. 1992. Four-Dimensional Assimilation in presence of baroclinic instability. *Q. J. R. Meteorol. Soc.* 118, 649-672
- [22] Thepaut, J. N. and Moll, P., 1990. Variational Inversion of simulated TOVS radiances using the adjoint technique. *Quart. J. Roy. Meteorol. Soc.*, 116, 1425-1448.
- [23] Tikhonov, A. V. and Arserin, V. 1997. Solution of Ill posed Problems. *Winston & Sons*, Washington DC. pp. 224
- [24] Vidard, P. A., Blayo, E., Le Dimet, F.-X. and Piacentini, A. 2000. 4D-variational data analysis with imperfect model. reduction of the size of control. *Journal of Flow, turbulence and combustion*, 65, 489-504.

- [25] Wang, Z., Navon, I. M., Le Dimet, F.-X. and Zou, X. 1992. The second order analysis: Theory and applications. *Meteorol. Atmos. Phys*, 50, 3-21.
- [26] Weaver, A., Courtier, P. 2001 Correlation modeling on the sphere using a generalized diffusion equation. *Q. J. Roy. Meteorol. Soc.*, 127, 1815-1846.
- [27] Wunsch, C. 1996. *The Ocean Circulation Inverse Problem*, (Cambridge University Press). pp 99
- [28] Yang, W., Navon, I. M. and Courtier, P. 1996. A New Hessian Preconditioning Method Applied to Variational Data Assimilation Experiment using Adiabatic Version of NASA/GEOS-1G CM. *Monthly Weather Review*, vol. 124, No 5, 1000-1017.
- [29] Zou, X., Navon, I. M. and Le Dimet, F.-X. 1992. An optimal Nudging Data Assimilation Scheme using Parameter Estimation, *Q. J. Roy. Meteorol. Soc.*, 118, 1163-1186

Table 1: norm of the error and CPU time for both raw and smoothed methods according to the form of  $\mathbf{G}$

Form of $\mathbf{G}$	Size of control	Raw		Interpolated	
		RMS	CPU	RMS	CPU
$\mathbf{G}$ is a full matrix	$E \times M \times N$	45.06	57.80s	38.49	57.69s
$\mathbf{G}$ is a diagonal matrix	$M \times N$	65.30	6.90s	17.57	7.43s
$\mathbf{G}$ is a scalar coefficient	$N$	64.23	9.74s	17.33	3.63s
4D-Var	$E$	73.73	5.82s		
Without assimilation		248.21			



