Model Error in Variational Data Assimilation

Yannick Trémolet

ECMWF
Shinfield Park, Reading RG2 9AX, United Kingdom

1 Introduction

Atmospheric models which are used in data assimilation are not perfect. Taking into account the errors of the model has been studied in the context of sequential data assimilation but has generally been ignored in variational data assimilation. We present here a general introduction to weak constraints 4D-Var which explicitely takes into account model error and some preliminary results for its implementation in the ECMWF data assimilation system.

2 Variational Data Assimilation

Two sources of information are available to perform data assimilation: theoretical knowledge of the atmosphere and observations. Our theoretical knowledge of the system can be represented, on one hand by equations governing the physical state of the system x:

$$\mathscr{G}(x) = 0,$$

and on the other hand, by equations relating the state of the system x to the observations y:

$$\mathscr{H}(x) = y.$$

Taking into account the uncertainties, these equations can be rewritten as:

$$\begin{aligned} \mathscr{G}(x) &= \varepsilon_g, \\ y - \mathscr{H}(x) &= \varepsilon_h, \end{aligned}$$

where ε_g and ε_h are the errors associated to $\mathscr{G}$ and $\mathscr{H}$. Combining the two sources of information, the *a posteriori* probability distribution for the state x given the observations y is (from Bayes theorem):

$$P(x|y) = \alpha \exp\left(-\frac{1}{2}[y - \mathscr{H}(x)]^T R^{-1}[y - \mathscr{H}(x)] - \frac{1}{2}\mathscr{G}(x)^T C_g^{-1}\mathscr{G}(x)\right)$$

where C_g and R represent the covariance matrices of the errors ε_g and ε_h. The problem of finding the maximum of the probability distribution can be replaced by the problem of finding the minimum of:

$$J(x) = -\ln(P_a(x)) = \frac{1}{2}[y - \mathscr{H}(x)]^T R^{-1}[y - \mathscr{H}(x)] + \mathscr{G}(x)^T C_g^{-1}\mathscr{G}(x).$$

In meteorology, a prior estimate of the state of the system (background x_b) is usually known, with error ε_b and background error covariance matrix B. Taking it into account, the cost function becomes:

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}[y - \mathscr{H}(x)]^T R^{-1}[y - \mathscr{H}(x)] + \frac{1}{2}\mathscr{F}(x)^T C_f^{-1}\mathscr{F}(x)$$

where $\mathscr{F}$ represents the remaining theoretical knowledge after background information has been accounted for. This type of formulation was introduced very early on by Sasaki (1970). It is important to note that it is very general and that no hypothesis regarding x has been made yet.

2.1 Simplifying hypothesis

A variety of choices of control vector x are possible. For example, x can be chosen to be the 3D state of the atmosphere at analysis time. In that case, $\mathscr{F}$ might include balance constraints and $\mathscr{H}$ is a (sophisticated) 3D operator. This algorithm is known 3 dimensional variational data assimilation or **3D-Var**.

Another possible choice is to consider x as the 4D state of the atmosphere during the assimilation window with the additional assumption that the forecast model is perfect. x is only a function of the initial condition x_0 and the size of the control variable is reduced according to:

$$x_i = \mathscr{M}_i(x_{i-1}) = \mathscr{M}_{0,i}(x_0).$$

where $\mathscr{M}$ is the model representing the evolution of the atmospheric flow and $\mathscr{M}_{i,j}$ represents the model integration from time-step i to time-step j. $\mathscr{F}$ now includes constraints other than the model (digital filter initialisation for example). $\mathscr{H}$ is a (sophisticated) 4D operator, accounting for the time dimension which makes it possible to use observations at the correct time and to account for serially correlated observations. This approach is known as **strong constraint 4D-Var** or, very often, simply **4D-Var**.

3D-Var and 4D-Var implementations were made possible by the use of the adjoint technique introduced by LeDimet and Talagrand (1986) which allows to evaluate the gradient of the cost function at a reasonable cost by one backward integration of the adjoint model. This is the algorithm which is used in most operational implementation of 4D-Var.

