*Tellus (2004), 56A, 177–188*
*Printed in UK. All rights reserved*


**TELLUS**

# Variational data analysis with control of the forecast bias

*By* P. A. VIDARD[1]*, A. PIACENTINI[1] and F.-X. LE DIMET[2], [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*

(Manuscript received 3 March 2003; in final form 11 December 2003)

ABSTRACT

We propose a methodology for the treatment of the systematic model error in variational data assimilation. The principle of the method is to add a systematic error correction term in the model equations and to include it in the variational assimilation control vector.

This method is applied to a simplified ocean circulation model in an identical twin experiment framework. It shows a noticeable improvement compared to the result of a classical variational assimilation scheme in which the systematic error is not corrected. The estimated systematic error correction term is sufficiently consistent with that needed by the model that it allows improvements not just to the analysis, but also during the forecast phase.

## 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 best estimation of the 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 methods (i.e. remote sensing, buoys, tomography) and computing resources have permitted a wide improvement in data assimilation methods and their efficiency.

The current operational method in many meteorological centres is variational data assimilation (VDA; see Sasaki, 1958; Le Dimet and Talagrand, 1986). It consists of assimilating all the observational and model information from the previous time sequence. The problem is formulated as an optimal control problem which criterion measures the misfit between the model predictions and the observations of the system state. One of the main assumptions of the VDA scheme used in operational centres is that the model describes exactly the system behaviour. However, in practice the model equations do not represent the exact evolution of the system and model errors arise due to the lack of resolution, inaccurate representation of small-scale physics or errors in boundary conditions, topography or forcing terms.

A likely VDA improvement in an operational application will be to consider that the model is not exact, i.e. for example, to introduce a model error correction term in the control vector (see Jazwinski, 1970; Derber, 1989; Cohn, 1997). The principle of the complete method is to add to the control vector a residual error correction term, which is added to the model equation at every time step. Because of 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. There have been recent attempts to reduce the size of the model error part of the control vector in the above formulation, by controlling the correction term only in certain privileged directions (see, for instance, Durbiano, 2001; Vidard et al. 2001) or by using information provided by the analysis residual vector (Zou et al. 1992; Vidard et al. 2003) or by controlling only the systematic and time correlated part of the error (Bell et al. 2001; D'Andrea and Vautard, 2001; Griffith and Nichols, 2001). Additionally, Dee and Da Silva (1998) pointed out the importance of considering the random and the systematic part of the model error separately, otherwise the reduction of one could lead to an increase in the other. Vidard et al. (2003) presented a low-cost method for dealing with the random model error in VDA. In the present paper, we present a VDA with control of the systematic error scheme and we show a first application in the context of the assimilation of sea surface height (SSH) observations in a (two-dimensional non-linear) shallow-water model. Twin experiments are performed, in one case using a 'true' reference model

*Corresponding author. Now at European Centre for Medium-Range Weather Forecast, Shinfield Park, Reading, Berkshire, RG2 9AX, United Kingdom.
e-mail: nes@ecmwf.int

and in the other a 'false' model using perturbed dissipation and forcing terms.

The paper is organized as follows. The classical formulation of VDA is presented in Section 2. In Section 3 we present the modified assimilation scheme used in this paper. The model and the assimilation statistics used in the experiments are discussed in the first part of Section 4, and in the second part of this section we present the results of the experiment. Finally, we give a summary and concluding discussion in Section 5.

## 2. Variational data assimilation, classical formulation

The behaviour of the system is modelled by a set of equations (a spatial discretization is assumed)

$$\begin{cases} \mathbf{x}(t_0) = \mathbf{x}^{\mathrm{b}} \\ \dfrac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = M(\mathbf{x}(t)) \end{cases} \tag{1}$$

where $\mathbf{x}^{\mathrm{b}}$ is the first guess initial condition, $\mathbf{x}(t) \in \mathcal{X} \subset \mathbb{R}^n$ is the model state at time $t$, and $M$: $\mathbb{R}^n \to \mathbb{R}^n$ is a non-linear model describing the time evolution of the state.

An important part of the information in any data assimilation scheme comes from the data used. Let us define the observed values at time $t_i$, $i = 1, \ldots, I$ as gathered in a time-dependent observation vector $\mathbf{y}^{\mathrm{o}}_{t_i} \in \mathcal{Y}_{t_i} \subset \mathbb{R}^{m_{t_i}}$. In order to be able to compare them with the state vector, let us introduce the so-called observation operator $H_{t_i}$: $\mathcal{X} \to \mathcal{Y}_{t_i}$. The starting point of any assimilation procedure is the measure of the discrepancies $\mathbf{d}$ between observation and the state vector $\mathbf{d}_{t_i} = \mathbf{y}_{t_i} - H_{t_i}(\mathbf{x}(t_i))$, where $\mathbf{d}_{t_i}$ is called the 'innovation' or 'analysis residual vector' whether it is computed before or after the analysis. In VDA, the innovation vector is used to control both the initial condition $\mathbf{x}_0 = \mathbf{x}^{\mathrm{b}} + \varepsilon$ and model error correction term $\eta(t) \in \mathcal{N} \subset \mathbb{R}^n$ which act in the model integration as follows:

$$\begin{cases} \mathbf{x}(t_0) = \mathbf{x}_0 \\ \dfrac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = M(\mathbf{x}(t)) + \eta(t). \end{cases} \tag{2}$$

This control is done through the minimization of the cost function defined by

$$\begin{aligned} J(\mathbf{x}_0, \eta) = {} & \overbrace{\frac{1}{2}\sum_{i=1}^{I} \left\langle \mathbf{R}^{-1}[H(\mathbf{x}(t_i)) - \mathbf{y}_{t_i}],\, H(\mathbf{x}(t_i)) - \mathbf{y}_{t_i} \right\rangle}^{J_{\mathrm{o}}} \\ & + \overbrace{\frac{1}{2}\int_{t_0}^{T} \langle \mathbf{Q}^{-1}(t)\eta(t), \eta(t)\rangle \mathrm{d}t}^{J_\eta} \\ & + \overbrace{\frac{1}{2}\left\langle \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^{\mathrm{b}}), \mathbf{x}_0 - \mathbf{x}^{\mathrm{b}} \right\rangle}^{J_{\mathrm{b}}} \end{aligned} \tag{3}$$

where $\mathbf{R}$, $\mathbf{Q}$ and $\mathbf{B}$ are the observation, model error and initial condition (background) error covariance matrices, respectively. These three types of error are assumed to be uncorrelated. This assumption is debatable because the background state usually comes from previous assimilation using the same model and the same type of data, but is commonly used. The goal of the assimilation is to reduce the discrepancy between the model state and the observation ($J_{\mathrm{o}}$) but it should not be done at any cost. The presence of both $J_{\mathrm{b}}$ and $J_\eta$ terms prevents the control variables from being too large and introduce some more information about the structural correlation of the respective control variable. This addition of information is done through the inverses of the error covariance matrices. The balance between the three terms is also done due to the $\mathbf{R}^{-1}$, $\mathbf{Q}^{-1}$ and $\mathbf{B}^{-1}$ matrices.

In meteorological and oceanographic applications, the number of degrees of freedom is too large to allow the use of the method as described above. Additional assumptions are needed, mainly for the model error correction term. Assuming the problem is discretized in time, the size of the control vector is the size of the state vector times the number of time-steps (typically $10^8$–$10^9$). In all the existing operational variational assimilation suites, the model is assumed to be perfect (free of error), so avoiding the control of the $\boldsymbol{\eta}$ terms, the $J_\eta$ term is set to 0 and the size of control becomes the size of the state vector.

The scheme described above takes account of the time dimension and is commonly called 4D-Var (spatial plus temporal dimensions). There is a sequential version, 3D-Var, for which the model $M$ is replaced by the identity operator and the assimilation window is shortened. Model errors are less of an issue in 3D-Var because the assimilation is performed more frequently but they should be addressed through background error statistics. The method proposed in the following is more applicable to 4D-var because model errors are damaging for this scheme. Because the model used as an example is only two dimensional, we made the choice not to use 3D-Var and 4D-Var names in this paper.

