4D Variational Data Analysis with Imperfect Model

Arthur Vidard, Eric Blayo, François-Xavier Le Dimet, Andrea Piacentini

4D Variational Data Analysis with Imperfect Model

P. A. Vidard[1][2] E. Blayo[1] F-X. Le Dimet[1] A. Piacentini[2]

referring author : P. A. Vidard : vidard@imag.fr

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

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

Abstract:

One of the main hypothesis made in variational data assimilation is to consider that the model is a strong constraint of the minimization i.e. that the model describes exactly the behavior of the system. Obviously the hypothesis is never respected. We propose here an alternative to the 4D-Var that takes into account model errors adding a non physical term into the model equation and controlling this term. A practical application is proposed on a simple case and a reduction of the size of control using preferred directions is introduced to make the method affordable for realistic applications.

keyword : adjoint methods, data-assimilation, optimal control, model errors, order reduction, Kalman filtering

Data Assimilation - an introduction

Among applications of adjoint methods, the variational data assimilation has known a strong development during the last years and it provides nowadays operational (or pre-operational) tools in meteorology and oceanography.

The 'data assimilation' term indicates a class of very useful mathematical methods which redraw the state or the evolution of a system using the information provided by the prediction of a model and by the observations of the system.

In oceanography and meteorology these methods have two main goals:

- 'Smoothing': the assimilation provides, for a given time period , an estimate of the system evolution which is optimally adjusted with both the observations and the model. It allows a historical analysis of the behavior of the system, the location of the specific behaviors, etc... It is also a relevant tool for testing physical assumption and for sensitivity studies.
- 'Filtering': here the aim of the assimilation is to estimate the current best state of the system starting from all the past and present information. Only the current state is optimal.

This second approach is widely used in operational methods in meteorology. It consists in finding, using informations provided by a field of observations over a given period, a model solution which is as close as possible to the observations in order to obtain good initial conditions (or boundary conditions) for the following forecast. Although this is not data assimilation in a strict sense, it is the direct prolongation of the most recent results in filtering.

One generally classifies data assimilation methods in two classes: variational methods on one side, and stochastic methods on the other. The first ones are based on optimal control theory; these are the ones currently operational (or pre-operational) in meteorological centers (4D-Var). The control variable is the initial condition of the assimilation window, and the model is a strong constraint. Optimization is carried out by a gradient method. However, although the computational cost is affordable for the current computers, it is still very high : several dozens of iterations (thus as many integrations of the model) are generally necessary for the convergence of the method.

The stochastic approach is based on the optimal statistical estimate theory, the most famous stochastic method being the Kalman filter (Kalman, 1960). This one requires, in its complete form, the handling of $N \times N$ matrices (multiplication, resolution of linear systems...) where N is the size of state vector, and this for each time step. This implies significant computation and storage capacities, and perhaps out of reach for real application in meteorology or oceanography. Then we generally have to make simplifying assumptions. These methods apply more particularly to the filtering problems (from where their name) although there are some alternative formulations adapted to smoothing.

(Lorenc, 1986 [11]) has shown that for a perfect linear model with linear observation operators, 4D-Var and Kalman filter provide the same results at the end of the assimilation period. However for an imperfect model, the Kalman filter takes into account model errors whereas classical 4D-Var not.

The next step for variational data assimilation development is to takes into account model errors (i.e the model becomes a weak constraint for the minimization). While the 4D-Var with imperfect model algorithm is seldom known (Cohn 1997 [5], Zupanski 1997 [20]) the development of low cost versions of this alternative of 4D-Var has just started recently. The entire method of model error control is very costly in practice for operational application therefore only one part of the error can be corrected. Several ways have been proposed to reduce the cost, Griffith et al. [7], Dee and Da Silva [6], D'Andrea and Vautard 1999 [1], focus on systematic model error (bias), Zou et al 1992 [19] propose an optimal determination of nudging coefficients. In this paper, we will introduce another technique for treating model error reducing the size of the control. The aim of this method is, starting from complete 4D-Var with imperfect model method, to correct the model following only a reduced number of well chosen directions. Using a very simple example and thanks to a Kalman filter relationship, we will show that the control of model error provides really better results than classical 4D-Var for an affordable over-cost (in its reduced form)

In this article we will adopt the unified notations for data assimilation introduced by Ide et al [8]

Classical method: initial condition control (4D-Var)

As we said before, the operational method for weather forecasting is the optimal control of the initial condition. A non-linear model is considered. Let $\mathbf{x} \in [\mathcal{C}(0,T)]^n$ be the variable describing the state of the atmosphere or the ocean.

$$\begin{cases} \dfrac{d\mathbf{x}}{dt} = M(\mathbf{x}) \\ \mathbf{x}(0) = \mathbf{x}^b + \delta\mathbf{x}_0 \end{cases} \tag{1}$$

where M is a non linear operator.