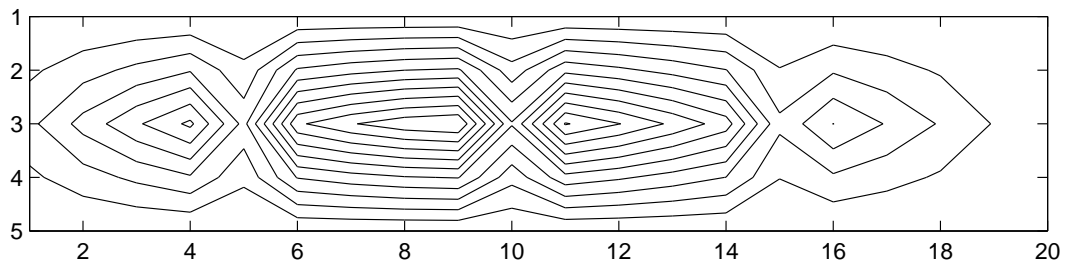


Figure 1: Transpose of the estimated  $\mathbf{G}$  as a full matrix

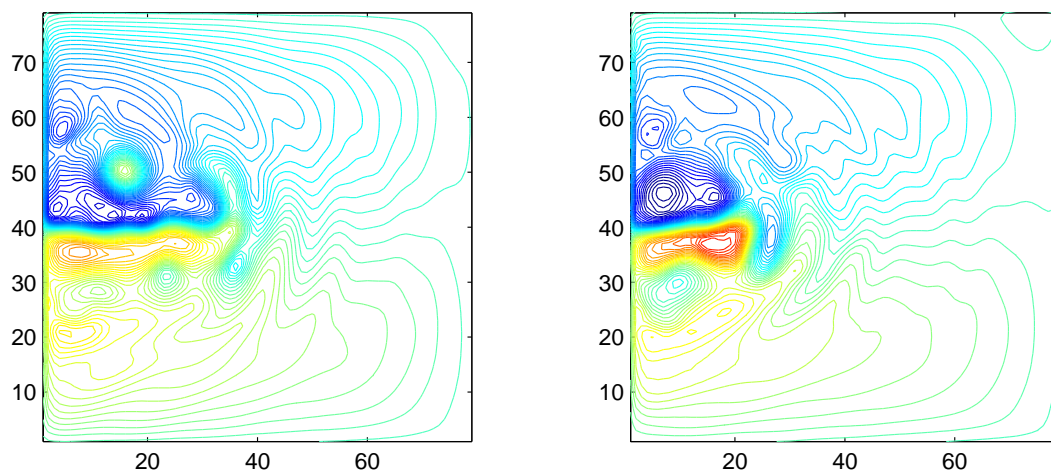


Figure 2: SSH for forecast model (left) and true state (right) at initial time step

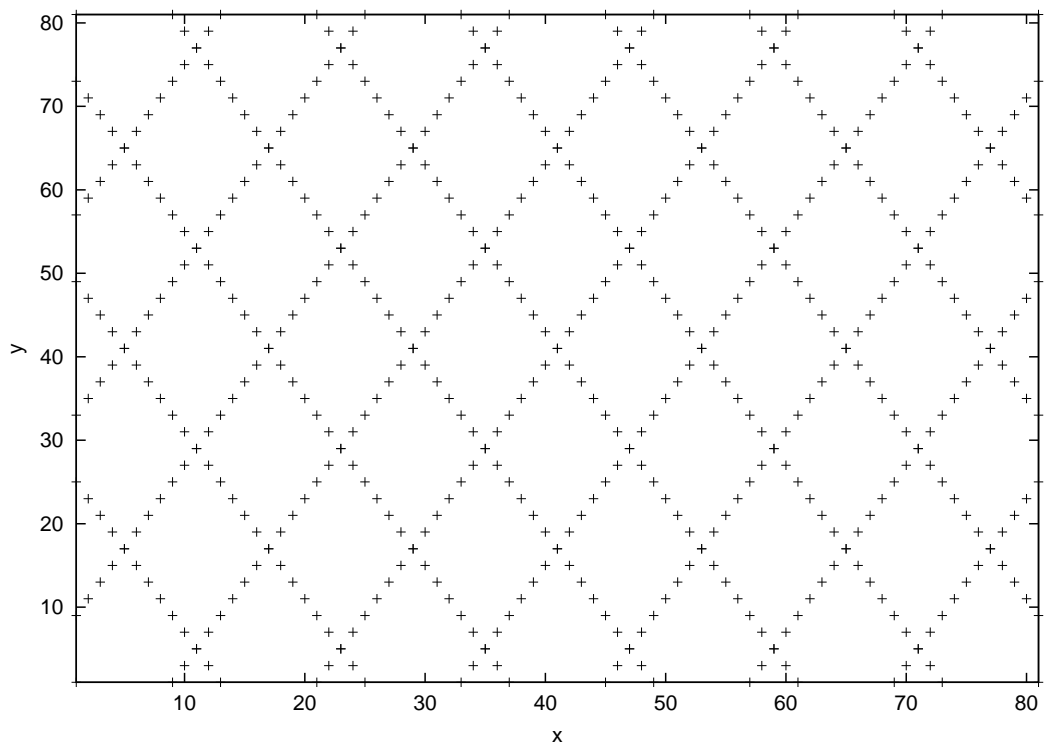


Figure 3: Location of the observations (ground track of satellites)

