TECHNICAL MEMORANDUM

520

Model error estimation in 4D-Var

Yannick Trémolet

Research Department

Submitted to Q. J. R. Meteorol. Soc.

March 2007

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications/

Contact: library@ecmwf.int

Abstract

Current operational implementations of 4D-Var rely on the assumption that the numerical model representing the evolution of the atmospheric flow is perfect, or at least that model errors are small enough to be neglected compared to other errors in the system. This paper describes a formulation of weak constraint 4D-Var that removes that assumption by explicitly representing model error as part of the 4D-Var control variable.

The consequences of this choice of control variable on the implementation of incremental 4D-Var are discussed. It is shown that the background error covariance matrix cannot be used as an approximate model error covariance matrix. Another model error covariance matrix based on statistics of model tendencies is proposed.

Experimental results are presented which show that this approach to accounting for and estimating model error does capture some known model errors and improves the fit to observations both in the analysis and in the background but it also captures part of the observation bias. We show that the model error estimated in this approach varies rapidly and cannot be applied to correct medium or long range forecasts. It is also shown that, because it relies on the tangent linear assumption for the entire assimilation window, the incremental formulation of weak constraint 4D-Var is not the best suited for long assimilation windows.

1 Introduction

Data assimilation comprises combining all available sources of information about the atmosphere to produce the best possible initial condition for weather forecasts. The European Centre for Medium-range Weather Forecasts (ECMWF) uses the four dimensional variational data assimilation system (4D-Var) described by Rabier *et al.* (2000), Mahfouf and Rabier (2000), Klinker *et al.* (2000) and Andersson *et al.* (2004). 4D-Var takes into account several sources of information to produce an estimate of the atmospheric state at analysis time. While errors in observations and background state are accounted for, the numerical model representing the evolution of the atmospheric flow is assumed perfect, or at least the model errors are assumed small enough to be neglected compared to other errors in the system.

As other aspects of the data assimilation process have progressed over the years, one might ask whether this assumption remains valid and whether the fact that we neglect model error degrades the quality of ECMWF's analysis and forecast. Is there any evidence of the presence of model error in the system or, is it still small enough that it remains legitimate to neglect it?

Following Dee (1995b), if the model is perfect and all other errors (observation error and initial conditions error) are Gaussian and independent, the expected value of the residual, that is the value of observation term J_o of the cost function at the end of the minimisation, should be equal to the number of observations p and thus satisfy:

$$(J_o)_{min}/p = 1.$$

When model error is present and Gaussian, this expression becomes:

$$(J_o)_{min}/p = 1 + s_N^2$$

where s_N expresses how much of the residual is due to accumulation of model error relative to that which is due to data error. This term depends on the length N of the assimilation window since, as it becomes longer, the accumulation of model error will become increasingly significant. Table 1 shows the evolution of $(J_o)_{min}/p$ with the length of the assimilation window in ECMWF's Integrated Forecasting System (IFS). The fact that the values are smaller than 1 is a sign of the overestimation of background error relative to observation error. As explained, for example, by Talagrand and Bouttier (1999) this is safer than underestimating it which could lead

Assimilation window	p	$(J_o)_{min}$	$(J_o)_{min}/p$
6h (4 cycles)	5931560	2749999.3	0.46
12h (2 cycles)	5917569	2865151.9	0.48
24h (1 cycle)	5879029	3119484.2	0.53

Table 1: Values of J_o at the end of the minimisation for several lengths of the assimilation window. The values were obtained by accumulating values from 1, 2 or 4 data assimilation cycles on 07-07-2004.

to a progressive drift of the assimilated fields away from observations. However, the table shows that $(J_o)_{min}/p$ increases with the length of the assimilation window which suggests that model error does affect ECMWF's data assimilation system.

Other evidence is given in section 4.2 (figure 10) which shows the fit of the analysis to observations over the length of the assimilation window. If the model were perfect, the fit should be constant in time since it would depend only on the accuracy of the observations. When model error is present, the model drifts away from the correct solution and the discrepancy with observations increases with time as explained, for example, by Talagrand (1998) and as can be seen on the figure.

There have been attempts to take model error into account in various data assimilation systems. In the context of Kalman Filter systems in particular by Dee (1995a) or Dee and Da Silva (1998). Bennett *et al.* (1996) proposed the representers method which reduces the size of the problem to the number of observations using the generalised inverse. The physical space assimilation system (4D-PSAS) defines the control variable in observation space and thus has the same property. When taking into account model error, the benefit of these algorithms would be greater since the control variable size would not change and this makes observation space algorithms attractive for taking into account model error. However, the size of the control variable is not the dominant factor in the cost of 4D-Var. The most expensive components of the 4D-Var system are the integrations of the nonlinear, tangent linear and adjoint models and observation operators. The overall cost thus mostly depends on the number of iterations performed during the minimisation, i.e. on the conditioning of the minimisation problem and not on the cost of the linear algebra in the minimisation algorithm itself. Courtier (1997) has shown that 4D-Var and 4D-PSAS are equivalent problems and that their conditioning is the same. Thus, in this study, we restrict ourselves to model error estimation in the 4D-Var context which is the current operational data assimilation system at ECMWF.

In that context, weak constraint 4D-Var has been introduced by Sasaki (1970) but has never been implemented fully with a realistic forecast model because of the computational cost and because of the lack of information to define and solve the problem. The underlying idea is that, since the model's equations are not exact, it is sufficient to satisfy them only approximately: they can be imposed as a weak constraint in the optimisation problem. Even with important approximations in the representation of model error itself and of model error statistics, promising results have been obtained, for example by Derber (1989), Zupanski (1993) or Zupanski (1997) with atmospheric models or Vidard *et al.* (2004) with an ocean model. Trémolet (2006) discusses several formulations of weak constraint 4D-Var. These formulations are being developed and evaluated at ECMWF. This paper presents some implementation details and experimental results with the model error forcing formulation.

The paper is organised as follows: section 2 is devoted to the theoretical formulation of variational data assimilation and weak constraint 4D-Var. Section 3 discusses the determination of a model error covariance matrix. Section 4 gives experimental results obtained in the IFS. Finally, some conclusions and perspectives are given.

2 Variational data assimilation

2.1 Four dimensional problem

The observations of the atmosphere represented by the vector $\mathbf{y}$ in observation space are one source of information about the state of the atmosphere. An observation operator $\mathcal{H}(\mathbf{x})$ represents knowledge of what the observations should be given the atmospheric state represented by the state variable $\mathbf{x}$. Errors in the observations and in the observation operator are assumed unbiased, Gaussian and uncorrelated with other sources of error. They are characterised by their covariance matrix $\mathbf{R}$.