The goal of this data assimilation method is to look for the correction of the initial condition $\delta\mathbf{x}_0$ so that the solution of the model is close to observations. Thus our problem will be to minimize the cost function $J(\delta\mathbf{x}_0)$ which is the norm of the difference between the solution of the model and the observations, plus a regularization term that restricts the degree of freedom of the variation of the initial condition.

$$J(\delta\mathbf{x}_0) = \frac{1}{2}\left(\parallel H(\mathbf{x}) - \mathbf{y}^o \parallel^2 + \varepsilon \parallel \delta\mathbf{x}_0 \parallel^2\right)$$

$\mathbf{y}^o$ is the vector of observed variables of the system. The operator H, called observation operator, is the projection of the system space onto the observation space. For example satellites measure the temperatures with radiances; the role of this operator is then of 'translating' the state variables (hygroscopy, temperature...) into its radiances equivalence. In addition the observations are neither done at every moment nor at all the space discretization node, which is also taken into account in the operator H. The ε parameter represents the relative weight of the regularization term compared to the variation of the observation term in the cost function.

In the linear case, provided the matrix $\mathbf{H}$ is semi-definite positive, J is α-convex on an convex set. It thus admits a single minimum $\delta\mathbf{x}_0^*$ (cf [13, LIONS]). We are in the case of an optimization without constraint. The optimal initial condition $\delta\mathbf{x}_0^*$ is given by $\nabla J(\delta\mathbf{x}_0^*) = 0$. The functional calculus J depends on $\mathbf{v}$ only implicitly. This implies that the calculation of its gradient becomes unaffordable by 'classical' direct methods. However the use of an adjoint method as described below allows, thanks to a double model integration (forward and backward), to obtain it.

Let us consider $\hat{\mathbf{x}}$ the directional derivative of $\mathbf{x}$ in the direction $h_{\delta\mathbf{x}_0}$ and $\hat{J}$ the directional derivative J.

If we notes $\mathbf{M}_{|\mathbf{x}} = [\frac{dM}{d\mathbf{x}}(\mathbf{x})]$, then

$$\begin{cases} \dfrac{d\hat{\mathbf{x}}}{dt} = \mathbf{M}_{|\mathbf{x}}.\hat{\mathbf{x}} \\ \\ \hat{\mathbf{x}}(0) = h_{\delta\mathbf{x}_0} \end{cases} \tag{2}$$

$$\hat{J}(\delta\mathbf{x}_0, h_{\delta\mathbf{x}_0}) = \int_0^T (\mathbf{H}^T(H(\mathbf{x}) - \mathbf{y}^o), \hat{\mathbf{x}})dt + \varepsilon(\delta\mathbf{x}_0, h_{\delta\mathbf{x}_0}) = (\nabla J, h_{\delta\mathbf{x}_0}) \tag{3}$$

System (2) is called the tangent linear system.
Let introduce $\mathbf{x}^*$, called the adjoint state, that we will define later by convenience.
Multiplying (2) by $\mathbf{x}^*$ and integrating between 0 and T, we obtain :

$$\int_0^T (\mathbf{x}^*, \frac{d\hat{\mathbf{x}}}{dt})dt = \int_0^T (\mathbf{x}^*, \mathbf{M}_{|\mathbf{x}}.\hat{\mathbf{x}})dt$$

After an integration by parts, we obtain :

$$-\int_0^T (\frac{d\mathbf{x}^*}{dt}, \hat{\mathbf{x}}) + (\mathbf{x}^*(T), \hat{\mathbf{x}}(T)) = \int_0^T (\mathbf{M}_{|\mathbf{x}}{}^T\mathbf{x}^*, \hat{\mathbf{x}})dt + (\mathbf{x}^*(0), h_{\delta\mathbf{x}_0})$$

We define $\mathbf{x}^*$ such as it satisfies

$$\begin{cases} -\dfrac{d\mathbf{x}^*}{dt} = \mathbf{M}_{|\mathbf{x}}{}^T\mathbf{x}^* + \mathbf{H}^T(H(\mathbf{x}) - \mathbf{y}^o) \\ \mathbf{x}^*(T) = 0 \end{cases} \tag{4}$$

Then, by identification with (3):

$$\int_0^T (\mathbf{H}^T((H)\mathbf{x} - \mathbf{y}^o), \hat{\mathbf{x}}) = (\mathbf{x}^*(0), h_{\delta\mathbf{x}_0})$$

Thus one can, by integrating backward the adjoint model (4), compute the gradient components : $\nabla J = \mathbf{x}^*(0) + \varepsilon\delta\mathbf{x}_0$

Thus we can deduce that $\delta\mathbf{x}_0$ is optimal for $\delta\mathbf{x}_0^* = -\frac{1}{\varepsilon}\mathbf{x}^*(0)$

So we have optimality system :

$$\begin{cases} \dfrac{d\mathbf{x}}{dt} = \mathbf{M}_{|\mathbf{x}}.\mathbf{x} \\ \mathbf{x}(0) = \mathbf{x}^b - \dfrac{1}{\varepsilon}\mathbf{x}^*(0) \\ \\ -\dfrac{d\mathbf{x}^*}{dt} = {\mathbf{M}_{|\mathbf{x}}}^T\mathbf{x}^* + \mathbf{H}^T(H(\mathbf{x}) - \mathbf{y}^o) \\ \mathbf{x}^*(T) = 0 \end{cases} \tag{5}$$

Model error control

Despite convincing results, the preceding approach comprises several significant limitations. The main one is the fact that the model is imposed as a strong constraint. In spite of the use of extremely sophisticated models, they remain a significant source of error (we will never manage to model perfectly the behavior of the atmosphere or the ocean). Two possible solutions seem to arise: on one hand a thorough analysis the model errors allowing an a priori correction of model equations , or on the other hand, a correcting term added directly inside the model, determined by optimal control. It is the second option which we are interested on.