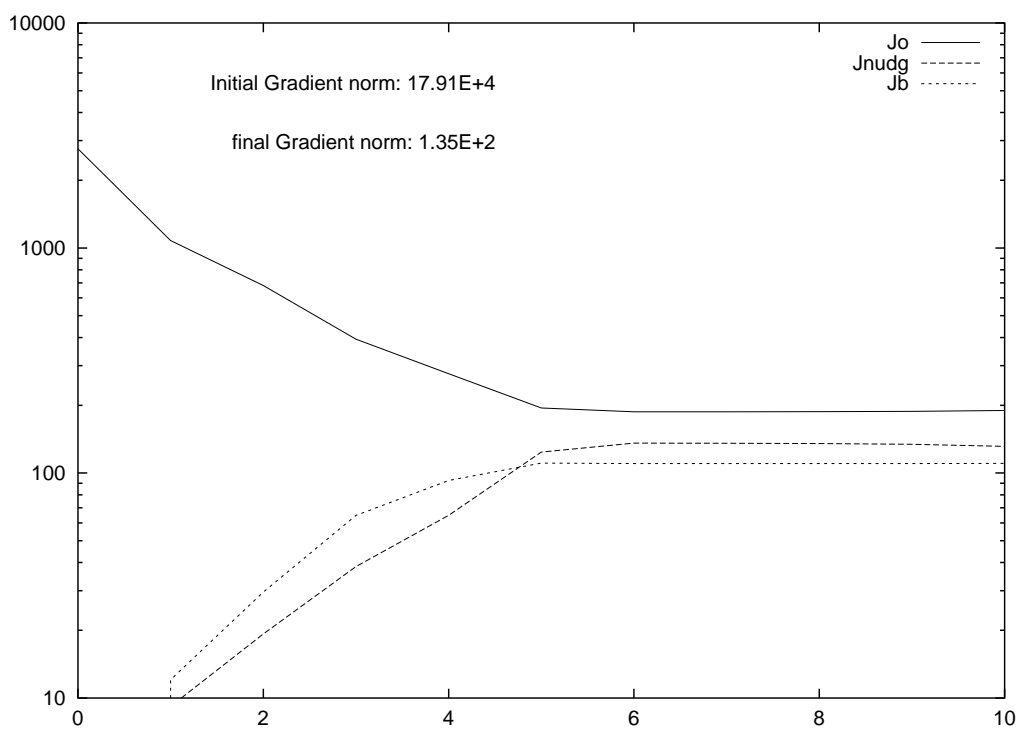


Figure 4: Evolution of the value of the different terms of the cost function

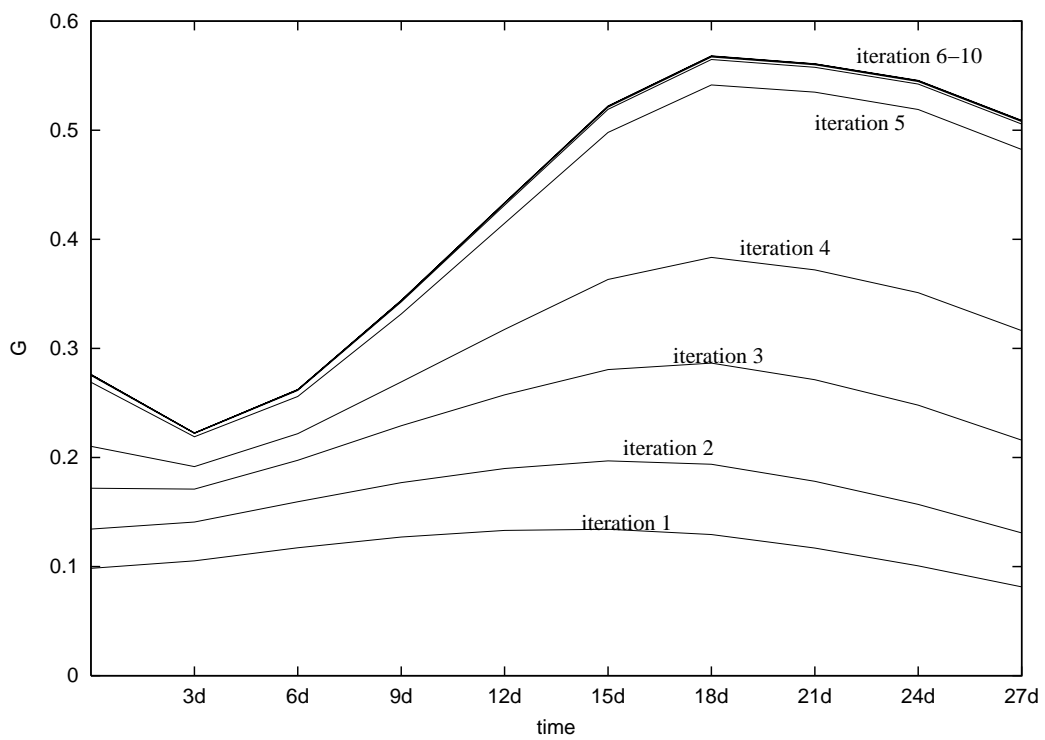


Figure 5: Evolutions of the nudging coefficient according to the number of iterations