A particular source of information available in meteorology is a prior estimate of the state of the system. In practice, in operational weather forecasting centres, it is a forecast from the most recent analysis. This represents prior knowledge we have about the state of the system without resorting to the current observations $\mathbf{y}$. The prior estimate of the mean of the state is represented by $\mathbf{x}_b$ and called background. We assume that background error is unbiased, uncorrelated with other errors in the problem and with background error covariance matrix $\mathbf{B}$.

Another source of information about the system is theoretical knowledge, represented by the equation $\mathcal{F}(\mathbf{x}) = 0$. In meteorological applications, $\mathcal{F}$ can include the equations governing the evolution of the flow as well as additional constraints such as balance equations or prior knowledge about the state of the system. Errors if $\mathcal{F}$ are assumed unbiased, Gaussian and uncorrelated with other sources of error. They are characterised by their covariance matrix $\mathbf{C}_f$.

Using these sources of information, four dimensional variational data assimilation comprises minimising the cost function:

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) \\ &+ \frac{1}{2}[\mathcal{H}(\mathbf{x})-\mathbf{y}]^T\mathbf{R}^{-1}[\mathcal{H}(\mathbf{x})-\mathbf{y}] \\ &+ \frac{1}{2}\mathcal{F}(\mathbf{x})^T\mathbf{C}_f^{-1}\mathcal{F}(\mathbf{x}) \end{aligned} \tag{1}$$

The cost function can be interpreted as a weighted measure of the distance from the state $\mathbf{x}$ to the various available sources of information, either observational or theoretical. It can be derived as a result of Bayesian estimation, more details on this result are presented for example by Jazwinski (1970) or Rodgers (2000).

The components of $\mathbf{x}$ are the physical variables describing the atmosphere (e.g. temperature, wind, humidity and surface pressure) discretised over the three spatial dimensions of the model's domain and the temporal dimension over the period for which observations are available. The assimilation window $[0,T]$ is discretised into $n+1$ time steps $\{t_i\}_{i=0,\ldots,n}$. The state vector $\mathbf{x}_i$ represents the three dimensional state of the atmosphere at time t_i. The observation operator will use the components of the state variable at the appropriate time to evaluate the observation term of the cost function and will make the most accurate use of available observations.

In practice, approximations are necessary in order to solve the variational data assimilation problem. In operational variational data assimilation implementations, model error is assumed small enough to be neglected compared to initial condition error and the atmospheric model is imposed as a strong constraint. The state variable is a solution of the model equation:

$$\mathbf{x}_i = \mathcal{M}_i(\mathbf{x}_{i-1})$$

where $\mathcal{M}_i$ represents the model describing the evolution of the atmospheric flow from time t_{i-1} to time t_i. The evolution of the atmosphere is then entirely determined by the initial condition $\mathbf{x}_0$, the control variable reduces to a three dimensional state and the constraint $\mathcal{F}$ disappears from the cost function. This reduction of the control variable, combined with the adjoint technique to compute the gradient of the cost function required by most minimisation algorithms, was introduced by Le Dimet and Talagrand (1986) and is usually referred to as strong constraint 4D-Var or simply 4D-Var. Even though the time dimension of the information brought by the observations and the forecast model is taken into account, the control variable is defined over a three dimensional space. The size of the control variable and the elimination of the model error covariance matrix make this algorithm achievable operationally with today's supercomputers.

A more general approach is to consider that the forecast model is not perfect. We can write:

$$\mathbf{x}_i = \mathcal{M}_i(\mathbf{x}_{i-1}) + \boldsymbol{\eta}_i \tag{2}$$

where $\boldsymbol{\eta}_i$ is the three dimensional model error at time step t_i. The constraint $\mathcal{F}$ can be defined by:

$$\mathcal{F}_i(\mathbf{x}) = \mathbf{x}_i - \mathcal{M}_i(\mathbf{x}_{i-1}).$$

In this formulation, the atmospheric model is only imposed as a weak constraint since the minimising solution $\mathbf{x}$ does not have to be an exact solution of the model. This formulation is known as weak constraint 4D-Var. In this case, $\mathbf{C}_f$ is the model error covariance matrix usually denoted by $\mathbf{Q}$, the associated term in the cost function will be denoted J_q. A more complete introduction to the formulation of 4D-Var accounting for an imperfect model is given by Trémolet (2006).

2.2 Control variable definition

Equation (2) defines a change of variable which allows the computation of $\mathbf{x}$ given $\mathbf{x}_0$ and $\boldsymbol{\eta} = \{\boldsymbol{\eta}_i\}_{i=1,\ldots,n}$ and vice versa. A discussion of various possible choices of control variable is given by Trémolet (2006). Here, we choose to use the initial condition $\mathbf{x}_0$ and model error forcing $\boldsymbol{\eta}$ as the control variable. There have been several attempts to take model error into account in variational data assimilation and this is the choice most authors have made.

At the current operational resolution of the ECMWF data assimilation system, implementing weak constraint 4D-Var as described above would require using a control variable with a dimension of the order of 10^9. Because of the computing cost and amount of information that would be required to solve this problem, it is not possible to implement weak constraint 4D-Var without simplifications in the representation of model error. A realistic compromise between the full four dimensional control variable and ignoring model error is necessary.

Model error can come from many sources. Some are systematic or even constant, like for example approximations in the definition of the orography used in the model. Systematic errors can also include errors coming from the discretisation of the equations based on resolution or the choice of coordinates. Some errors are sensitive to the particular state of the atmosphere, this is the case of errors in the physics of the model which is active only when certain conditions are met. Some errors in the radiation scheme or related to solar heating will be linked to the diurnal cycle and will be approximately periodic with a 24h period. Some errors are flow dependant and highly correlated in time although not constant because of the general circulation of the flow. On an even longer time scale, seasonal errors can be present in the model, like misrepresentation of sea-ice particularly in the spring and autumn, or errors affecting the stratosphere of polar regions in winter as described by McNally (2003).

Derber (1989) introduced the variational continuous assimilation scheme in which he controlled systematic model error rather than the initial condition. In this case, the model error control variable is the same size as the state variable and the size of the problem is unchanged. Zupanski (1993) kept the initial condition component of the control variable and defined model error as $\boldsymbol{\eta}_i = \lambda_i \boldsymbol{\Phi}$ where $\boldsymbol{\Phi}$ is a three dimensional control variable and

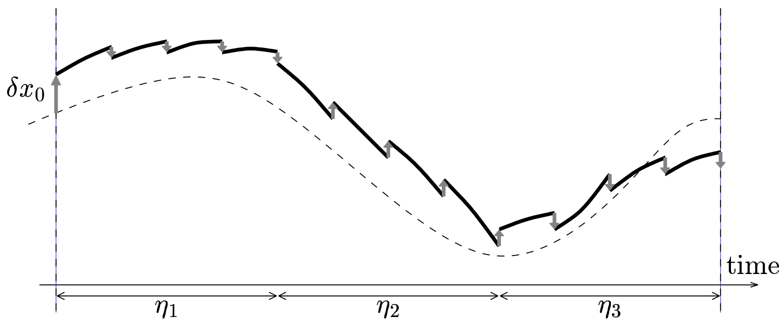

Figure 1: Weak constraint 4D-Var with model error forcing control variable. In the schematic example presented here, the forcing is constant for each of three intervals in the assimilation window (denoted $\boldsymbol{\eta}_1$, $\boldsymbol{\eta}_2$ and $\boldsymbol{\eta}_3$). It is applied at every time step, for illustrative purposes four times in each interval.