In this way, we will add a correction term to the model, intended to deal with uncertainties of the model at every time step. This term will then become our control variable and we will minimize, by a variational method, the difference between the solution of the 'weak' model and observations.

Optimality system

One adds in (1) the term $B\eta$ intended to correct the errors of model.

$$\begin{cases} \dfrac{d\mathbf{x}}{dt} = M(\mathbf{x}) + B\eta \\ \mathbf{x}(0) = \mathbf{x}^b + \delta\mathbf{x}_0 \end{cases} \tag{6}$$

with $\eta \in [\mathcal{C}(0,T)]^m$.

Now, the goal of the assimilation is to determinate this estimation of the error so that the solution of the model will be close to observations. Our problem will be thus to minimize the functional $J(\eta)$ which is the norm of the difference between model solution and observations plus a regularization term that restricts the freedom of the model error correction term.

$$J(\eta) = \frac{1}{2}\left(\parallel H(\mathbf{x}) - \mathbf{y}^o \parallel^2 + (N\eta, \eta)\right)$$

The operator H is the observation operator and N is used to weight the importance of regularization term of the cost function. In the linear case, provided the matrix $\mathbf{H}$ is semi-definite positive and that $\mathbf{N}$ is symmetrical definite positive, J is α-convex on a convex set. It thus admits a single η^* minimum (cf [13, Lions]).

Let us considering $\hat{\mathbf{x}}$ the directional derivative of $\mathbf{x}$ in the direction h_η, and $\hat{J}$ the directional derivative of J. By modifying calculations of paragraph 1.2, one obtains :

$$\begin{cases} \dfrac{d\hat{\mathbf{x}}}{dt} = \mathbf{M}_{|\mathbf{x}}.\hat{\mathbf{x}} + Bh_\eta \\ \\ \hat{\mathbf{x}}(0) = 0 \end{cases} \tag{7}$$

$$\hat{J}(\eta, h_\eta) = \int_0^T (\mathbf{H}^T(H(\mathbf{x}) - \mathbf{y}^o), \hat{\mathbf{x}}) + (N\eta, h_\eta)dt = (\nabla J, h_\eta) \tag{8}$$

we introduce $\mathbf{x}^*$ the adjoint state which is the same as in the previous case :

$$\begin{cases} -\dfrac{d\mathbf{x}^*}{dt} = \mathbf{M}_{|\mathbf{x}}{}^T\mathbf{x}^* + \mathbf{H}^T(H(\mathbf{x}) - \mathbf{y}^o) \\ \mathbf{x}^*(T) = 0 \end{cases} \tag{9}$$

In the same way as in the previous paragraph, one obtains : $\nabla J = B^T\mathbf{x}^* + N\eta$.
That is to say : η is optimal for $\eta = -N^{-1}B^T\mathbf{x}^*$
So we have the optimality system :

$$\begin{cases} \dfrac{d\mathbf{x}}{dt} = M(\mathbf{x}) - BN^{-1}B^T\mathbf{x}^* + g(t) \\ \mathbf{x}(0) = U \\ \\ -\dfrac{d\mathbf{x}^*}{dt} = \mathbf{M}_{|\mathbf{x}}{}^T\mathbf{x}^* + \mathbf{H}^T(H(\mathbf{x}) - \mathbf{y}^o) \\ \mathbf{x}^*(T) = 0 \end{cases} \tag{10}$$

In the case of linear model, a comparison between model error control and the Kalman filtering (cf [18]) showed that an adequate choice for N and B adequate was, respectively, $\mathbf{R}$ (observational error covariance matrix) and $\mathbf{H^T}$

Results

In this section we compare the two methods introduced in the preceding using the very simple case of a reactive-diffusive equation with homogeneous Dirichlet boundary conditions.

One is interested in the model : $x \in]0,1[$, $t > 0$, $\nu = 1.1$, $\alpha = 0.9$

$$\begin{cases} \dfrac{\partial u}{\partial t} + \alpha\dfrac{\partial u}{\partial x} - \nu\dfrac{\partial^2 u}{\partial x^2} = g \\ u(0,t) = u(1,t) = 0 \\ u(x,0) = \sin \pi x \end{cases} \tag{11}$$

for which an exact solution is known.

Let us choose the solution of (11) to be $u = \mathrm{e}^{-t}\sin \pi x$. Then the exact right hand side member expression is $g = \mathrm{e}^{-t}(\pi\alpha \cos \pi x + (-1 + \nu\pi^2)\sin \pi x)$.