## 3. Control of systematic errors: mathematical formulation

As mentioned previously, the number of degrees of freedom in realistic oceanographic or meteorological applications is too large to allow the application of the method introduced in the previous section. A drastic reduction of the size of the control vector is needed. Moreover, according to Dee and Da Silva (1998), systematic and random parts of the error should be accounted for separately.

Following Griffith and Nichols (2001), we assume that the time evolution of the model error can be modelled as a function of the state vector and the initial bias as follows

$$\begin{cases} \eta(t_0) = \eta_0 \\ \dfrac{\mathrm{d}\eta}{\mathrm{d}t} = \Phi(\eta(t), \mathbf{x}(t)) + \varepsilon(t), \end{cases} \tag{4}$$

with $\varepsilon(t)$ being stochastic, unbiased, serially uncorrelated, normally distributed random vectors. In order to reduce the size of

control in the following we assume that this stochastic process can be neglected. If the random part of the model error is not negligible, ensemble prediction methods (see, for instance, Molteni et al. 1996) should be able to cope with it. In the following, we are only interested in the deterministic part of the model error. This deterministic evolution is given by $\Phi : \mathbb{R}^l \times \mathbb{R}^n \to \mathbb{R}^l$. Equation (2) can then be rewritten as

$$\begin{cases} \mathbf{x}(t_0) = \mathbf{x}_0 \\ \dfrac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = M(\mathbf{x}(t)) + \mathbf{T}(\eta(t)) \end{cases} \tag{5}$$

where $\mathbf{x}(t) \in \mathcal{X} \subset \mathbb{R}^n$, and $\eta(t) \in \mathcal{N} \subset \mathbb{R}^l$, $\mathbf{T} : \mathbb{R}^l \to \mathbb{R}^n$ is an operator that projects the model error term on to the model space. Usually $l = n$ and $\mathbf{T}$ is the identity function but, in some cases, the error may be expressed in a smaller space than the model state (for example, it could be negligible on some component of the state vector or in some location of the domain).

Now, we assume $\mathbf{\Phi}$ and $\mathbf{T}$ known, and the control vector consists of both initial condition $\mathbf{x}_0$ and initial model error $\eta_0$

$$\begin{aligned} J(\mathbf{x}_0, \eta_0) = {} & \frac{1}{2} \sum_{i=1}^{I} \left\langle \mathbf{R}_{t_i}^{-1} \left[ H_{t_i}(\mathbf{x}(t_i)) - \mathbf{y}_{t_i} \right], H_{t_i}(\mathbf{x}(t_i)) - \mathbf{y}_{t_i} \right\rangle \\ & + \frac{1}{2} \left\langle \mathbf{Q}_0^{-1}(\eta_0 - \eta^{\mathrm{b}}), \eta_0 - \eta^{\mathrm{b}} \right\rangle \\ & + \frac{1}{2} \left\langle \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^{\mathrm{b}}), \mathbf{x}_0 - \mathbf{x}^{\mathrm{b}} \right\rangle \end{aligned} \tag{6}$$

where $(\mathbf{x}^{\mathrm{b}}, \eta^{\mathrm{b}})$ is the first guess of the optimization problem.

This constrained minimization problem is first reduced to an unconstrained problem using a method of Lagrange. The Lagrange function is defined by

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \eta, \mathbf{x}^*, \eta^*) = {} & J(\mathbf{x}_0, \eta_0) \\ & + \int_{t_0}^{T} \left\langle \mathbf{x}^*, \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} - M(\mathbf{x}(t)) - \mathbf{T}(\eta(t)) \right\rangle \mathrm{d}t \\ & + \int_{t_0}^{T} \left\langle \eta^*, \frac{\mathrm{d}\eta}{\mathrm{d}t} - \Phi(\eta(t), \mathbf{x}(t)) \right\rangle \mathrm{d}t \end{aligned} \tag{7}$$

where $\mathbf{x}^*$ and $\boldsymbol{\eta}^*$ are the Lagrange multiplier vectors. The Lagrange multipliers are not specified but computed in determining the best fit. We can easily compute the partial derivative of the Lagrange function with respect to $\mathbf{x}$, $\boldsymbol{\eta}$, $\mathbf{x}^*$ and $\boldsymbol{\eta}^*$ (see Appendix A). At the minimum point, the gradient must be zero.

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = 0 \qquad \frac{\partial \mathcal{L}}{\partial \eta} = 0 \qquad \frac{\partial \mathcal{L}}{\partial \mathbf{x}^*} = 0 \qquad \frac{\partial \mathcal{L}}{\partial \eta^*} = 0. \tag{8}$$

The vector $(\mathbf{x}_0, \eta_0, \mathbf{x}^*, \eta^*)$ that satisfies eq. (8) is called a stationary point of $\mathcal{L}$. An important relation between the gradient of the cost function with respect to the control vectors $\mathbf{x}_0$ and $\eta_0$ and the partial derivative of the Lagrange function is

$$\begin{aligned} \nabla_{\mathbf{x}_0} J(\mathbf{x}_0, \eta_0) &= \left. \frac{\partial \mathcal{L}}{\partial \mathbf{x}} \right|_{\text{station.pt}} \\ \nabla_{\eta_0} J(\mathbf{x}_0, \eta_0) &= \left. \frac{\partial \mathcal{L}}{\partial \eta} \right|_{\text{station.pt}} . \end{aligned} \tag{9}$$

That is, the gradient of the cost function can be obtained by integrating forward the direct models (2) and (4), and then integrating backward in time the adjoint models (10) and (11):

$$\begin{cases} \mathbf{x}^*(T) = 0 \\ -\dfrac{\mathrm{d}\mathbf{x}^*}{\mathrm{d}t} = \left[ \dfrac{\partial \mathbf{M}}{\partial \mathbf{x}} \right]^T \mathbf{x}^* + \left[ \dfrac{\partial \mathbf{\Phi}}{\partial \mathbf{x}} \right]^T \eta^* - \delta(t - t_i) \displaystyle\sum_{i=1}^{I} \\ \qquad \times \mathbf{H}_{t_i}^T \mathbf{R}_{t_i}^{-1} \left[ H_{t_i}(\mathbf{x}(t_i)) - \mathbf{y}_{t_i} \right] \end{cases} \tag{10}$$

$$\begin{cases} \eta^*(T) = 0 \\ -\dfrac{\mathrm{d}\eta^*}{\mathrm{d}t} = \left[ \dfrac{\partial \mathbf{\Phi}}{\partial \eta} \right]^T \eta^* + \mathbf{T}^T \mathbf{x}^* \end{cases} \tag{11}$$

and

$$\begin{aligned} \nabla_{\mathbf{x}_0} J &= \mathbf{B}^{-1} \left( \mathbf{x}_0 - \mathbf{x}^{\mathrm{b}} \right) + \mathbf{x}^*(t_0) \\ \nabla_{\eta_0} J &= \mathbf{Q}_0^{-1} \left( \eta_0 - \eta^{\mathrm{b}} \right) + \eta^*(t_0). \end{aligned} \tag{12}$$

The size of the control vector is only doubled compared to the classical 'exact model' assumption scheme. The extra cost induced by the control of the model errors depends directly on the complexity of the error model $\mathbf{\Phi}$ (Derber, 1989). Bell et al. (2001) propose a more physical approach whereby $\mathbf{\Phi}$ and $\mathbf{T}$ are defined to maintain some physical balances. In the following experiment, $\Phi(\mathbf{x}, \eta) = \eta$ and $\mathbf{T}$ is assumed to be the identity operator.

## 4. Experiment with a two-dimensional non-linear model

### *4.1. Experiment framework*

*4.1.1. Model.* For this test experiment, we use a non-linear one-layer two-dimensional shallow-water model on a square basin with flat bottom (such a model is often 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.