the λ_i are predefined coefficients describing the evolution in time of model error. Griffith and Nichols (1998) proposed to use a spectral representation of model error such as:

$$\boldsymbol{\eta}_i = \boldsymbol{\gamma}_0 + \boldsymbol{\gamma}_1 \sin(i\pi/n\tau) + \boldsymbol{\gamma}_2 \cos(i\pi/n\tau)$$

where τ is the timescale on which the model error is expected to vary, for example 24h and $\boldsymbol{\gamma}_0$, $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are three dimensional fields composing the control variable. Errors with longer timescale are represented by the constant term over the assimilation period. This approach could be generalised to any order of spectral expansion.

Another relatively simple approach is to consider model error to be constant by intervals. This is the approach chosen here. The two extreme cases being with only one interval which means that model error is constant for the whole assimilation window or, at the other end, when the interval becomes as short as a model time step, this is the full four dimensional problem.

The weak constraint 4D-Var cost function is defined in the most general form by equation (1). It can be written more explicitly as a function of the components of the control variable $\mathbf{x}_0$ and $\boldsymbol{\eta} = \{\boldsymbol{\eta}_i\}_{i=1,\ldots,n}$ as follows:

$$\begin{aligned} J(\mathbf{x}_0, \boldsymbol{\eta}) &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x}_0 - \mathbf{x}_b) \\ &+ \frac{1}{2}\sum_{i=1}^{n} \boldsymbol{\eta}_i^T \mathbf{Q}_i^{-1} \boldsymbol{\eta}_i \\ &+ \frac{1}{2}\sum_{i=1}^{n} [\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i]^T \mathbf{R}_i^{-1} [\mathcal{H}_i(\mathbf{x}_i) - \mathbf{y}_i] \end{aligned} \tag{3}$$

where $\mathbf{x}_i = \mathcal{M}_{i,0}(\mathbf{x}_0, \boldsymbol{\eta})$ represents state at time t_i resulting from the forced model integrated from time t_0 to time t_i and observation and model errors are assumed uncorrelated in time. Time correlation can be taken into account at the cost of using a non block diagonal model error covariance matrix and of determining the appropriate statistics. A better model for model error could also be used, for example Zupanski (1997) defined model error as a first order Markov variable in which the random component is defined on a coarser resolution both in time and space, assuming that there is no systematic error in the model.

2.3 Incremental formulation

At ECMWF and in other operational implementations, the incremental formulation of strong constraint 4D-Var introduced by Courtier *et al.* (1994) is used. In this approach, the nonlinear problem is treated as a succession

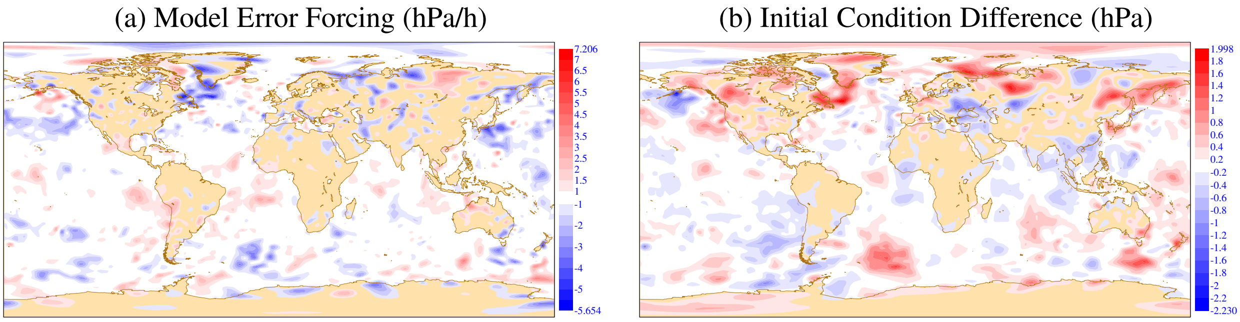

Figure 2: Surface pressure model error obtained with constant forcing and $Q = 10^{-1}B$ on 01-05-2004 at 0UTC on the left. The corresponding change in the initial condition increment is shown on the right.

of quadratic problems, approximating the nonlinear problem around a guess $\mathbf{x}^g$. This is achieved by using the initial condition increment $\delta\mathbf{x}_0 = \mathbf{x}_0 - \mathbf{x}_0^g$ as the control variable and by linearising the model and observation operator around the guess. This allows the use of very efficient minimisation algorithms such as the preconditioned Lanczos conjugate gradient method developed by Fisher (1998). The incremental strong constraint 4D-Var cost function is:

$$J(\delta\mathbf{x}_0) = \frac{1}{2}(\delta\mathbf{x}_0 + \mathbf{b})^T\mathbf{B}^{-1}(\delta\mathbf{x}_0 + \mathbf{b}) + \frac{1}{2}\sum_{i=0}^{n}(\mathbf{H}_i\mathbf{M}_{i,0}\delta\mathbf{x}_0 + \mathbf{d}_i)^T\mathbf{R}_i^{-1}(\mathbf{H}_i\mathbf{M}_{i,0}\delta\mathbf{x}_0 + \mathbf{d}_i)$$

where $\mathbf{M}_{i,0}$ represents the tangent linear model integrated from time t_0 to time t_i, i.e. $\mathbf{M}_{i,0} = \mathbf{M}_i\mathbf{M}_{i-1}\ldots\mathbf{M}_1$ and $\mathbf{M}_{0,0} = \mathbf{I}$, $\mathbf{M}_i$ and $\mathbf{H}_i$ are the tangent linear model and observation operator at time t_i, $\mathbf{d}_i = \mathcal{H}_i(\mathbf{x}_i^g) - \mathbf{y}_i$ and $\mathbf{b} = \mathbf{x}_0^g - \mathbf{x}_b$. The gradient of this cost function with respect to the initial condition increment is:

$$\nabla J_0 = \mathbf{B}^{-1}(\delta\mathbf{x}_0 + \mathbf{b}) + \sum_{i=0}^{n}\mathbf{M}_{i,0}^T\mathbf{H}_i^T\mathbf{R}_i^{-1}(\mathbf{H}_i\mathbf{M}_{i,0}\delta\mathbf{x}_0 + \mathbf{d}_i). \tag{4}$$

In practice, the rightmost terms in equation (4), $\tilde{\mathbf{y}}_i = \mathbf{H}_i\mathbf{M}_{i,0}\delta\mathbf{x}_0 + \mathbf{d}_i$, are computed sequentially in one forward integration of the tangent linear model and stored. The adjoint model is integrated backwards according to:

$$\delta\mathbf{x}_i^* = \mathbf{M}_{i+1}^T\delta\mathbf{x}_{i+1}^* + \mathbf{H}_i^T\mathbf{R}_i^{-1}\tilde{\mathbf{y}}_i$$

where $\delta\mathbf{x}^*$ is the adjoint variable and $\delta\mathbf{x}_n^* = \mathbf{H}_n^T\mathbf{R}_n^{-1}\tilde{\mathbf{y}}_n$. Using the linearity of the adjoint model, one shows that $\delta\mathbf{x}_0^*$ is equal to the sum in equation (4), i.e. to the gradient of the observation term of the cost function. Each evaluation of the cost function and its gradient are therefore obtained by one integration of the tangent linear model and one backward integration of the adjoint model.