One will implement the second method by considering the model:

$$\begin{cases} \dfrac{\partial u}{\partial t} + \alpha\dfrac{\partial u}{\partial x} - \nu\dfrac{\partial^2 u}{\partial x^2} = g + f(\eta) \\ u(0,t) = u(1,t) = 0 \\ u(x,0) = \sin \pi x \end{cases} \tag{12}$$

Numerical results : part I

Both 4D-Var and MEC 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 Piacentini et al [15]), therefore it has been of a great help for the construction and the handling of this two methods.

In order to compare the effectiveness of our two methods we will act, initially, on three parameters: by disturbing the forcing term g (either by a correlated noise - a multiplicative factor -, or by an uncorrelated white noise) and by under-sampling the observations. The results obtained by the model errors control method (MEC) will be compared:

- With the results obtained without control.
- With the results obtained by controlling only the initial condition.
- With the exact solution $e^{-t}\sin(\pi x)$.

During these experiments, we will use 5000 time steps ($\Delta t = \frac{1}{5000}$) and 20 discrete space steps ($\Delta x = \frac{1}{20}$).

The minimization method used here is a home made conjugate gradient and the Hessian of the cost function is used as a preconditioner. The Hessian vector product is computed thanks to the second order adjoint (see Charpentier and Le Dimet [4]). The stopping criterion is $\| \nabla J \| < 10^{-8}$.

Disturbing the right hand side member and the initial condition and under-sampling the observations

Instead of initializing the model with the exact initial condition $\sin \pi x$ and forcing by the second member corresponding to the exact solution, we will add a 10% white noise on both of them. Here the weight of the cost function has been fixed $N = 0$ (the observation are exact so we there is no necessity to restrict the control vector) and $B = Identity$.

In addition, one has, observations only for one time step over 100 and for one space discretization step over 2. Under these conditions the MEC algorithm converges in 25 iterations towards a result which remains very satisfying even for times which are not observed, whereas the traditional method converges in six iterations but towards a result definitely less interesting.

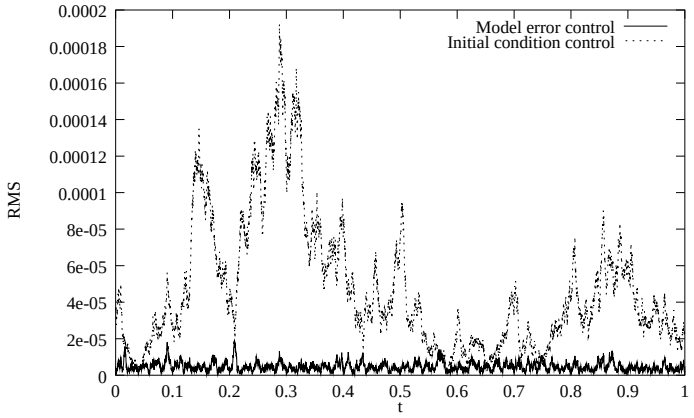

Figure 1: RMS according to time for both methods

A first assessment

This experiment shows us that, in the skewed model case the MEC method makes possible to redraw the system evolution accurately, which is not the case of the traditional method of control of the outline. It should be noticed that whatever the disturbance added to the 'exact' model, the results go in the right direction: what we saw here for an uncorrelated white noise remains true for a correlated noise(not shown).Unfortunately, the traditional method is already reaching the limit of current computational means , the computational over-cost makes any practical application difficult. However in the last experiment one sees that the under-sampling of the observations, which is always present in practice, reduces enormously the number of iterations.

It is true that the results obtained are more than encouraging. However until now, the observations were **exact** which obviously never occurs in practice.

Taking into account of the noises of observation

In the previous paragraph it is showed that the model error control algorithm gives us good results under acceptable disturbances of the second member or the initial state. The relaxation of the constraint imposed by the model makes the analyzed solution fit more closely to the observation. Thus one of the problems of this method comes from the inexactness of observations that are **never** exact in practice. Therefore, if we do not take care, the control of model error will make the model closely follow the evolution of observations . Thus a very noisy result will be obtained. This is illustrated by the figure 2 where we assimilate observations comprising a with a 1% amplitude white noise with no regularization term in the cost function ($N = 0$). The controlled solution tends to join the observations and then induce a calamitous result.

In practice it is thus necessary to avoid that the norm of the correction term η becomes too large, i.e. to increase the weight of regularization term ($N\eta$, η) in the cost function to be minimized. The choice of N may be made using the optimal control/Kalman filter analogy introduced in first section i.e $N \longleftrightarrow \mathbf{R}$ (where $\mathbf{R}$ is the observation error covariance matrix). The noise applied to observations being here a 1% uncorrelated white noise, thus a direct application of Kalman comparison e.g. $N \longleftrightarrow \mathbf{R}$ would give $N = 10^{-4} * I$. In practice, one obtains better results with a more significant weight.

Here two examples (figure 2) by noising second member only (10% white noise) with N=0 (left hand side) and $N = 4.10^{-2} * I$ (right hand side). Whereas in the first case the solution is of impressive bad quality, a more judicious choice for matrix N makes possible to obtain an acceptable result (error bounded with 5.10^{-5}).