The framework of this experiment is a twin experiment one. That is to say, observations do not come from reality but from a version of the model which is slightly different to the model used in assimilation. The first one, used to represent the reality, provides a simulated 'true' state evolution and then, using the observation operator (and possibly a white noise), synthetic observations can be obtained. The second one represents the model (called the forecast model in the following, as opposed 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

$$\mathbf{x}_{\text{true/forecast}} = \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \\ \mathbf{h} \end{bmatrix}$$

are described by

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

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

(i) wind : $\tau_{\text{forecast}} = \tau_0 \frac{\sin(2\pi(y-\frac{L}{4}))}{L}\vec{i}$, and $\tau_{\text{true}} = \tau_{\text{forecast}} \times [1 + 0.8 \times \sin(2\pi t/\Delta t \times 480)]$ where $L$ is the basin length and the boundary conditions are

(ii) $\mathbf{v} = 0$ on north/south boundaries and $\mathbf{u} = 0$ on east-west boundaries;

(iii) no-slip boundary conditions.

In this experiment the numerical values are $L = 2000$ km, $f_0 = 0.7 \times 10^{-4}$ s$^{-1}$, $\beta = 2 \times 10^{-11}$ m$^{-1}$ s$^{-1}$, $\nu = 15$ m$^2$ s$^{-1}$ (forecast), $\nu = 0.9 \times 15$ m$^2$ s$^{-1}$ (true), $r = 10^{-7}$ s$^{-1}$ (forecast), $r = 0.9 \times 10^{-7}$ s$^{-1}$ (true), $\rho_0 = 10^3$ kg m$^{-3}$, $g^* = 0.02$ m s$^{-2}$, $\tau_0 = 0.015$ N m$^{-2}$, and the Coriolis factor $f = f_0 + \beta\, y$.

For the spatial discretization, a second-order centred scheme is used on an Arakawa C-grid with $\Delta x = \Delta y = 25$ km, and for the time discretization a leap-frog scheme with a time-step $\Delta t = 30$ mn and an Asselin time filter (coefficient 0.02) is added in the equations.

The initial condition of our data assimilation window is provided using a 6-yr spin-up of the forecast model whereas true initial state and observations are computed with a true model spin-up. In this section the results of data assimilation using the scheme introduced below are compared to the 'exact model assumption scheme' that is the most common operational variational algorithm. These schemes are called respectively variational data assimilation initial condition and systematic error (VDA-ICSE) and variational data assimilation initial condition (VDA-IC ).

Because of the use of two (slightly) different models (the 'true' model and the forecast model) we can simulate the presence of error in the forecast model. An estimation of this error can be computed integrating the forecast model starting from the true initial condition and relaxing toward the true trajectory at each time-step (see Fig. 1). The result of this computation is shown in Fig. 2.

*4.1.2. Assimilation.* The assimilation is carried out for 30 d. This time window represents a relevant time-scale of the ocean

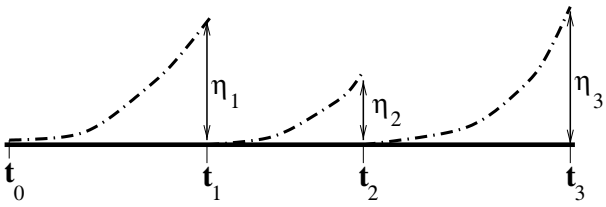



*Fig 1.* Computation of the error by integration of the forecast model (dashed line) starting from the 'true' trajectory (plain line).

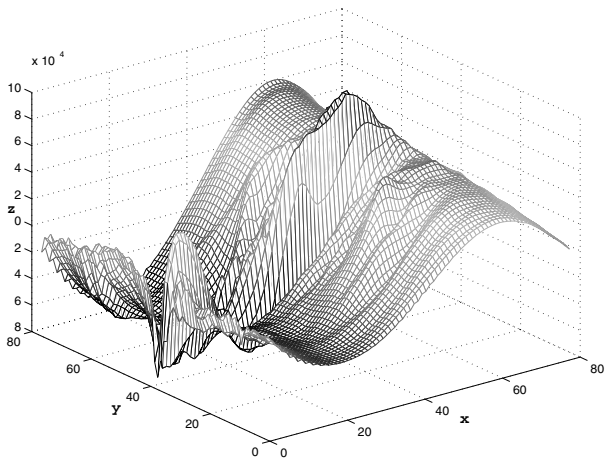



*Fig 2.* Systematic part of the model error on the SSH component on the whole domain.

model. Because of the large number of degrees of freedom, the optimization is not likely to converge in a number of iterations that makes the assimilation feasible. The number of iterations is therefore limited to 100 for the first experiment (high observation coverage) and 30 for the second (low coverage). In both cases, this choice represents the optimal number of iterations for the VDA-IC to achieve best results. In the fist experiment, 100 iterations are enough for VDA-IC to reach convergence because of the simplicity of the model. The results of VDA-ICSE can be slightly improved by increasing the number of iterations, but we chose to keep the same number of iterations, as we wanted to show the improvement at constant cost. If the observation coverage is sparse in space and time (second experiment), reaching the convergence is not suitable. Indeed, usually a VDA scheme resolves first the large scales and then goes down to small scales as the minimization proceeds (see Vidard, 2001). Therefore, after a certain number of iterations (about 30 in this case), both VDA-IC and VDA-ICSE start to act on scales that are not represented by the observations. The minimization method used in the following experiment is M1QN3 quasi-Newton with limited memory developed at INRIA by Gilbert and Lemaréchal (1989) and based on Nocedal (1980).

Because of the great discrepancy between the dimension of different components of the state vector ($\mathbf{u}$ and $\mathbf{v}$ are about 1 m s$^{-1}$ and $\mathbf{h}$ is about 500 m) the minimization problem is very ill-conditioned. To avoid this, a preconditioning is performed using

the square root of the inverse of the background and model error covariance matrices (in the $\mathbf{B} = \mathbf{B}^{1/2}\,\mathbf{B}^{T/2}$ sense). The new control vector is

$$\mathbf{v} = \begin{bmatrix} \mathbf{B}^{-(1/2)} & 0 \\ 0 & \mathbf{Q}^{-(1/2)} \end{bmatrix} \begin{bmatrix} \mathbf{x}_0 - \mathbf{x}^{\mathrm{b}} \\ \eta_0 - \eta^{\mathrm{b}} \end{bmatrix}. \tag{14}$$

$\mathbf{v}$ is a non-dimensional vector, and minimizing eq. (6) becomes the equivalent problem to minimizing

$$J(\mathbf{v}) = \underbrace{\frac{1}{2}\sum_{i=0}^{n} \left(H_{t_i}(\mathbf{x}(t_i)) - \mathbf{y}_i^{\mathrm{o}}\right)^T \mathbf{R}_i^{-1} \left(H_{t_i}(\mathbf{x}(t_i)) - \mathbf{y}_i^{\mathrm{o}}\right)}_{J_{\mathrm{o}}} + \overbrace{\mathbf{v}^T\mathbf{v}}^{J_{\mathrm{b}}+J_\eta} \tag{15}$$

with the gradient

$$\nabla_{\mathbf{v}} J = \mathbf{v} + \begin{bmatrix} \mathbf{B}^{T/2} & 0 \\ 0 & \mathbf{Q}^{T/2} \end{bmatrix} \begin{bmatrix} \nabla_{\mathbf{x}_0} J \\ \nabla_{\eta_0} J \end{bmatrix}. \tag{16}$$

In addition to non-dimensionalizing the control vector, because of this change of variable the eigenvalue of the Hessian matrix is at least 1 and the ratio between the highest and the lowest eigenvalue is reduced (Courtier, 1997). Moreover, as it is actually the inverse of eq. (14) that is performed at the beginning of each minimization iteration, it is the square roots of the covariance matrices and their adjoint that are used and not their inverses. Because of the lack of knowledge on the errors and the size of the problem, the determination on $\mathbf{B}$ and $\mathbf{Q}$ is not obvious. In this paper, both $\mathbf{B}$ and $\mathbf{Q}$ are multivariate operators that are constructed as described in Appendix B. The observation error covariance matrix $\mathbf{R}$ is assumed to be diagonal (error uncorrelated in space) and constant in time. In a twin experiment where the observation take place exactly at model grid points (no representativeness error) and the measurement error is simulated by white noise, this hypothesis is appropriate. In a realistic case, representativeness errors have to be taken into account; moreover, the measurement errors are likely to be correlated in space and time.