2.2 Weak constraint 4D-Var

We now consider that x is a four-dimension control vector, representing the three-dimension state of the atmosphere over the assimilation period. The successive three-dimension states are not independent and have to satisfy the equations governing the evolution of the atmospheric flow:

$$x_i = \mathscr{M}_i(x_{i-1})$$

where $\mathscr{M}_i$ is the nonlinear atmospheric model. In practice, these equations are known and solved inexactly, both because of our imperfect knowledge of the atmosphere and because of the discrete representation being used. x verifies:

$$x_i = \mathscr{M}_i(x_{i-1}) + \eta_i \tag{1}$$

where η_i is the model error at time-step i. We can define $\mathscr{F}(x)$ in the last term of the cost function by:

$$\mathscr{F}_i(x) = x_i - \mathscr{M}_i(x_{i-1}).$$

The choices of control vector $x = \{x_i\}_{i=0,\ldots,n}$ and $\chi = \{x_0, (\eta_i)_{i=1,\ldots,n}\}$ are equivalent according to equation (1). For small perturbations of the control vector, equation (1) can be linearised and the perturbation evolves according to:

$$\delta x_i = M_i \delta x_{i-1} + \delta\eta_i = M_{0,i}\delta x_0 + \sum_{j=1}^{i} M_{j,i}\delta\eta_j$$

where M is the linearised model. The corresponding variation of the total cost function is:

$$\begin{aligned}\delta J(x_0,\eta) \quad = \quad & \delta x_0^T B^{-1}(x_0 - x_b) + \delta x_0^T \sum_{i=1}^{n} M_{i,0}^T H_i^T R_i^{-1} d_i \\ & + \sum_{j=1}^{n} \delta\eta_j^T \sum_{i=j}^{n} M_{i,j}^T H_i^T R_i^{-1} d_i + \sum_{j=1}^{n}\sum_{k=1}^{n} \delta\eta_j^T Q_{j,k}^{-1}\eta_k\end{aligned}$$

The gradient $\frac{\partial J}{\partial \eta_j}$ is obtained by accumulating contributions from the adjoint at steps $i = j, \ldots, n$ and the total gradient of the cost function is obtained by one backward integration of the adjoint, as was the case in strong constraints 4D-Var.

3 Weak constraint 4D-Var: Practice

3.1 Size of the problem

At the current operational resolution at ECMWF, the 4D-Var control variable is represented in spectral space at the horizontal truncation of T159, with 60 levels in the vertical and a time-step of 30 minutes over a 12 hour assimilation window. For the operational strong constraints 4D-Var, this leads to a control variable of size 7.7×10^6. In the case of weak constraints 4D-Var, the size of the control variable would increase to 1.9×10^8. The corresponding model error covariance matrix would have 1.9×10^{16} elements and occupy 131,331 Tb of memory. This cannot be achieved on today's supercomputers.

In addition to this technical problem, one has to consider the following: 3×10^6 observations are available each day to estimate the 1.9×10^{16} elements of Q. Assuming that all parameters of Q are observable, that there is no redundancy in observed quantities and that model error can be separated from other sources of error, it would take 6 billion years to gather as many observations as there are parameters in Q. To gather meaningful statistics, it would require an order of magnitude more data. There is clearly not enough information available to solve this problem even if we had enough computer power.

3.2 Representing Model Error

Model error comes from several sources, some of which are constant (errors related to orography) while others can be almost periodic (errors related to diurnal cycle) or flow dependent (errors in physical processes). Discretisation and numerical errors may be more random. Overall, model error is correlated in time (in addition to growth). In the incremental 4D-Var context, model error as seen by 4D-Var also includes the error between the inner and outer loops which is the difference between the high resolution nonlinear model and the low resolution linearised model with limited physics used for the minimisation of the cost function. This is not negligible as shown by Trémolet (2003).

The simplest possible approximation is to consider that model error is constant over the assimilation period. Other choices have been proposed such as the use of a Markov chain by Zupanski (1997):

$$\eta_i = \frac{\mu}{\mu + (1-\mu^2)^{1/2}} \eta_{i-1} + \frac{(1-\mu^2)^{1/2}}{\mu + (1-\mu^2)^{1/2}} r_k$$

where r_k is a random variable at lower resolution than η_i. Another possible choice would be a Fourier series expansion which would represent the diurnal cycle well as proposed by Griffith and Nichols (1998). One could also want to include a vanishing term to represent model spin-up/down.

In the example presented below, we will assume that model error is a constant forcing over the assimilation window, keeping in mind that other, less restrictive assumptions will have to be considered in the future.

3.3 Model Error Covariance Matrix

As for the choice of control variable, many approximations are possible to define the model error covariance matrix. Statistics for model error could be defined from the model's implementation, taking into account uncertainties in each aspect of the model (numerics and all physical processes). A practical approach is to use an approximation based on the background error covariance matrix B as described by Zupanski (1993). Other possibilities have been proposed such as the one based on statistics on innovation and residual by Daley (1992) which relies on Kalman filtering and is not (easily) applicable to 4D-Var. It is also possible to described model error and the associated statistics by an ensemble of slow modes as introduced by Phillips (1986) and Cohn and Parrish (1991). Once a parameterisation for Q has been chosen which depends on some parameters α, it is possible to determine the parameters online as shown by Dee (1995). The Lanczos algorithm which is already used in the 4D-Var minimisation at ECMWF could also be used for that purpose (M. Fisher, personal communication).