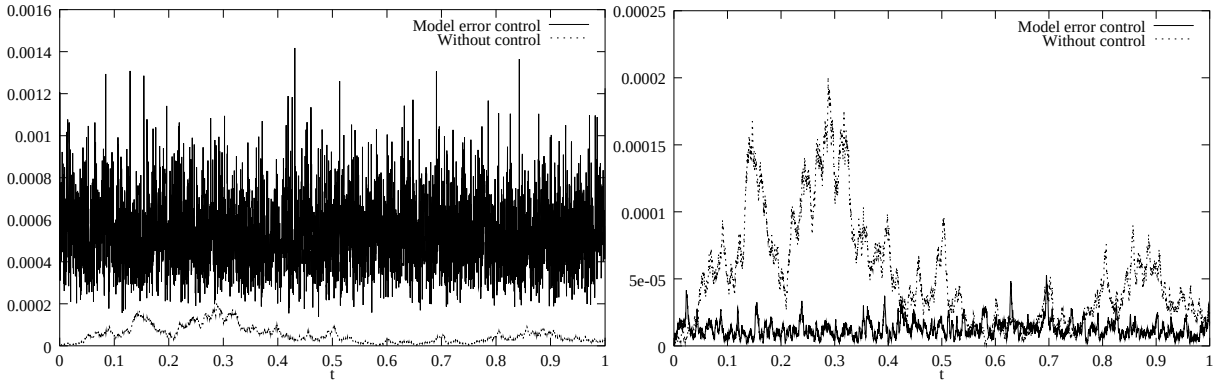

Figure 2: RMS according to time for $N = 0$ and $N = 4.10^{-2}$*Identity

Let us make a further step towards the real applications with partial coverage of observations(cf fig 3) the method gives catastrophic results. If we remember the relationship with the filter Kalman we take $B = \mathbf{H}^T$. The results so obtained are convenient again (fig 4 and 5). It can be noticed however that, at the end of the period, results are not satisfactory. That undoubtedly due to the fact that, for large value of t, the errors without control are already very small (about 10^{-5}) and the method 'hesitate' between minimizing the gap with observations and the regularization term .

Making N vary we can obtain different results over the various periods. A better knowledge of the observation error would undoubtedly make possible to refine the choice of N.

Moreover taking $B = \mathbf{H}^T$ has the interest to reduce the size of the control vector . Indeed this becomes of the size of the observation vector.

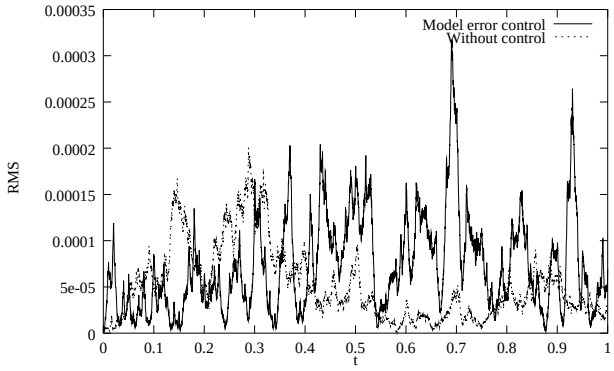

Figure 3: RMS according to time for $N = 10^{-3}$*Identity and observing only one time step per 50, with B =Identity

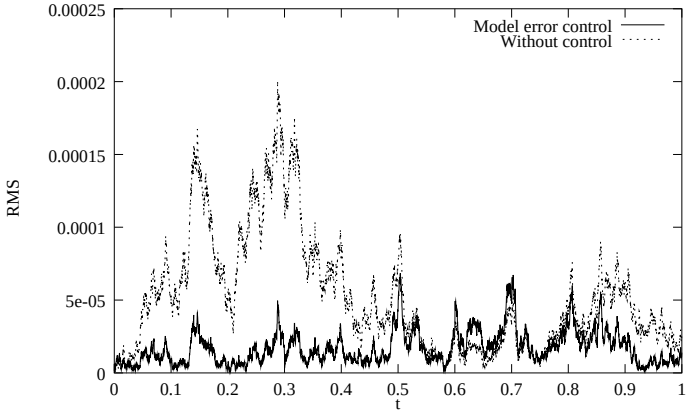

Figure 4: RMS according to time for $N = 10^{-3}$*Identity and observing only one time step per 50, with $B = \mathbf{H}^T$

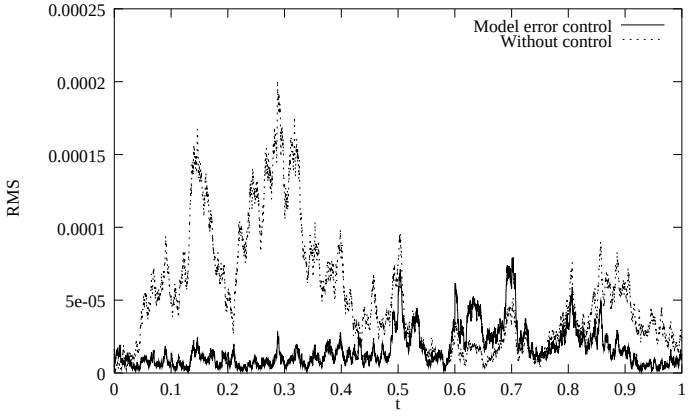

Figure 5: RMS according to time for $N = 5.10^{-4}$*Identity and observing only one time step per 50, with $B = \mathbf{H}^T$

Size reduction of control vector

Principle

Until now there was in our case a control vector η with size $20 \times 5000 = 100000$. By considering the simplicity of our model, one can imagine what that would give for models of meteorology or oceanography (typically 10^8 - 10^{10}). It is thus necessary, for obvious reasons of memory size and computing cost, to reduce the size of this vector.