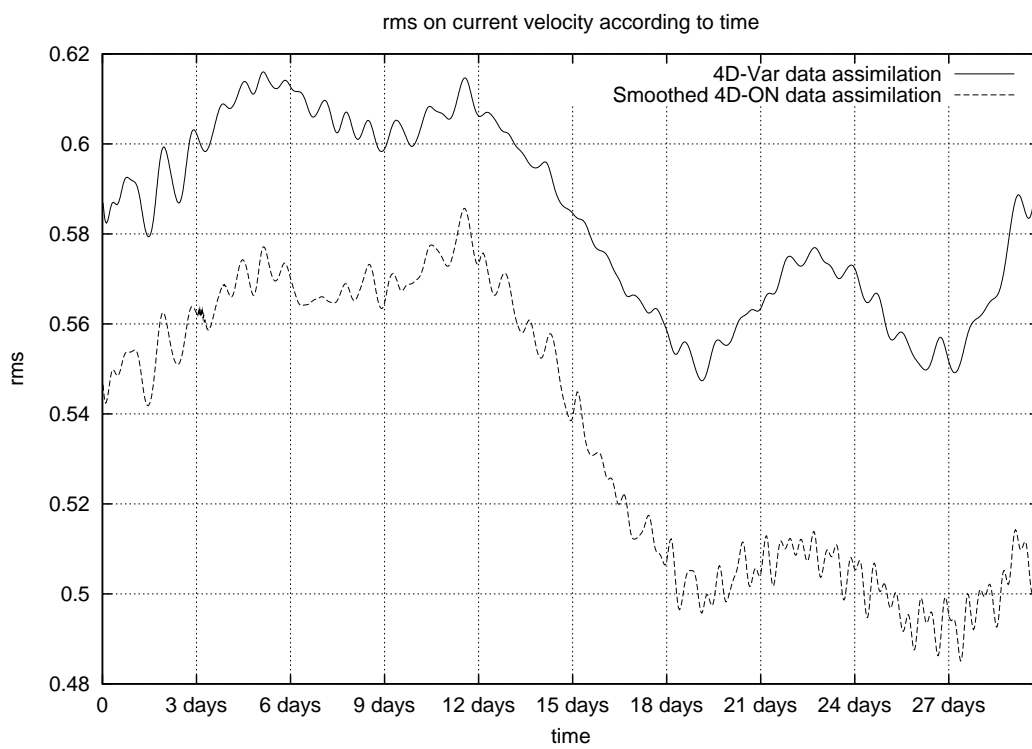


Figure 6: RMS of the error on the current velocity for smoothed version of the 4D-ON compared to 4D-Var