4 Preliminary results

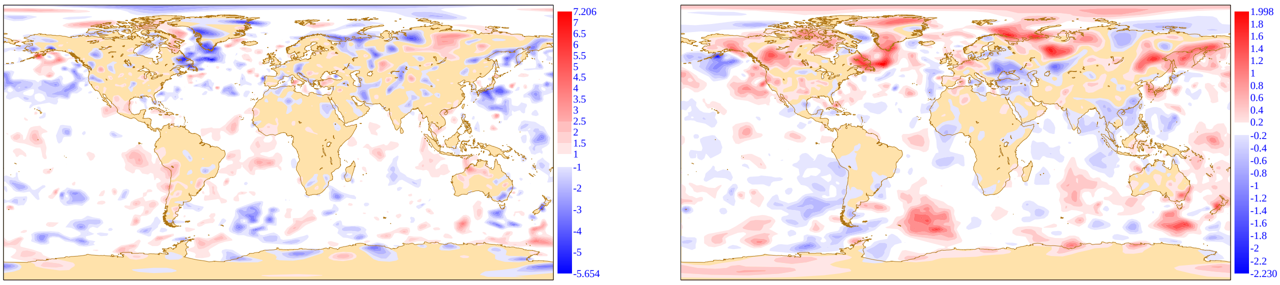

Figure 1: Surface pressure model error obtained with constant forcing and $Q = 10^{-1}B$ on the left and corresponding impact on the initial condition increment on the right.

The left panel of figure 1 shows the surface pressure component of model error obtained in the IFS in the simple case where model error is represented by a constant forcing term and the model error covariance matrix is chosen as $Q = \alpha B$ with $\alpha = 10^{-1}$. The right panel on the figure shows the change in the surface pressure initial condition when model error is used. Although the scale is different on both panels, the patterns in both are very similar and of opposite signs. In fact, the cumulated effect of model error forcing will compensate for the difference in initial condition, and there is no visible impact in the forecast (not shown). This result is disappointing but can be explained by looking at the impact of model error on a forecast.

4.1 Model Error Evolution

In this section, we try to evaluate the impact of model error forcing on the forecast. Knowing x_0 and η, x_i can be computed for all time-steps according to:

$$x_i = \mathscr{M}_i(x_{i-1}) + \eta_i = \mathscr{M}_i\left(\mathscr{M}_{i-1}\left(\ldots\left(\mathscr{M}_1(x_0) + \eta_1\right)\ldots\right) + \eta_{i-1}\right) + \eta_i$$

If model errors are small enough, $\mathscr{M}_i$ can be linearised and we obtain:

$$x_i = x_i^m + \sum_{j=1}^{i} M_i \ldots M_{j+1} \eta_j \tag{2}$$

where $x_i^m = \mathscr{M}_i(\ldots(\mathscr{M}_1(x_0))\ldots)$ is the perfect model forecast. Early components of model error will dominate the resulting impact on the forecast for two reasons: they have an impact on all the later steps and this impact can grow with time according to the model dynamics. In a weak-constraints 4D-Var, this means that the cost function will be dominated by the early components of model error, or in other terms, that the early component will be better determined. If we add the hypothesis that model error is constant over the assimilation window, it becomes clear the model error will mainly represents the early errors.

In the incremental formulation of 4D-Var, we can write x_i as a function of the initial condition increment δx_0 and model error:

$$x_i = \mathscr{M}_i\left(\mathscr{M}_{i-1}\left(\ldots\left(\mathscr{M}_1(x_b+\delta x_0)+\eta_1\right)\ldots\right)+\eta_{i-1}\right)+\eta_i$$

This expression can be linearised in the same way as (2), we obtain:

$$x_i = x_i^b + M_i \ldots M_1 \delta x_0 + \sum_{j=1}^{i} M_i \ldots M_{j+1} \eta_j$$

and δx_0 can be identified with η_0. In addition to that, since we assumed that Q was proportional to B, the initial condition increment δx_0 and the model error η are constrained in the same directions. Only their relative amplitudes which are controlled by α differ. As a consequence, they both predominantly retrieve the same information which explains the lack of impact in this preliminary experiment.

4.2 Model Bias control variable

We have considered so far that model error was defined by

$$\eta_i = x_i - \mathscr{M}_i(x_{i-1}).$$

It is also possible to define it relatively to the perfect model forecast:

$$\beta_i = x_i - x_i^m.$$

In that case too, we can simplify the problem by assuming β_i constant in time. This means that β represents the bias of the model. Since the model error term in the cost function is the sum of the difference at each time-step between the perfect model forecast and the corrected forecast, all the components in time now have equal influence on the cost function. It is a good representation of the time averaged error. One should note that the choices of control vector $x = (x_i)_{i=0,\ldots,n}$, $\chi = (x_0, (\eta_i)_{i=1,\ldots,n})$ or $\chi' = (x_0, (\beta_i)_{i=1,\ldots,n})$ are equivalent. However, the approximations $\eta_i = \eta$ and $\beta_i = \beta$ are not equivalent.