We are always interested in the problem

$$\begin{cases} \dfrac{d\mathbf{x}}{dt} = M(\mathbf{x}) + B\eta \\ \mathbf{x}(0) = \mathbf{x}_0 + \delta\mathbf{x}_0 \end{cases} \tag{13}$$

But we reduce the size of the control vector. For space contribution we only control on a subspace spanned by some main direction. The time dimension is reduced using the first terms of a Fourier series development.

The choice of the directions along which we control η is directly guided by the model. Let $\mathbf{x}_1$ be the solution to the model on which we added a small disturbance h at $t = t_i$, i.e.:

$$\begin{cases} \dfrac{d\mathbf{x_1}}{dt} = M(\mathbf{x}_1) + B\eta \\ \mathbf{x}_1(t_i) = \mathbf{x}_{t_i} + h \end{cases} \tag{14}$$

Then the error $\hat{\mathbf{x}} = \mathbf{x}_1 - \mathbf{x}$ will be solution of equation :

$$\begin{cases} \dfrac{d\hat{\mathbf{x}}}{dt} = \mathbf{M}_{|\mathbf{x}}\hat{\mathbf{x}} \\ \\ \hat{\mathbf{x}}(t_i) = h \end{cases} \tag{15}$$

It is easy to see that the directions h along which the error grows most rapidly are those maximizing the amplification factor $A(h)$

$$A(h) = \frac{< \mathbf{M}h, \mathbf{M}h >}{< h, h >} = \frac{< \mathbf{M}^T\mathbf{M}h, h >}{< h, h >}$$

The operator $\mathbf{M}^T\mathbf{M}$ being symmetric the fastest growing perturbations are the eigenvectors of $\mathbf{M}^T\mathbf{M}$ associated with the largest eigenvalues (Lacarra and Talagrand [9]). Thus, if we decompose the error in the base spanned by these eigenvectors, the components associated with the

eigenvectors with eigenvalues higher than 1 will be amplified by the model during its integration while the others tend to vanish.

Application to our problem

We can use this result in the following way:

- We choose the first m eigenvectors χ_i of ${\mathbf{M}_{|\mathbf{x}}}^T\mathbf{M}_{|\mathbf{x}}$ corresponding to the m greatest eigenvalues of this matrix . For example, we can see in fig (6 what we have in our case eigenvalues of ${\mathbf{M}_{|\mathbf{x}}}^T\mathbf{M}_{|\mathbf{x}}$ for 20 space discretization steps.

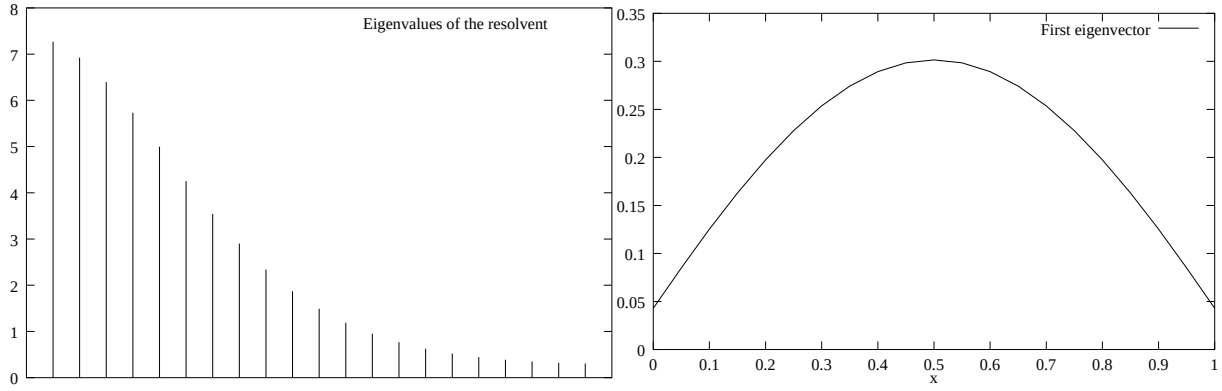

Figure 6: Eigenvalues of $\mathbf{M}_{|\mathbf{x}}^{\mathbf{T}}\mathbf{M}_{|\mathbf{x}}$ and the first eigenvector

  Eigenvectors and associated eigenvalues are computed using a Householder reduction method and a QL algorithm.