The incremental 4D-Var formulation can be extended to include the model error part of the control variable $\boldsymbol{\eta}$. The control variable for the incremental formulation is the departure $(\delta\mathbf{x}_0, \delta\boldsymbol{\eta})$ from a first guess $(\mathbf{x}_0^g, \boldsymbol{\eta}^g)$. For small perturbations $\delta\mathbf{x}_{i-1}$ of $\mathbf{x}_{i-1}^g$ and $\delta\boldsymbol{\eta}_i$ of $\boldsymbol{\eta}_i^g$, the model can be linearised and the perturbation evolves according to:

$$\delta\mathbf{x}_i = \mathbf{M}_i\delta\mathbf{x}_{i-1} + \delta\boldsymbol{\eta}_i. \tag{5}$$

The gradients of the quadratic approximation of the cost function with respect to the initial condition increment

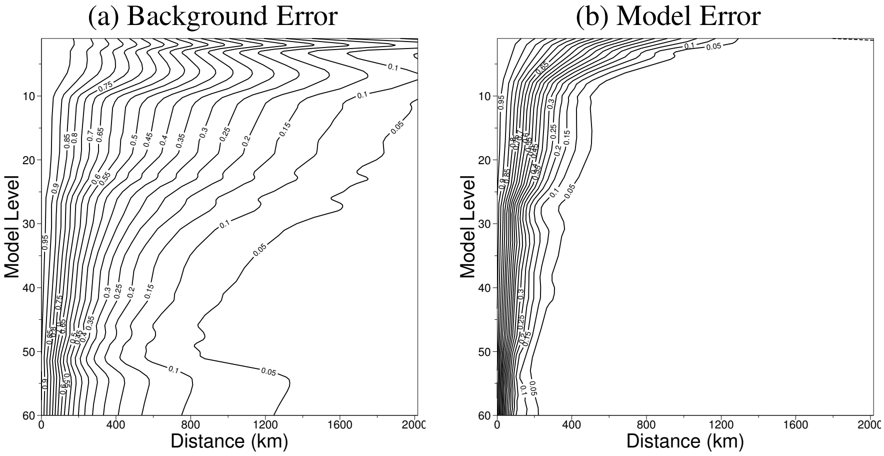

Figure 3: Temperature background and model error horizontal correlations. Contour interval is 0.05 for both figures.

and to the forcing increment at time t_i are respectively:

$$\nabla J_0 = \mathbf{B}^{-1}(\delta\mathbf{x}_0 + \mathbf{b}) + \sum_{j=0}^{n} \mathbf{M}_{j,0}^T \mathbf{H}_j^T \mathbf{R}_j^{-1}(\mathbf{H}_j \delta\mathbf{x}_j + \mathbf{d}_j)$$

$$\nabla J_i = \mathbf{Q}_i^{-1}(\delta\boldsymbol{\eta}_i + \boldsymbol{\eta}_i^g) + \sum_{j=i}^{n} \mathbf{M}_{j,i}^T \mathbf{H}_j^T \mathbf{R}_j^{-1}(\mathbf{H}_j \delta\mathbf{x}_j + \mathbf{d}_j)$$

where $\mathbf{M}_{i,j}$ represents the tangent linear model integrated from time t_j to time t_i and the nonlinear departures from observations $\mathbf{d}_j$ are evaluated from the guess using the forced nonlinear model. The gradient with respect to the initial condition is the same as in strong constraint 4D-Var except for the fact that the forward integration is carried out with the forced linear model defined by equation (5). The observation term gradient with respect to the forcing increment at time t_i has the same form as the gradient with respect to the initial condition increment. The difference is that the sum is restricted to the steps $j \geq i$. The gradient with respect to the forcing increment at time t_i results from the backward integration of the adjoint from the end of the window to time t_i. It is an intermediate value already available during the computation of the gradient with respect to the initial condition. The total gradient of the cost function is still obtained by one forward integration of the tangent linear model and one backward integration of the adjoint, as was the case in strong constraint 4D-Var. This requires only limited modifications to an adjoint model designed for use in a strong constraint 4D-Var system and the computational cost per iteration of the minimisation algorithm will be very similar, as already pointed out by Derber (1989).

Figure 1 gives an overview of the incremental implementation of this formulation of the weak constraint 4D-Var problem. The background is obtained from a forecast from an earlier analysis (dashed line). The minimisation gives a new estimate of the initial condition and of the model error forcing. The solution (thick solid line) seems discontinuous when the forcing is applied. The forcing is applied at every time step and should be interpreted as a source term in the equations to correct for errors in each time step. The solution is not more discontinuous in principle than any discrete solution of the model's equations. In practice, the solution may present small amplitude discontinuities if the forcing is not applied at every time step for practical reasons but they should remain small enough not to create spin-up problems. Furthermore, balance constraints can be incorporated in the model error covariance matrix as is already done for the background error covariance matrix.

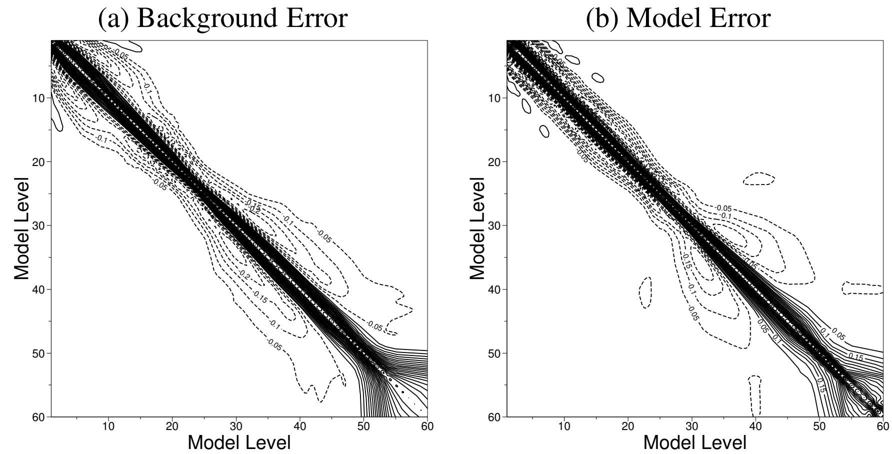

Figure 4: Temperature background and model error average vertical correlations. Contour interval is 0.05 for both figures.