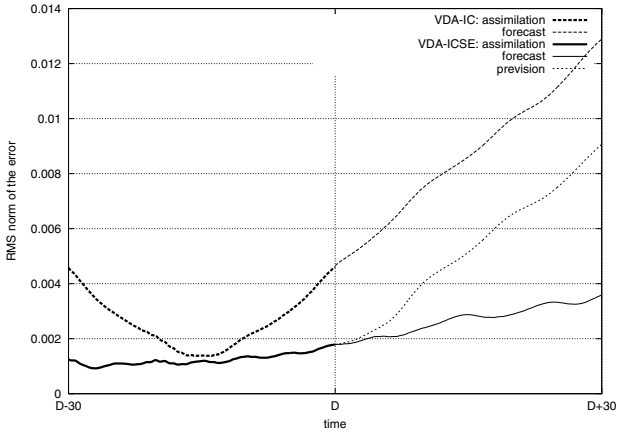
### 4.2. Results

*4.2.1. Assimilation of sea surface height maps.* First, every three days, the whole field of SSH derived from the 'true' model is taken to serve as observations for the assimilation experiment. In this case, as the number of observations is larger than the number of degrees of freedom, the minimization problem is highly overestimated. Because of this assimilation we obtain an optimal (or analysed) trajectory on the whole one month time window. The analysed state at the end of the assimilation period is subsequently used as an initial condition for a one month forecast (during this period the forcing is assumed known). Figure 3 shows the time evolution of the root-mean-square (rms) error of each component of the analysed state vector ($\mathbf{h}$, $\mathbf{u}$ and $\mathbf{v}$) on the whole domain, for the assimilation period (thick lines from D-30 day to D) and for the forecast period (from D day onward).

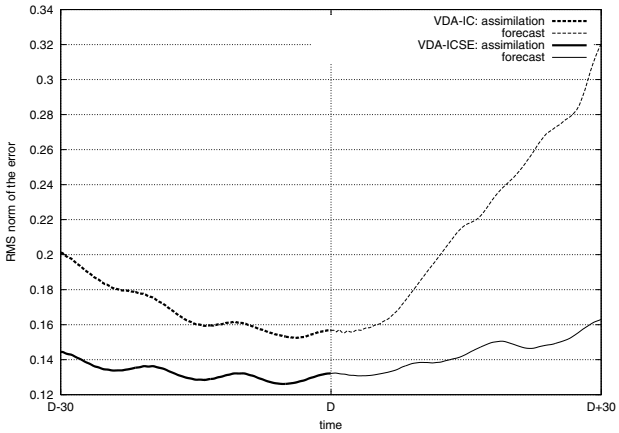
We can recognize the typical u-shaped (or j-shaped) error curve of the control of initial conditions only (thick dashed line) on the observed component in the presence of the model error. Indeed, in this case the VDA-IC method has to introduce errors in the initial field in order to obtain a trajectory closer to the observations on the whole assimilation time window. This phenomenon is illustrated by Fig. 4a which represents the error on the initial conditions increment. This is the difference between the ideal correction to the background initial conditions in order to reconstruct the 'true' initial conditions ($\delta\mathbf{x}^{\mathrm{true}}$) and the analysed increment provided by the VDA-IC ($\delta\mathbf{x}_{\mathrm{IC}}^{\mathrm{opt}}$): $\delta\mathbf{x}^{\mathrm{true}} - \delta\mathbf{x}_{\mathrm{IC}}^{\mathrm{opt}}$. Therefore, this field represents the error introduced by the VDA-IC on the initial conditions. Similarities with the systematic model error shown in Fig. 2 are obvious, as we can recognize the 'two-phase' shape: negative in one part of the domain, positive in the other. Because the initial condition is the only controlled field in VDA-IC, in order to correct the effect of the model error on the whole trajectory, this scheme has to introduce the same type of error in the initial field. Once the assimilation period is over, the forecast model drifts gently from the simulated reality (true model), because of the error in the forecast model.

The introduction of the control of the systematic model error (thick plain line) reduces the global level of error during the assimilation window, and moreover prevents the analysed error from being u-shaped. The fact that the evolution curve of the analysed error is almost horizontal means that a large part of the model error has been corrected. Moreover, in the VDA-ICSE, the initial conditions can be estimated independently of the correction of model errors. In Fig. 4 the same $\mathbf{x}^{\mathrm{true}} - \delta\mathbf{x}_{\mathrm{ICSE}}^{\mathrm{opt}}$ increment error as for VDA-IC is plotted. Except at some locations on the boundary, the error in the initial state is highly reduced compare to VDA-IC and the 'two-phased' structured error disappears and is replaced by a low amplitude noise. This is similar to what is obtained by VDA-IC in a 'perfect model' case (not shown).

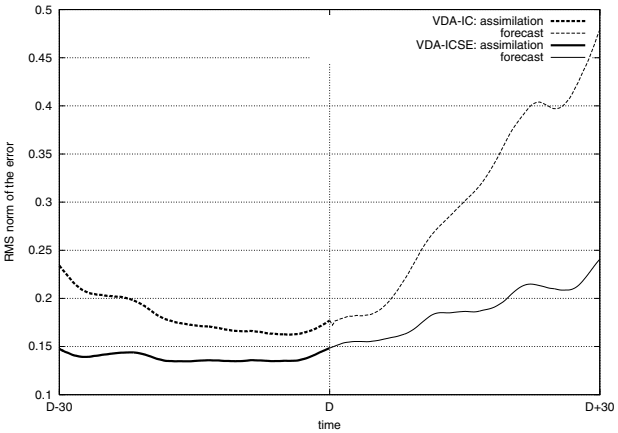
The forecast originating from VDA-IC (dashed thin line) evenly drifts; so does the forecast originating from VDA-ICSE if the forecast model is the same (thin dotted line). However, if the model error correction term $\eta_0$ computed by the assimilation is still present in the forecast equation, i.e. if the forecast integration is continuously corrected by the estimated systematic error (plain thin line), the drift is highly reduced. This may indicate that the estimated model error correction term is representative of the systematic part of the error present in the model. Comparing the 'true' and the estimated SSH component of the systematic model error (see Fig. 5) we can notice the similarities between these two fields. Although they are not identical, the 'two-phase' shapes are quite well reconstructed in the estimated field. In operational applications, applying the systematic error correction during the forecast phase may not provide such a significant improvement. Indeed, the forcing fields usually come




(a) SSH

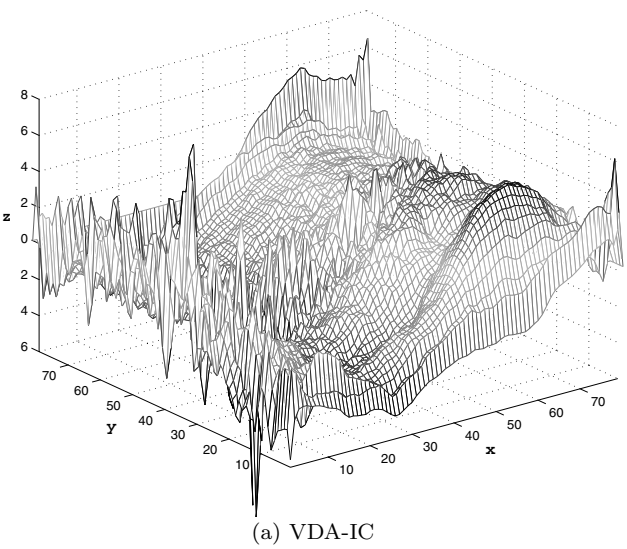



(b) meridional current velocity

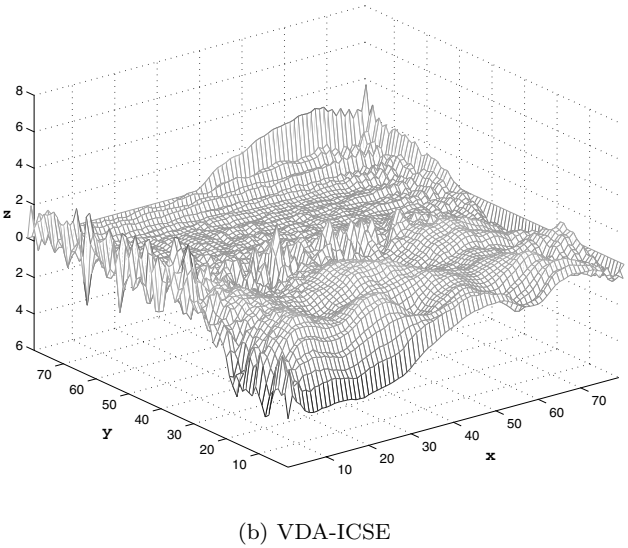



(c) zonal current velocity

*Fig 3.* Time evolution of the sea surface elevation, zonal and meridional current rms error for the assimilation and forecast windows.