- We choose a time dependent function basis $\lambda_j(t)$, for instance the l first terms of a development in Fourier series.

$$\lambda_j(t) = \sin(2\pi jt) \text{ or } \cos(2\pi jt)$$

And we seek a model error of the form

$$\eta(x,t) = \sum_{i=1}^{m}\sum_{j=1}^{l} \alpha_{ij}\lambda_j(t)\chi_i(x)$$

For the sake of simplicity, we note $\eta = \Lambda(t)\alpha$ where α is the column vector made of α_{ij} and Λ is the matrix the columns of which are the $\lambda_j\chi_i$ products.

Now we control only $\alpha_{ij}\ \forall i,j$. Then the cost function becomes:

$$J(\alpha) = \frac{1}{2}\left(\int_0^T \parallel \mathbf{Hx} - \mathbf{x}_{obs} \parallel^2 + (N\Lambda\alpha, \Lambda\alpha)dt\right)$$

Gradient computation

In the same way as for the complete methods, and using the same adjoint model, one obtains the gradient:

$$\nabla J = \int_0^T (\Lambda^T\mathbf{x}^* + \Lambda^T N\Lambda\alpha)dt$$

Numerical results: Part II

The experiment conditions are the same as before , i.e a 10% uncorrelated white noise applied on the second member and a 1% white noise on the observations. The choice of the number of eigenvectors and time functions is not obvious because the more is not necessary the best: a small number makes possible to avoid to restrict control freedom but choosing it too small prevents it from correcting sufficiently. Empirically: with the reading of table 1 it appears that a good compromise is about 2-3 eigenvectors and about fifteen time functions. Thus we have brought back the size of the control vector from 100000 to 30!! We can notice (see fig 7 right) that with this choice one obtains results slightly better than the method without reduction! This result is quite surprising but could be explained : the decomposition onto a reduced basis might filter the observation errors and then gives better results. Recall that the stopping criterion is $\| \nabla J \| < 10^{-8}$.

Number of time functions	Number of eigenvector							
	2		3		4		5	
5	24	$1.54e^{-5}$	36	$1.50e^{-5}$	46	$1.51e^{-5}$	57	$1.53e^{-5}$
10	28	$1.4e^{-5}$	44	$1.41e^{-5}$	59	$1.45e^{-5}$	74	$1.49e^{-5}$
15	**39**	$\mathbf{1.33}e^{-5}$	53	$1.35e^{-5}$	63	$1.40e^{-5}$	84	$1.45e^{-5}$
20	47	$1.35e^{-5}$	57	$1.4e^{-5}$	82	$1.46e^{-5}$	87	$1.53e^{-5}$

Table 1: Number of iterations and RMS for some value of the number of eigenvector and the size of time basis

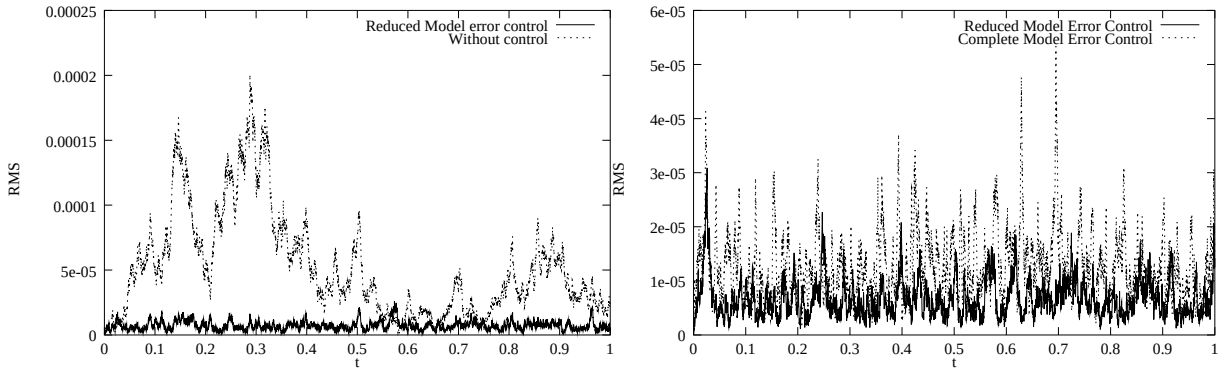

Figure 7: RMS according to time with 15 time function and 2 eigenvector

As we have seen before, the method tolerate the under-sampling combined with noised observations. It is significant to note that the choice of 'good' matrices N and B gives almost **exactly** the same result as in complete methods (fig 8) for half less iterations.