3 Model error covariance matrix

3.1 Preliminary experiment

Very little is known about model error statistics and the model error covariance matrix. Implementing 4D-Var without any approximation in the current ECMWF data assimilation system would require a model error covariance matrix with of the order of 10^{18} elements. This is more than the total number of observations of the atmosphere taken since routine upper air observations started in the 1940s. Since approximately 6×10^6 observations are available each day, it would take 250 million years to gather as many observations as there are parameters in $\mathbf{Q}$. To gather meaningful statistics, it would require orders of magnitude more data, assuming that one could properly distinguish model error from other sources of error. There is clearly not enough information available to define this problem without important simplifications. Assuming a simplifying model can be found for the model error covariance matrix, as is already done for the background error covariance matrix, the number of parameters defining $\mathbf{Q}$ would be reduced and could be determined.

An approximation where the model error covariance matrix is chosen proportional to the background error covariance matrix $\mathbf{B}$ has been used in weak constraint 4D-Var studies, for example by Derber (1989), Zupanski (1993), Zupanski (1997) or Vidard *et al.* (2004). This choice can be justified by the fact that background error is a short term forecast error and thus includes a component of model error but also, and maybe mostly, from the practical fact that the background error covariance matrix is readily available.

Experiments were performed with the IFS in the simple case where model error is constant over the assimilation window and $\mathbf{Q} = \alpha\mathbf{B}$ with α a scalar chosen empirically. The left panel of figure 2 shows an example of the surface pressure component of model error obtained with $\alpha = 0.1$. The right panel shows the difference in surface pressure initial condition increment in this weak constraint 4D-Var experiment compared to the strong constraint 4D-Var reference. Although the scale is different on both panels, the patterns are very similar and of opposite signs. Similar results are obtained for other variables. There is little noticeable impact in the ensuing forecast, at least in the short term, as the cumulated effect of model error forcing approximately compensates for the difference in initial condition. In the longer term, the forecast is degraded as the forcing keeps acting on the model. It was also noted that the choice of α, within reasonable limits, does not change the shape of the solution but only its amplitude.

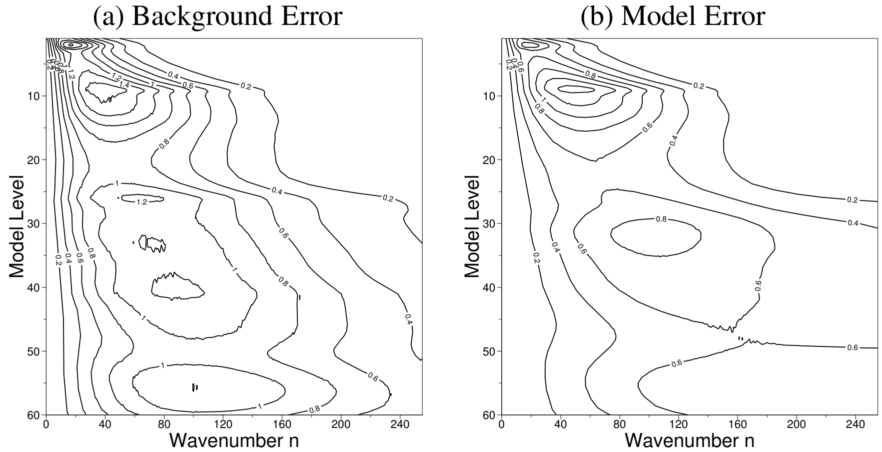

Figure 5: Divergence background and model error standard deviations as a function of wavenumber. Values are scaled by a factor of 10^{-6} *and contour interval is 0.2 for both figures.*

The temperature profile of the analysis, initial condition increment and model error forcing obtained using a model error covariance matrix proportional to $\mathbf{B}$ and another model error covariance matrix that will be described in the next section are shown on figure 9. In both cases, the initial condition increment shows oscillations in the stratosphere (level 30 and above). The profiles of model error forcing are however different. When the model error covariance matrix is proportional to $\mathbf{B}$, the forcing shows oscillations similar to those in the initial condition increment. As for surface pressure, the forcing is compensating for part of the initial condition increment. This result does not prove that the oscillations are caused by model error. They are believed to be the result of a combination of model bias in the stratosphere, vertical interpolations in the radiative transfer model which is part of the observation operator, the long tails of the radiance weighting functions in the vertical and the vertical structure of the $\mathbf{B}$ matrix.

In the incremental formulation of 4D-Var, a compact notation for the solution of the analysis equation is:

$$\delta\mathbf{z} = \mathbf{A}\mathbf{H}^T(\mathbf{R}+\mathbf{H}\mathbf{A}\mathbf{H}^T)^{-1}(\mathcal{H}(\mathbf{x}^g)-\mathbf{y})$$

$$\text{with } \mathbf{z} = \begin{pmatrix} \mathbf{x}_0 \\ \boldsymbol{\eta} \end{pmatrix} \text{ and } \mathbf{A} = \begin{pmatrix} \mathbf{B} & 0 \\ 0 & \mathbf{Q} \end{pmatrix}.$$

For this equation, $\mathbf{H}$ represents the generalised linearised observation operator, including the tangent linear model and function of $\mathbf{x}_0$ and $\boldsymbol{\eta}$. In this expression, $\mathbf{A}$ is the leftmost operator. This means that $\mathbf{B}$ and $\mathbf{Q}$ are the last operators to be applied in the expression for $\delta\mathbf{x}_0$ and $\delta\boldsymbol{\eta}$. The solutions thus lie in the subspaces spanned respectively by $\mathbf{B}$ and $\mathbf{Q}$. However, in this experiment, the model error covariance matrix $\mathbf{Q}$ was chosen proportional to the background error covariance matrix $\mathbf{B}$. The consequence is that the initial condition increment $\delta\mathbf{x}_0$ and the model error $\delta\boldsymbol{\eta}$ are constrained in the same directions. Model error is restricted to the same subspace as the initial condition increment, only their relative amplitudes which are controlled by α differ.

Being constrained in the same directions, the initial condition increment and the model error term both predominantly retrieve the same solution which explains the lack of impact in this preliminary experiment. More precisely, part of the initial condition increment is transfered into model error. The initial condition increment and the model error forcing could even compensate each other. In the worst case, they could even drift away from the truth in opposite directions. This choice of model error covariance matrix proportional to the background error covariance matrix has introduced two degrees of freedom in the same direction.