(a) VDA-IC




(b) VDA-ICSE

*Fig 4.* Error on the initial condition increment for VDA-IC (left) VDA-ICSE (right) schemes.

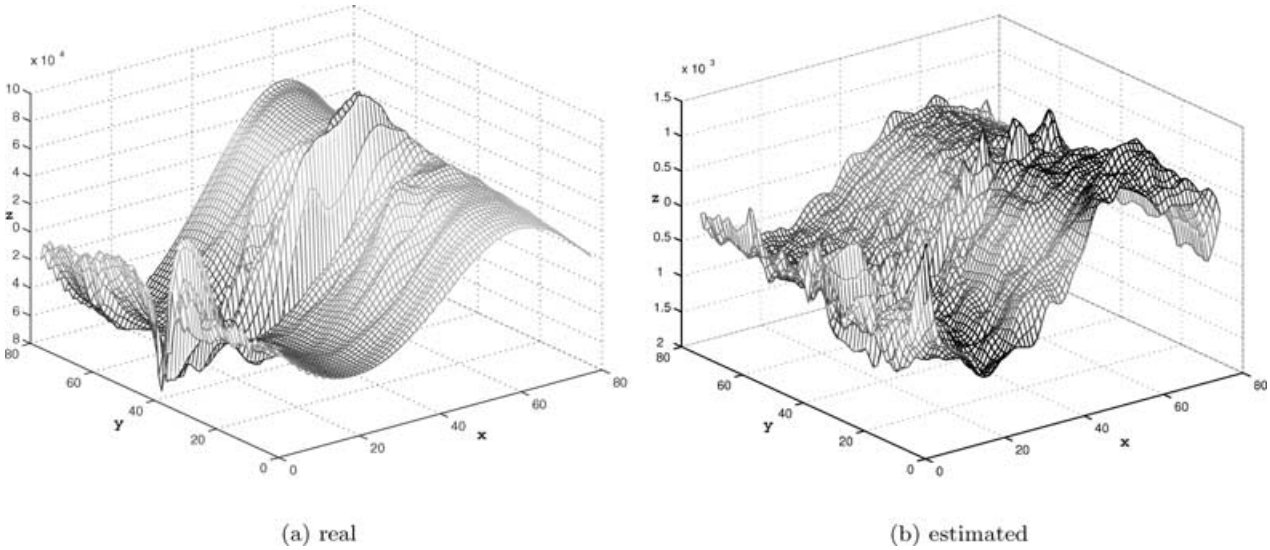



*Fig 5.* Known model error (left) and estimated error by the VDA-ICSE observing the whole SSH field every three days (right).

from the 'best' available atmospheric source, such as reanalysis, whereas the forecast phase is a coupled ocean–atmosphere integration. In this case the systematic error may not be the same during the forecast and the assimilation period. However, the correction term can still be kept in the framework of a hindcast experiment and of coupled data assimilation

Figures 3(b) and (c) represent the rms error of the analysed trajectory for the non-observed component of the state vector **u** and **v**. Results of VDA-ICSE on these components preserve a good quality: the level of error during the assimilation period is lower than in the VDA-IC case and the good estimation of the model error correction term on these components leads to a reduction in the forecast drift as well. The quality of this result is due mainly to the fact that the operator of multiplication by **Q** is multivariate (see Appendix B).

Starting from the VDA-IC scheme, which is the operational data assimilation algorithm in several weather centres, the introduction of the control of the systematic part of model errors does not require drastic algorithm modifications. It is sufficient to add the correction term into the direct model equation and to sum the adjoint state vector at each time-step to compute the corresponding gradient term. Because the adjoint model is already integrated for the VDA-IC, one iteration of VDA-ICSE is about the same cost. However, as the size of control is twice that of VDA-IC we can expect the number of iterations needed for convergence to be higher. Figure 6 shows that, for the same number of iterations, the discrepancy between the observation and model state in VDA-ICSE is at least as good as VDA-IC and gradually improves as the number of iterations increases. At the end of the fixed number of iterations, VDA-IC no longer seems to be able to improve the results, whereas the VDA-ICSE analysed model state is still getting closer to observations.

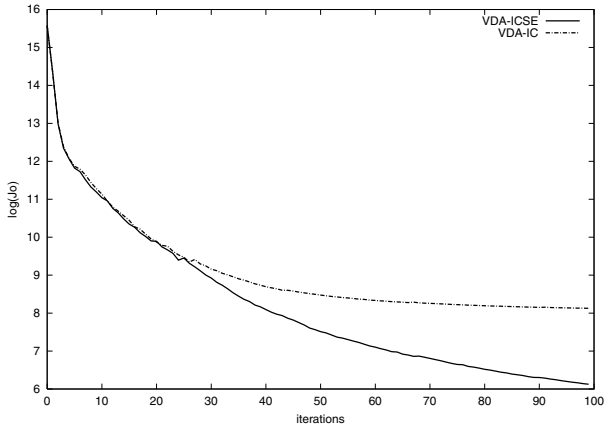



*Fig 6.* Comparison of the logarithm of $J_o$ terms.

*4.2.2. Assimilation along 'satellite' tracks.* In the previous section, the observations were the whole SSH field every three days. This may be related in real life to the assimilation of satellite maps interpolated from tracks. One of the advantages of variational methods is to allow the use of observations directly when and where they are produced. Therefore, the observation data set is now reduced to a style of 'along satellite tracks' (see Fig. 7).

The number of observations is now only 25% of the number of degrees of freedom of our minimization problem. This remains an overestimated problem, but now the majority of the information comes from the background terms of the cost function $J_b$ and $J_\eta$. Because the background covariance matrices are not usually well known, the lack of observations may degrade the results. Indeed, in an unobserved area, the methods may introduce spurious corrections in areas without observations in order to improve the fit in an area where there are observations. These

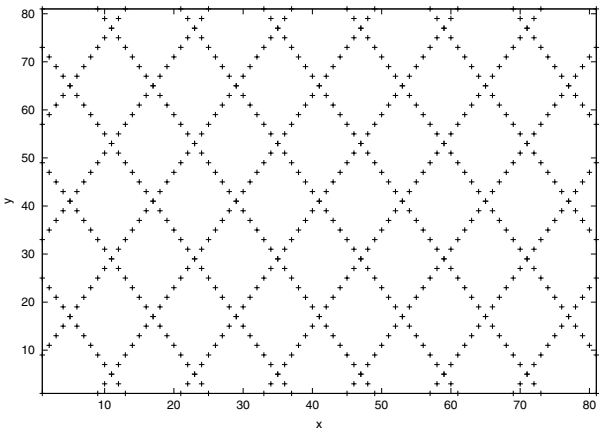



*Fig 7.* Location of the observations (ground track of satellites).