Conclusions

The new method introduced in the present article provides a strong enhancement of the classical 4D-Var. On the simple example of the reactive-diffusive equation, it gives better results tolerating

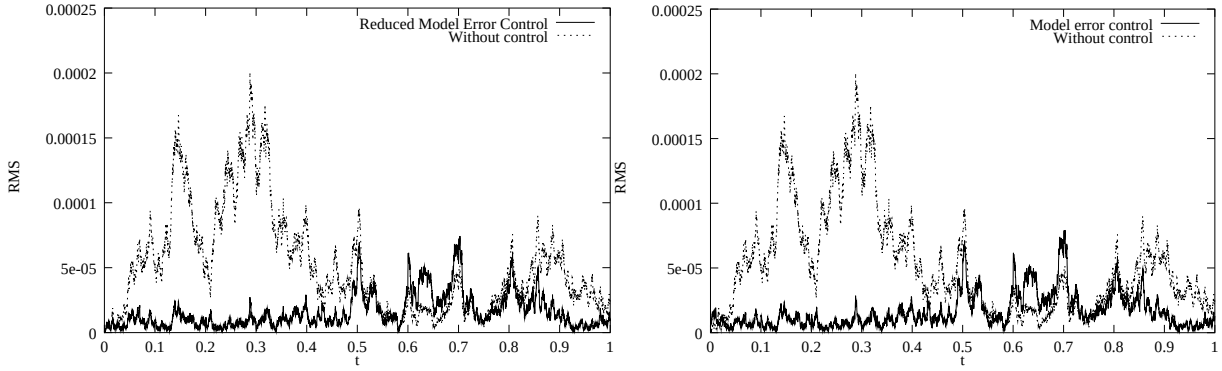

Figure 8: RMS according to time with 5000 time steps with noisy under-sampled observations. Comparison entire/reduced MEC method

an inexact model. Moreover a longer assimilation window does not alter its efficiency. The entire version of the Model Error Control method does not allows realistic applications due to its huge size of control vector but the reduction of the size of control permits to implement this methods with a very little extra cost compared to 4D-Var.

Although this first results are encouraging, many further development have to be made for realistic applications. Other basis of preferred directions (Liapunov, breeding, EOF) and other time functions (wavelet) have to be tested.

References

[1] F. D'andrea, R. Vautard : **Reducing Systematic Error by Empirically Correcting Model Errors** Tellus A, *1999*

[2] A.F. Bennett : **Inverse Methods and Data Assimilation** Lecture notes of Oregon State University summer school *1997*

[3] Eric Blayo, Jacques Blum, Jacques Verron : **Assimilation Variationnelle de Données en Océanographie et Réduction de la dimension de l'Espace de Contrôle** in Equations aux Dérivées Partielles et Applications, p 199-219, Gauthier-Vilars, *1998*

[4] Isabelle Charpentier, François-Xavier Le Dimet : **Méthodes du second ordre en assimilation de données**in Equations aux Dérivées Partielles et Applications, Gauthier-Vilars, *1998*

[5] S.E. Cohn : **Estimation theory for data assimilation problems: Basic conceptual framework and some open questions.** J. Meteorol. Soc. Japan, 75, 1B, 257-288, *1997*

[6] D.P. Dee, A.M. Da Silva: **Data assimilation in presence of forecast bias**, Q. J. R. Meteorol. Soc. 124, 269-295, *1998*

[7] A. K. Griffith, N. K. Nichols: **Accounting for model error in data assimilation using adjoint methods.** Computational Differentiation : Techniques, Application and tools, SIAM, pp 195-204 *1996*

[8] K. Ide, P. Courtier, M. Ghil, A.C. Lorenc : **Unified notation for data assimilation: operational, sequential, and variational.** J. Meteor. Soc. Japan, 751B, 181-189, *1997*

[9] J-F Lacarra, O. Talagrand: **Short-Range evolution of small perturbation in a barotropic model.** Tellus 40A, 81-95, *1998*

[10] J.-M. Lellouche, J.-L. Devenon, I. Dekeyser : **Boundary control of Burgers' Equation-A Numerical Approach** Computers Math Applic. 28 pp.33-44, *1994*

[11] A. C. Lorenc : **Analysis methods for numerical weather prediction.** Q. J. Roy. Meteorol. Soc. 112, 1177-1194,*1986*

[12] Bruno Luong, Jacques Blum, Jacques Verron : **Variational assimilation of altimeter data into a non-linear ocean model : temporal strategies** ESAIM Proceedings, *1998*

[13] Jacques-Louis Lions : **Optimal Control of Systems Governed by Partial Differential Equations**, Springer-Verlag,*1970*

[14] Richard Ménard : **Kalman Filtering of Burgers' Equation and its Application to Atmospheric Data Assimilation** PhD thesis, McGill University, Montréal *1993*

[15] Andrea Piacentini and the PALM group : **PALM : A modular data assimilation system** in proceeding of the third WMO symposium on data assimilation, Québec *1999* in press.

[16] George N. Saridis : **Stochastic Processes, Estimation and Control. The entropy approach** Wiley-interscience *1995*

[17] Olivier Talagrand : **The use of adjoint equation in numerical modeling of the atmospheric Circulation.** in SIAM workshop on Automatic differentiation of algorithms : theory, implementation and application, Beckenridge, Colorado, *1991*

[18] Patrick Vidard : **Assimilation de Données avec contrôle de l'erreur modèle.** DEA report Univ. Grenoble 1 *1998*

[19] X. Zou, I. M. Navon, F. X. Le Dimet : **An optimal Nudging Data Assimilation Scheme using Parameter Estimation.**, Q. J. R. Meteorol. Soc. *1992*, 118, pp. 1163-1186

[20] D. Zupanski: **A general weak constraint applicable to operational 4DVAR data assimilation system.** Mon. Weather. Rev. 125, 2274-2292