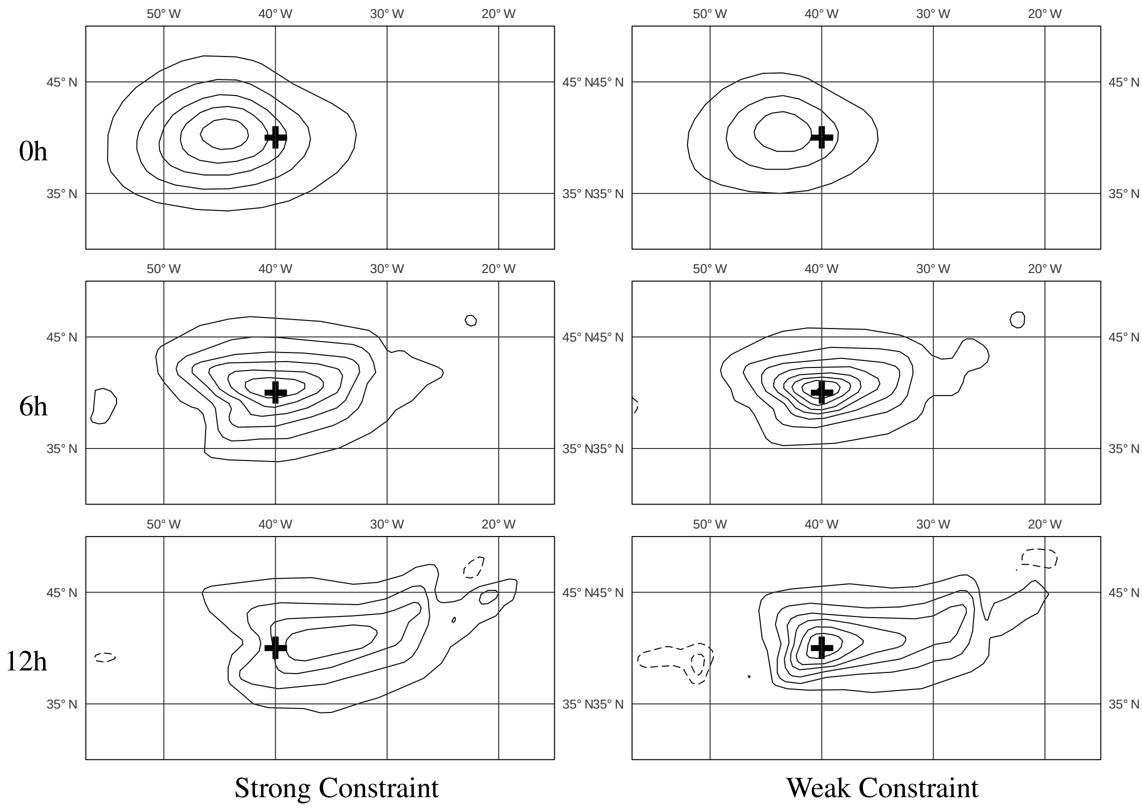

Figure 6: Evolved temperature increments for strong and weak constraint 4D-Var with temperature observations at every hour, located at 40N, 40W (as marked by the + sign) and with departures of 1K. Contour interval is 0.1K.

The discussion above shows that it is not possible to avoid determining a more appropriate model error covariance matrix. Using non proportional covariances matrices would give 4D-Var more freedom to explore various directions to fit the data. It is better to extend the space of directions being explored, even if the added directions are not optimal rather than constrain the algorithm in the directions already explored and thereby introduce under-determination along those directions.

3.2 Tendency based model error covariance matrix

In the current ECMWF data assimilation system, the background error covariance matrix $\mathbf{B}$ is estimated from an ensemble of 4D-Var assimilations as described by Fisher (2003). Considering the ensemble of forecasts run from the analyses produced by the 4D-Var ensemble, at any given step, the ensemble of model states represents a proxy for the probability distribution for the true atmospheric state. The model tendencies derived from the states of ensemble members should represent a distribution of the possible evolutions of the atmosphere from the true state. The differences between these tendencies can be interpreted as possible uncertainties in the model forcing or an ensemble of possible realisations of model error. This is the basis for constructing a model error covariance matrix.

This set of model error realisations can be fitted to a statistical model similar to the one used to represent $\mathbf{B}$. The statistical model used here is isotropic, homogeneous and non separable. It is the same model (and code) that was used to specify the background error covariance matrix for the operational strong constraint 4D-Var at ECMWF until April 2005 and described by Courtier *et al.* (1998) and references therein. Other covariance

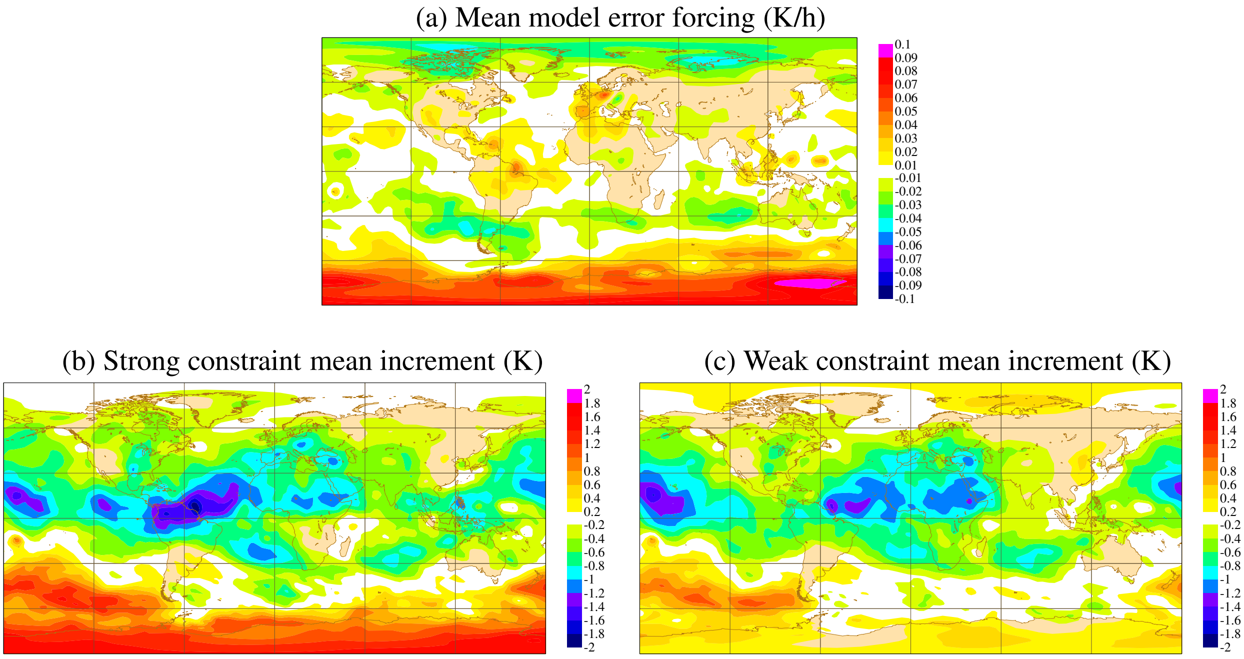

Figure 7: Average temperature model error forcing and initial condition increments for the month of July 2004 at model level 11 (≈5hPa).