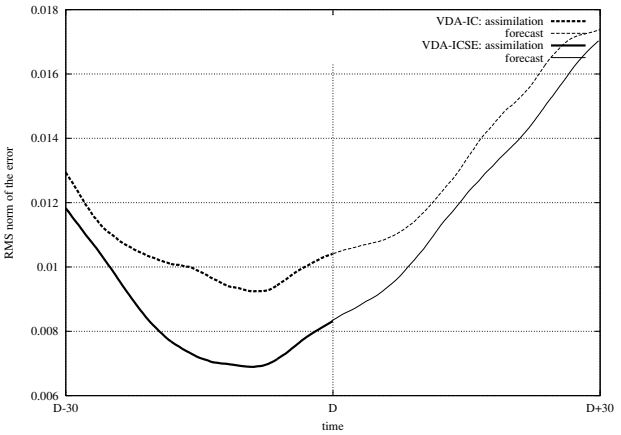
spurious corrections damage the overall balance of the model fields and can lead, for instance, to the creation of unexpected and unwanted currents.

Figure 8 shows the corresponding rms error evolution plots as Fig. 3. It can be seen that the VDA-ICSE still improves the results compared to VDA-IC during the assimilation period on every component of the state vector. However the SSH error curve remains u-shaped, which may indicate that model errors are still present. Moreover, the improvement of the drift of the model is not as clear as in the previous experiment.

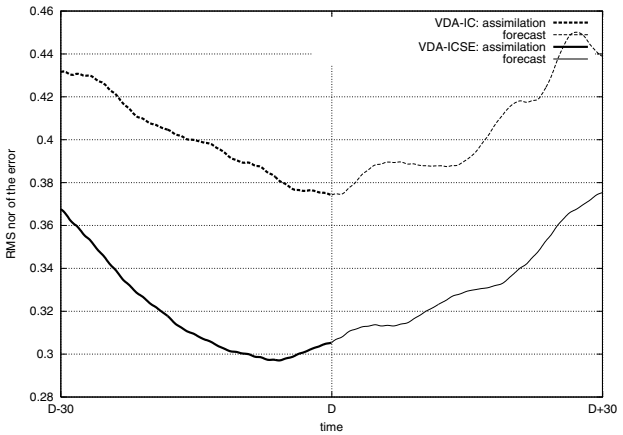
The u shape is even more pronounced in the ICSE experiment, which is better thanks to more degrees of freedom, but the systematic error is no longer well estimated (see Fig. 9).

Even if the introduction of the control of systematic error reduces the level of error during the assimilation window, the number of observations is no longer sufficient to allow a correct estimation of the systematic part of the error. Indeed, the estimated correction term may include information about other types of error such as the non-systematic part of model errors or a remaining part of the initial condition error, or even a term reducing the difference between observation and model but without any consistency with the reality in non-observed areas (aliasing).

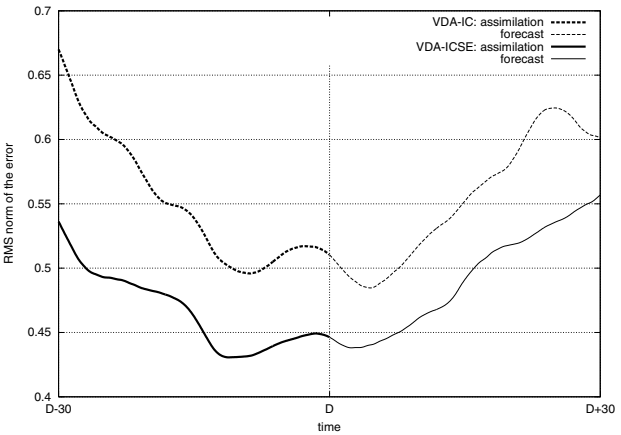
If we want to retrieve the good quality of the previous experiment, we need to introduce more information. This may be done using prior knowledge of the systematic error. One of the




(a) SSH




(b) meridional current velocity




(c) zonal current velocity

*Fig 8.* Time evolution of the sea surface elevation, zonal and meridional current rms error for the assimilation and forecast window.

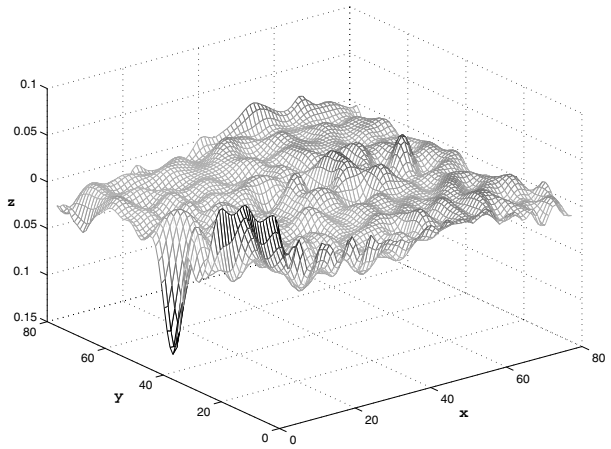



*Fig 9.* Estimated error by the VDA-ICSE observing SSH along satellite tracks.

possible ways to do this is to modify the systematic error background term $J_\eta$ of the cost function, and more precisely using the **Q** operator. This means that the size of the ensemble of eligible states for the minimization can be reduced in order to obtain an analysed state more consistent with the physics of the system.

In the current case, we can introduce in the matrix **Q** a constraint on the two-main-phase structure that appears on the 'theoretical' error (Fig. 2). In other words, we force the method to estimate the error only along the larger structures (these two phases) because there is not enough information to correct smaller structures in an efficient way. This can be done by adding a new operator **T** into the change of variable eq. (14) and the subsequent gradient computation eq. (16) by

$$\eta_0 = \mathbf{Q}^{1/2}\mathbf{T}\alpha \tag{17}$$

$$\nabla_\alpha J = \mathbf{T}^T\mathbf{Q}^{T/2}\nabla_{\eta_0} J. \tag{18}$$

In this case **T** can be built using the 'sinusoidal' shape of the systematic part of the model error to be estimated, in a simplistic way:

$$(\mathbf{T}\alpha)(i,j) = \alpha(j)\sin(-\pi i/I) \quad i = 1\ldots I \quad j = 1\ldots J. \tag{19}$$

In addition, this approach reduces dramatically the size of the systematic error part of the control vector (from $I \times J$ to $I$). Constraining the minimization to focus on larger structures permits us to estimate a correction term more consistent with the error present in the model. This leads directly to a better quality forecast (see Fig. 10). During the assimilation time window, the gain provided by the 'great structure' version is not obvious, but the fact that the error curves are more horizontal may prove that the correction term is better estimated.

*4.2.3. Remark on incremental formulation.* In order to have a quadratic cost function, most of the operational VDA schemes use the incremental formulation (Courtier et al. 1994) where the direct model $M$ is replaced by its tangent linear (linearized with respect to a reference trajectory). The reference trajectory is updated at regular intervals during the minimization to keep the tangent linear hypothesis[1] valid. The VDA-ICSE can be applied in an incremental formulation provided the model error correction term is treated as an increment, and the model error model **T** is linearized.

## 5. Conclusions

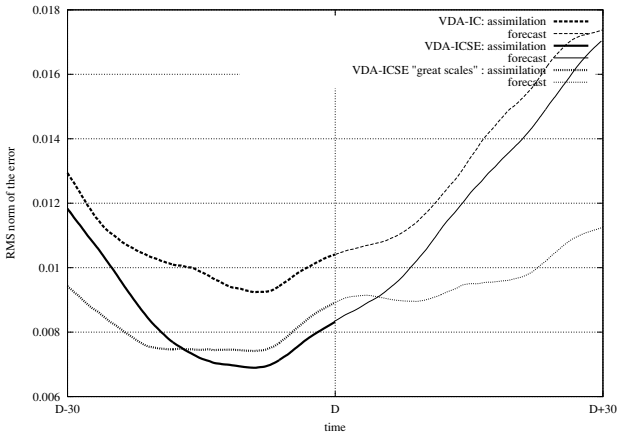
To control the model error is an important issue. In the presence of the model error, the control of the initial condition only is not enough. Indeed, the presence of model errors forces the assimilation to introduce errors in the initial conditions to correct for them. Treating separately initial condition errors and model errors may allow a better estimation of both corrections. Nevertheless, the control of the model error in its classical form is too expensive for current computational means. In addition, the systematic and random components of the error should be treated separately. Therefore, controlling the systematic part of the error has several advantages. First, the additional computational cost is negligible compared to the control of the initial condition only. Indeed, the gap between analysis and observation is smaller in the VDA-ICSE case compared to the VDA-IC case, even for a small number of minimization iterations. Moreover, the computational cost of one iteration is comparable in both cases. Another interest of this approach is to allow us to keep the estimated model error correction term to constrain the forecast, and this leads to a significant improvement on the forecast.