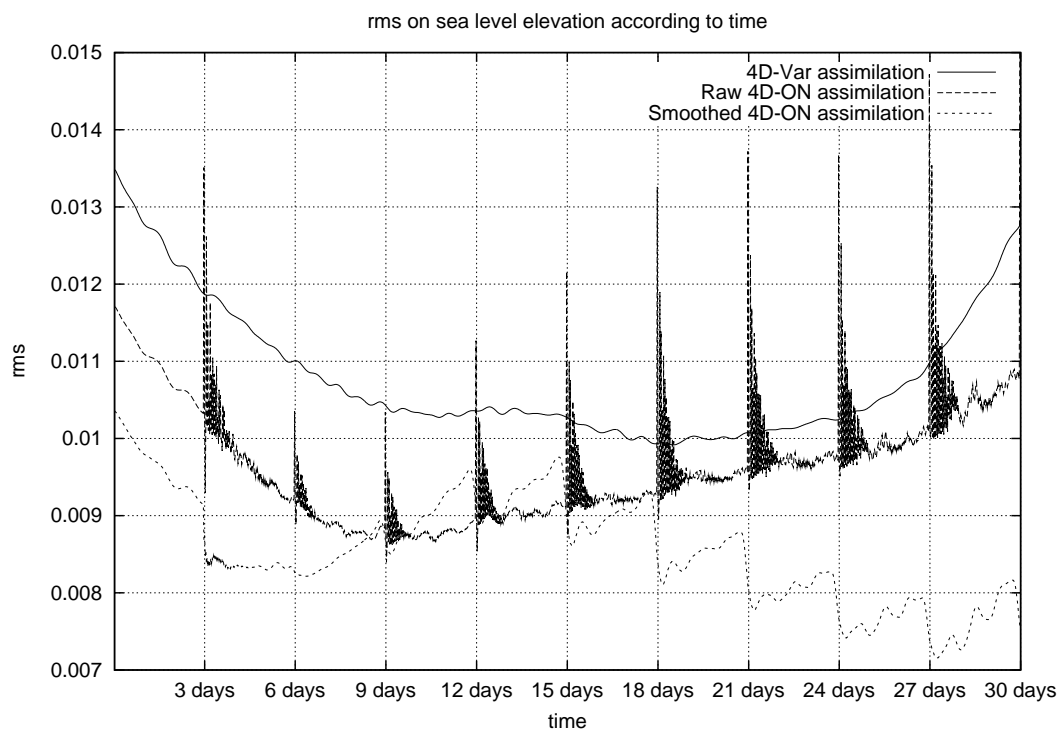


Figure 7: RMS of the error of SSH for smoothed version of the 4D-ON compared to 4D-Var and raw 4D-ON



Figure 8: RMS of the error on current velocity for smoothed version of the 4D-ON compared to 4D-Var on assimilation window and one month forecast



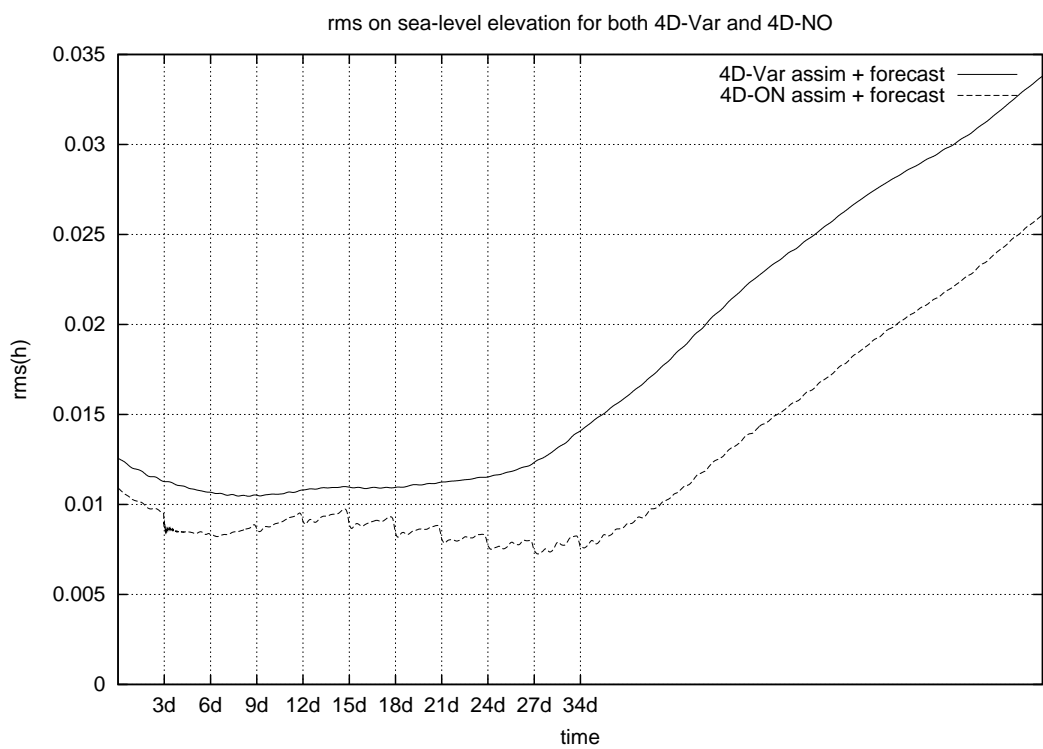


Figure 9: RMS of SSH for smoothed version of the 4D-ON compared to 4D-Var on assimilation window and one month forecast