The table below summarises the possible choices of control variable and simplifying assumptions which can be made in variational data assimilation.

x_0	x_0	x_0	x_0	x_0
η_i or β_i	$\eta_i = \eta$	$\beta_i = \beta$	$\eta_i = 0$	$\eta_i = 0$
$x_i = \mathscr{M}_i(x_{i-1}) + \eta_i$ or $x_i = \mathscr{M}_{0,i}(x_0) + \beta_i$	$x_i = \mathscr{M}_i(x_{i-1}) + \eta$	$x_i = \mathscr{M}_{0,i}(x_0) + \beta$	$x_i = \mathscr{M}_i(x_{i-1})$	$x_i = x_0$
$\Downarrow$	$\Downarrow$	$\Downarrow$	$\Downarrow$	$\Downarrow$
Weak constraint 4D-Var	Constant Forcing	Model Bias	4D-Var	3D-Var

4.3 Computational cost

The cost per iteration of the weak constraints 4D-Var can be estimated. In the case of constant forcing, the size of the control vector is double that of the strong constraints case. The cost of the linear algebra is thus doubled (at most). We have already seen that one backward integration of the adjoint model gives access to the gradient of the cost function as was the case in strong constraints 4D-Var. The cost of adding the forcing in the linear and adjoint models and the cost of linear algebra is negligible when compared with the model integrations as can be verified in table 1.

The number of iterations of the minimisation algorithm is fixed in the first minimisation. In the subsequent minimisations, it depends on the conditioning which is used. ECMWF 4D-Var relies on an efficient preconditioning using the Lanczos algorithm and it is expected that the conditioning of the second and following minimisations will remain the same (M. Fisher, personal communication).

Elapsed time	Strong constraint	Weak constraint
Dot product	3.3 ms	4.9 ms
Cost function	89.2 sec	91.9 sec
Memory	703 Mb	769 Mb

Table 1: Measured elapsed time and memory for T159 minimisation with the IFS on IBM SP, 128 CPUs. The cost function times include the evaluation of the cost function and its gradient which are obtained by integrating the tangent linear and adjoint models.

5 Conclusion and future developments

Weak constraints 4D-Var has been described and a very preliminary application to ECMWF data assimilation system presented. It has been shown that this implementation is too crude to have a beneficial impact. However, the reasons for this have been identified and developments are under way to remove some of the excessive simplifications of this preliminary implementation.

Weak constraint 4D-Var is a generalisation of the more widely developed strong constraints 4D-Var where one simplifying assumption, namely the assumption that the forecast model is perfect, has been removed. In addition to lifting a questionable assumption, model error is valuable information which can be used in several ways. It can be added as forcing in the forecast model or at the post-processing stage if a model bias was determined. It might help identify model deficiencies and improve the model. Finally, model error term can be used in sensitivity computations as well as data assimilation and help determine whether bad forecasts are a consequence of errors in the initial condition or errors in the forecast model.

References

S. Cohn and D. Parrish. The behavior of forecast error covariances for a kalman filter in two dimensions. *Mon. Wea. Rev.*, 119:1757–1785, 1991.

R. Daley. Estimating model-error covariances for application to atmospheric data assimilation. *Mon. Wea. Rev.*, 120:1735–1746, 1992.

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

A. Griffith and N. Nichols. Adjoint methods for treating model error in data assimilation. In *Numerical Methods for Fluid Dynamics*, volume VI of *Institute for CFD, Oxford*, pages 335–344, 1998.

F.-X. LeDimet and O. Talagrand. Variational algorithms for analysis and assimilation of meteorological observations : Theorical aspects. *Tellus*, 38A:97–110, 1986.

N. Phillips. The spatial statistics of random geostrophic modes and first-guess errors. *Tellus*, 38A:314–332, 1986.

Y. Sasaki. Some basic formalisms in numerical variational analysis. *Mon. Wea. Rev.*, 98:875–883, 1970.

Y. Trémolet. Diagnostics of linear and incremental approximations in 4D-Var. *Accepted for publication in Q. J. R. Meteorol. Soc.*, 2003.

D. Zupanski. A general weak constraint applicable to operational 4dvar data assimilation systems. *Mon. Wea. Rev.*, 125:2274–2291, 1997.

M. Zupanski. Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Wea. Rev.*, 121:2396–2408, 1993.