However, because controlling the model error doubles the number of degrees of freedom of the minimization, additional care has to be taken when observations are sparse in time and/or in space. Such a lack of information can be filled by improving the model error covariance matrix and/or by focusing on the largest scales of the correction only. In other words, if there is not enough information brought by the observations to obtain a good estimation of the control vector, either one adds information (in the error covariance matrices for instance) or one can restrict the control domain.

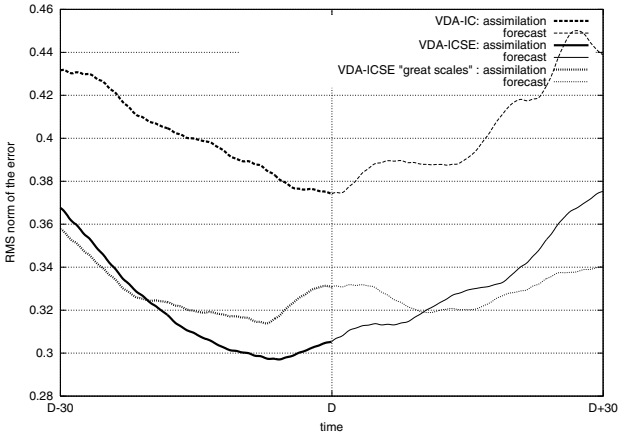
Because the model error shape is seldom known in any realistic application, the way of reducing the size of control used at the end of the previous section cannot be used directly. One has to gather information coming from prior knowledge of the model errors such as weakness of the model, which approximations are done in its equations, and/or statistical information coming from previous analysis.

This problem of the number of observations versus the number of degrees of freedom of the minimization problem is not restricted to the control of the model error only. In a more general way, we can ask the question: 'Is the resolution of our model

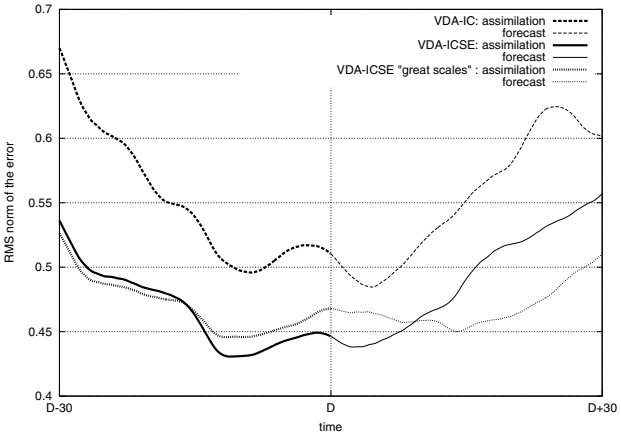
[1] $M_{(t_0,t_N)}(\mathbf{x}^{\mathrm{b}} + \delta\mathbf{x}_0) = M_{(t_0,t_N)}(\mathbf{x}^{\mathrm{b}}) + \mathbf{M}_{(t_0,t_N)}\delta\mathbf{x}_0$.




(a) SSH




(b) meridional current velocity




(c) zonal current velocity

*Fig 10.* Time evolution of the sea surface elevation, zonal and meridional current rms error for the assimilation and forecast window.

the best choice given our observation data set?' Let us suppose that several models are available with different grid scales, approximations and parametrizations to describe one given system behaviour. For a given observation data set, the highest resolution model may not be the best choice for assimilating this data set. Indeed, if it models physical phenomena that are not well represented by observations because of too small scales, the assimilation might not be able to control them and therefore the analysis might be degraded. On the other hand, the use of a too coarse model does not lead to a high-quality analysis. The choice of the resolution of the model should be related to the observation network used for the data assimilation.

Because of the simplicity of the model and the twin-experiment framework, the conclusion drawn in this paper serves only as guidance to more realistic experiments. An early implementation on the Ocean General Circulation model OPA (Madec et al. 1998) of the VDA-ICSE scheme has been carried out (Vidard, 2001). It shows that despite the low coverage of the observation data set, the control of systematic model errors can improve the results of the VDA-IC.

## 6. Acknowledgment

We wish to thank Sophie Durbiano, Eric Blayo and Anthony Weaver for helpful discussions.

## 7. Appendix A: derivation of the lagrangian function

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}^*} = \int_{t_0}^{T} \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} - M(\mathbf{x}(t)) - \mathbf{T}(\eta(t))\,\mathrm{d}t \tag{A1}$$

$$\frac{\partial \mathcal{L}}{\partial \eta^*} = \int_{t_0}^{T} \frac{\mathrm{d}\eta}{\mathrm{d}t} - \Phi(\eta(t), \mathbf{x}(t))\,\mathrm{d}t \tag{A2}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \sum_{i=1}^{I} \mathbf{H}_{t_i}^T \mathbf{R}_{t_i}^{-1} \left[ H_{t_i}(\mathbf{x}(t_i)) - \mathbf{y}_{t_i} \right] + \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^{\mathrm{b}}) + \int_{t_0}^{T} -\frac{\mathrm{d}\mathbf{x}^*}{\mathrm{d}t} - \left[ \frac{\partial \mathbf{M}}{\partial \mathbf{x}} \right]^T \mathbf{x}^* + \left[ \frac{\partial \mathbf{\Phi}}{\partial \mathbf{x}} \right]^T \eta^*\,\mathrm{d}t \tag{A3}$$

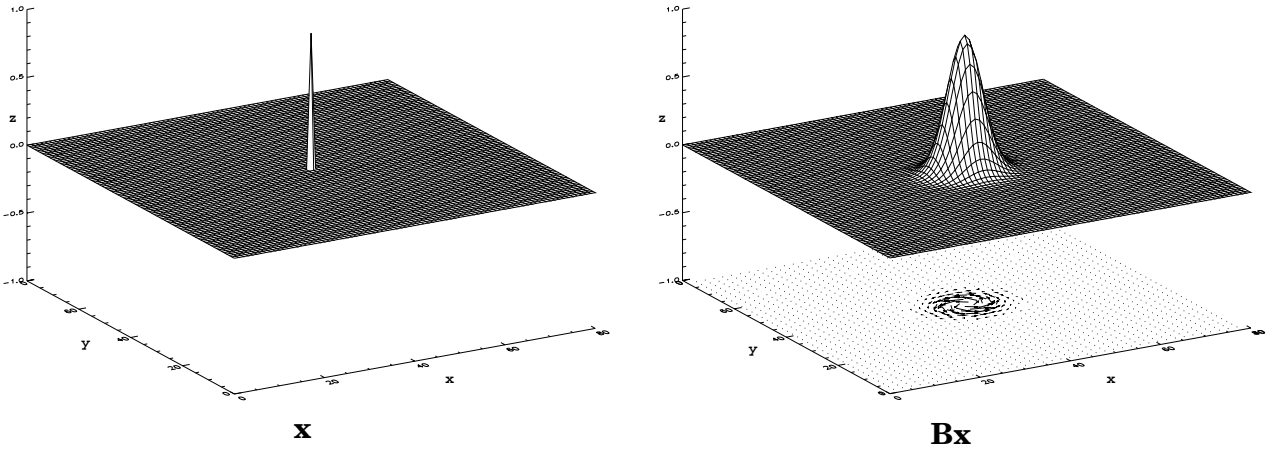



*Fig 11.* Sea surface elevation (surface shade) and current velocity field (vector field) for one single Dirac on the sea surface elevation vector(left) and its result from the multiplication by **B** (right).

$$\frac{\partial \mathcal{L}}{\partial \eta} = \mathbf{Q}_0^{-1}\left(\eta_0 - \eta^{\mathrm{b}}\right) + \int_{t_0}^{T} -\frac{\mathrm{d}\eta^*}{\mathrm{d}t} - \left[\frac{\partial \mathbf{\Phi}}{\partial \eta}\right]^T \eta^* + \mathbf{T}^T \mathbf{x}^* \,\mathrm{d}t. \tag{A4}$$

## 8. Appendix B: B and Q operators

The background and model error covariance matrices (**B** and **Q**) play an important role in determining the spatial structure of the analysis initial condition increment and systematic model error correction term. Unfortunately, because of the size of the problem, it is not possible to store explicitly these matrices in computer memory. Moreover, in a realistic application these covariances are seldom known and, because of the sparsity of the external information (observations), it is difficult, if not impossible, to obtain complete and accurate estimates of them.