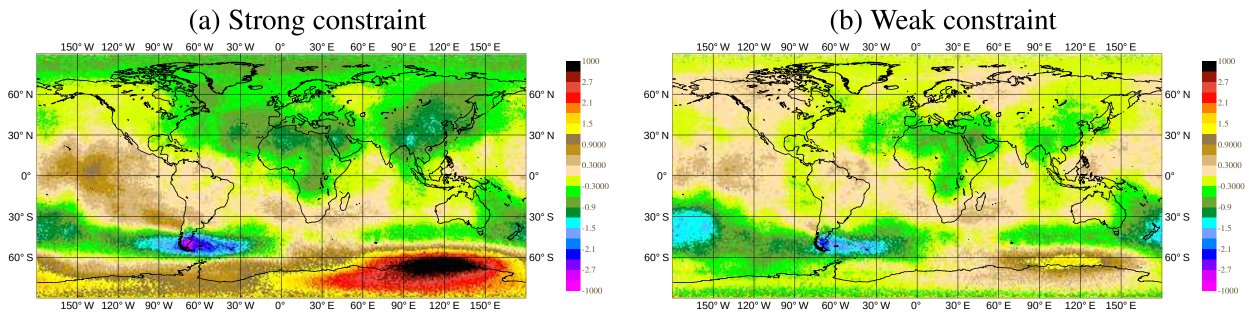

Figure 8: NOAA-16 AMSU-A channel 14 mean first guess departures (K) for July 2004, in strong and weak constraint 4D-Var.

models could be developed in the future but, as for the use of $\mathbf{B}$ in the past, this particular choice is readily available.

To determine the model error covariance matrix, forecasts were run from the 4D-Var ensemble at the spectral resolution of T319 and tendencies saved after 12h, 18h, 24h and 30h. Four tendencies spread over 24 hours were saved in order to avoid any diurnal cycle signal in the statistics and to increase the sample size. The first hours of the forecast were also avoided to reduce the dependence on the initial state and spin-up issues. The 4D-Var ensemble which was used had 10 members and was available for 26 days. This gave 936 realisations of model errors to be used as inputs to the statistical model.

Figures 3, 4 and 5 show some examples of the characteristics of the model error covariance matrix obtained by this method, next to the corresponding characteristics for background error. Comparison of horizontal correlations for temperature (figure 3) shows that model error correlations are narrower than background error ones.

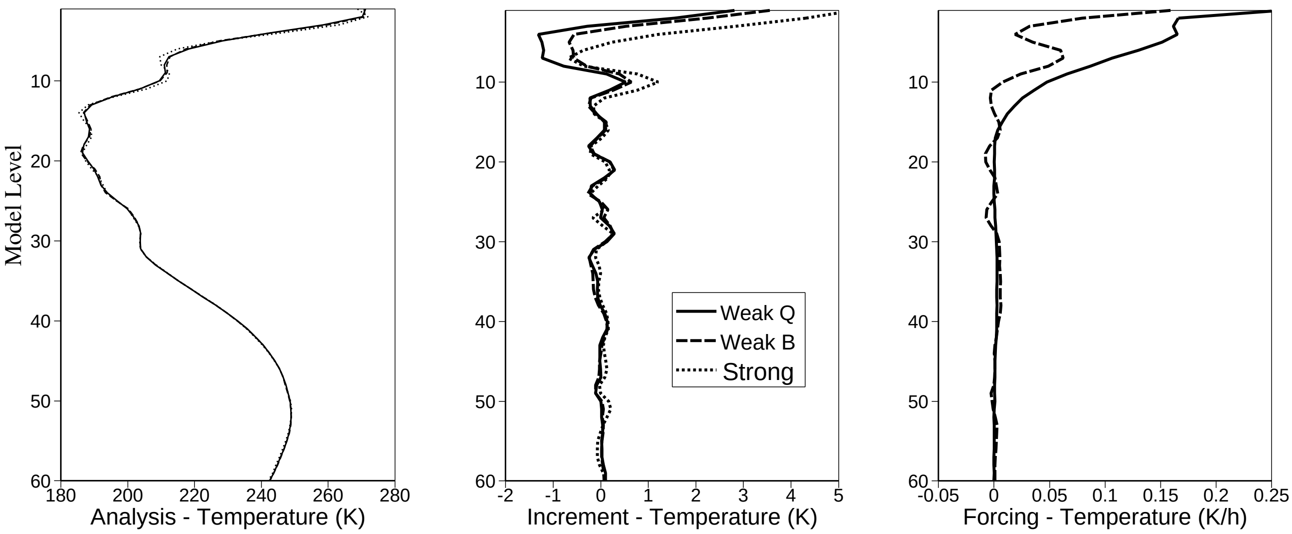

Figure 9: Profiles of the average temperature analysis, initial condition increments and model error forcing over Antarctica on 05/07/2004 at 12UTC with strong and weak constraint 4D-Var, with the tendency-based model error covariance matrix (weak Q) and with $Q = \alpha B$ (weak B).

As for background error, model error tends to be correlated over longer distances at higher levels. Figure 4 shows the average vertical correlations for background and model errors for temperature. Model error correlations are generally narrower than background errors and the negative lobes around the diagonal are reduced. Figure 5 shows the standard deviations for divergence for each model level and wavenumber. The standard deviations for background and model errors show similar patterns, although shifted towards higher wavenumbers (smaller scales) for model error. Generally, standard deviations peak at lower wavenumbers at higher levels, this is consistent with the wider correlations near the top of the model seen on figure 3. The statistics for model error are slightly less noisy because four realisations of model error are obtained for each of the assimilation cycles in the ensemble against one for background error.

Overall, this covariance matrix represents small scale structures in both the horizontal and vertical directions which is consistent with the fact that it is supposed to represent the characteristics of instantaneous model error.

4 Experimental results

Weak constraint 4D-Var with a model error forcing term has been implemented in the IFS. All the experiments described below were performed with the operational system at the time of this study, with a 12-hour assimilation window and where the resolution of the forecast and 4D-Var outer loops is set to T255. Two minimisations are performed, the first one at T95 without physics, the second one at T159 with linear physics. The model error covariance matrix described in section 3.2 was used for all the experiments presented below and the model error control variable is chosen constant over each 12-hour assimilation window.