**B** can be expressed as

$$\mathbf{B} = \begin{pmatrix} \mathbf{B_{uu}} & \mathbf{B_{uv}} & \mathbf{B_{uh}} \\ \mathbf{B_{vu}} & \mathbf{B_{vv}} & \mathbf{B_{vh}} \\ \mathbf{B_{hu}} & \mathbf{B_{hv}} & \mathbf{B_{hh}} \end{pmatrix} \tag{B1}$$

with

$$\mathbf{B_{xx}} = \Sigma_{\mathbf{xx}} \mathbf{C_{xx}} \Sigma_{\mathbf{xx}}. \tag{B2}$$

$\Sigma_{\mathbf{xx}}$ are the matrices of standard deviation of the unbalanced error on variable **x** (**u**, **v**, or **h**) and $\mathbf{C}_{\mathrm{xx}}$ are the corresponding correlation matrices. In this paper, $\mathbf{C}_{\mathrm{xx}}$ are assumed to be Gaussian and are modelled using a diffusion operator. This is a direct application of Weaver and Courtier (2001). Following Derber and Bouttier (1999), the **B** matrix can be written as

$$\mathbf{B} = \mathbf{K}_{\mathrm{b}} \hat{\mathbf{B}} \mathbf{K}_{\mathrm{b}}^T \tag{B3}$$

where

$$\hat{\mathbf{B}} = \begin{pmatrix} \mathbf{B_{uu}} & 0 & 0 \\ 0 & \mathbf{B_{vv}} & 0 \\ 0 & 0 & \mathbf{B_{hh}} \end{pmatrix}$$

and $\mathbf{K}_{\mathrm{b}}$ is the so-called balance operator. In the test case presented in this paper, we can use the geostrophic balance (eqs B4 and B5) and the hydrostatic hypothesis (B6) to define the relationship between **u**, **v** and **h**:

$$\rho f \mathbf{v}^{eq\mathbf{h}} = \frac{\partial \mathbf{p}}{\partial x} \tag{B4}$$

$$\rho f \mathbf{u}^{eq\mathbf{h}} = -\frac{\partial \mathbf{p}}{\partial y} \tag{B5}$$

$$\mathbf{p} = \rho g \mathbf{h}. \tag{B6}$$

The balance operator is then defined by

$$\mathbf{K}_{\mathrm{b}} = \mathbf{K}'_{\mathrm{b}} + \mathbf{I} = \begin{pmatrix} \mathbf{I} & 0 & -\frac{g}{f}\frac{\partial .}{\partial y} \\ 0 & \mathbf{I} & \frac{g}{f}\frac{\partial .}{\partial x} \\ 0 & 0 & \mathbf{I} \end{pmatrix}. \tag{B7}$$

Figure B1 shows the result (right) of the multiplication of a one-state vector with a sea surface elevation of 1 in the middle of the domain and 0 everywhere and with no current at all (left) by the **B** operator constructed as described before. We can see the Gaussian correlation of the sea surface elevation and the vortex in current velocity resulting from this elevation.

## References

Adcroft, A. and Marshall, D. 1998. How slippery are piecewise-constant coastlines in numerical ocean models. *Tellus* **50A**, 95–108.

Bell, M. J., Martin, J. M. and Nichols, N. K. 2001. *Assimilation of data into an ocean model with systematic errors near the equator*. Ocean Applications Technical Note No 27, UK Met Office. 27 p.

Cohn, S. E. 1997. An introduction to estimation theory. *J. Meteor. Soc. Japan* **75**, 257–288.

Courtier, P. 1997. Dual Formulation of Four–Dimensional Variational Assimilation. *Q. J. R. Meteorol. Soc.* **123**, 2449–2461.

Courtier, P., Thépaut, J. N. and Hollingsworth, A. 1994. A startegy for operational implementation of 4d-var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**, 1367–1387.

D'Andrea, F. and Vautard, R. 2001. Reducing Systematic Errors by Empirically Correcting Model Errors. *Tellus* **52A**, 21–41.

Dee, D. P. and Da Silva, A. M. 1998. Data assimilation in the presence of forecast bias. *Q. J. R. Meteorol. Soc.* **124**, 269–295.

Derber, J. C. 1989. A Variational Continuous Assimilation Technique. *Mon. Wea. Rew.* **117**, 2437–2446.

Derber, J. and Bouttier, F. 1999. A reformulation of the background error covariance in the ecmwf global data assimilation system. *Tellus* **51A**, 195–221.

Durbiano, S. 2001. *Vecteurs caractéristiques de modèles océaniques pour la réduction d'ordre en assimilation de données.* Reduced order strategy for 4d-var ocean data assimilation. PhD thesis, Université Joseph Fourier (Grenoble I), December, 214 p. In French.

Gilbert, J.-C. and Lemaréchal, C. 1989. Some numerical experiments with variable-storage quasi-Newton algorithms. *Math. Prog.* **45**, 407–435.

Griffith, A. K. and Nichols, N. K. 2001. Adjoint methods in data assimilation for estimating model error. *J. Flow Turbulence Combustion* **65**, 469–488.

Jazwinski, A. H. 1970. *Stochastic Processes and Filtering Theory*. Academic Press, New York.

Le Dimet, F.-X. and Talagrand, O. 1986. Variational algorithms for analysis and assimilation of meteorological observation: theroetical aspects. *Tellus* **38A**, 97–110

Madec, G., Delecluse, P., Imbard, M. and Lévy, C. 1998. OPA 8.1 ocean general circulation model: Reference manual. Note du ple de modélisation 11, IPSL, Paris, France. 91 p.

Molteni, F., Buizza, R., Palmer, T. N. and Petroliagis, T. 1996. The new ECMWF ensemble prediction system: methodology and validation. *Q. J. R. Meteorol. Soc.* **122**, 73–119.

Nocedal, J. 1980. Updating quasi-Newton matrices with limited storage. *Math. Comp.* **24**, 773–782

Sasaki, Y. 1958. An objective analysis based on the variational method. *J. Meteorol. Soc. Japan* **36**, 77–88.

Vidard, A. 2001. *Vers une prise en compte des erreurs modèle en assimilation de données 4D-variationnelle. Application à un modèle réaliste d'océan*. Toward the correction of model error in 4D variational data assimilation. Application to a realistic ocean model. PhD thesis, Université Joseph Fourier (Grenoble I), 196 p. In French.

Vidard, A, Blayo, E., Le Dimet, F.-X. and Piacentini, A. 2001. 4d-variational data analysis with imperfect model. reduction of the size of control. *J. Flow Turbulence Combustion* **65**, 489–504.

Vidard, A., Le Dimet, F.-X. and Piacentini, A. 2003. Optimal determination of nudging coefficient. *Tellus* **55A**, 1–15.

Weaver, A. and Courtier, P. 2001. Correlation modelling on the sphere using a generalized diffusion equation. *Q. J. R. Meteorol. Soc.* **127**, 1815–1846.

Zou, X., Navon, I. M. and Le Dimet, F.-X. 1992. An optimal nudging data assimilation scheme using parameter estimation. *Q. J. R. Meteorol. Soc.* **118**, 1163–1186.