4.1 Single observation experiments

In order to illustrate the difference between strong and weak constraint 4D-Var, two experiments were performed in simplified settings, with a single observation and with one observation at every hour in the assimi-

Assimilation Method	1 obs.	12 obs.
Strong Constraint	0.48	0.15
Weak Constraint	0.35	0.05

Table 2: Final value of J_o/p for simplified observations experiments in strong and weak constraint 4D-Var configurations.

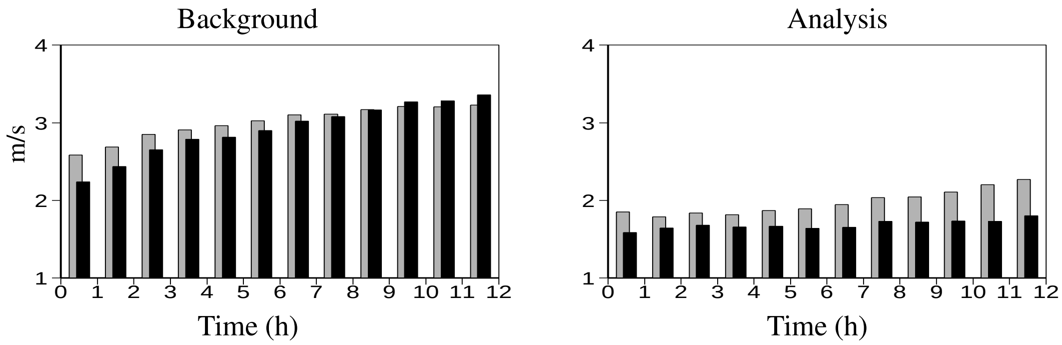

Figure 10: Profiler wind-speed departure standard deviation for North America for each hour bin in the assimilation window. Strong constraint 4D-Var is shown in grey, weak constraint 4D-Var in black.

lation window. The observations are of temperature, with a departure of 1K from the background, all located in the same position, at 40N, 40W and at the height of 500hPa. Figure 6 shows the increments obtained with strong and weak constraint 4D-Var when one observation every hour is assimilated. In the case of weak constraint 4D-Var, they include the cumulated effect of the change in initial condition and of model error forcing up to the time considered. This effect is shown after 6 and 12 hours on the figure. These increments are not exactly structure functions since 12 observations have been used and they show a combination of the structures in $\mathbf{B}$ and $\mathbf{Q}$.

The increments in strong constraint 4D-Var evolve according to the atmospheric flow. The fit to the observations is generally improved but mostly in the middle of the assimilation window as seen on the figure. The initial condition increment (top panels) is smaller with weak constraint 4D-Var since weak constraint 4D-Var also relies on the model error part of the control variable to fit the observations. In the weak constraint case, the initial increment evolves along the atmospheric flow but the forcing is added at every step and modifies the pattern as it evolves. The fit to the observations is improved as the resulting evolved increment is not just passing by the observations but the forcing term acts as a constant heating term at the observations location.

Overall, the weak constraint 4D-Var analysis is closer to the data than the strong constraint 4D-Var analysis. This is shown by the final values of J_o/p given in table 2. In the single observation experiment, with the observation in the middle of the assimilation window, weak constraint 4D-Var still can fit the observation better but the improvement is more limited, there is less benefit in a constant source term than with 12 observations.

These first experiments illustrate the properties of weak constraint 4D-Var. They are consistent with the expected results and allow continuation of the experimentation in the more complex operational environment.

4.2 Operational environment

Another experiment was performed with the operational observations for the month of July 2004. The forcing determined in one assimilation cycle is used as a first guess for the next cycle but there is no penalty term

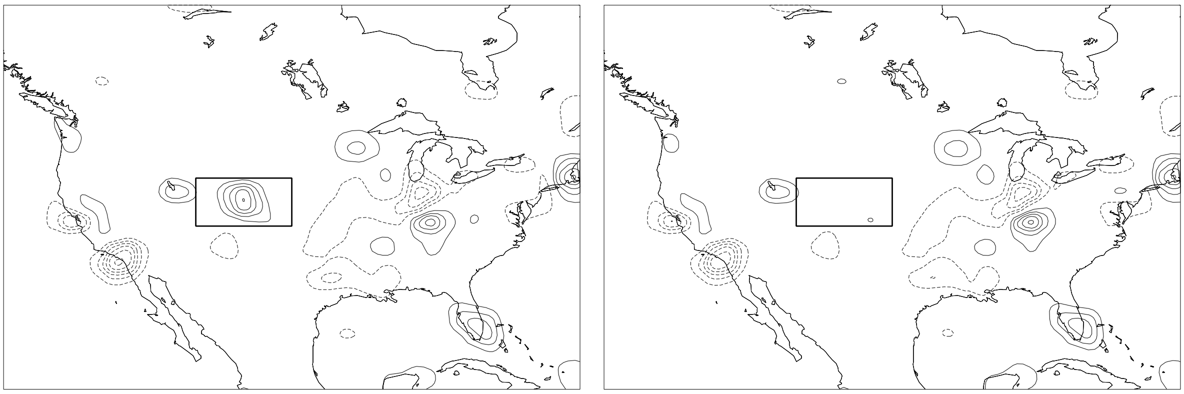

Figure 11: Average temperature forcing at the lowest model level over North America, with all data on the left, without aircraft data in the area shown on the picture on the right. Contour interval is 0.01 K/h.

Model level	Approx. Pressure (hPa)
1	0.1
11	5.2
20	35.8
30	202.1
39	500.0
50	884.3
60	1012.0

Table 3: Approximate pressure level corresponding to some selected model levels.

in the cost function to constrain it towards that first guess. (The model error term in the cost function only penalizes large model error as measured by the $\mathbf{Q}^{-1}$ metric.) This has the benefit of including the impact of the forcing in the background observation departure statistics. As expected from the description of the implementation, the cost per iteration of running this formulation of weak constraint 4D-Var is approximately the same as in strong constraint 4D-Var. The condition number of the minimisation problem and the number of iterations in the second minimisation have increased slightly. On average four more iterations were necessary, which corresponds to an average increase of the total run time of 4D-Var of 6%. This would make this system affordable for operational use in the future.

Since model error fields are not full fields, it would require defining a reference state to convert it to, for example, pressure coordinate. This applies to plotting model error but also to the use of model error to force the model. In addition to raising questions about the choice of the reference state, the change of vertical coordinate would have a cost that would make it impractical for operational use. In order to be consistent with the use being made of model error, all the figures presented in this section use model levels as the vertical coordinate. For reference, the approximate height of a few model levels are given in table 3.

As an example, the average temperature model error forcing and initial condition increments at model level 11 ($\approx$ 5hPa) for the month of July 2004 are shown on figure 7. Taking model error into account reduces the average initial condition increment, most dramatically over Antarctica. This is confirmed by the mean observation departures, an example is given on figure 8. Observation statistics also show that, in the southern hemisphere, the background departures standard deviation is reduced and more data is used when model error is taken into account.

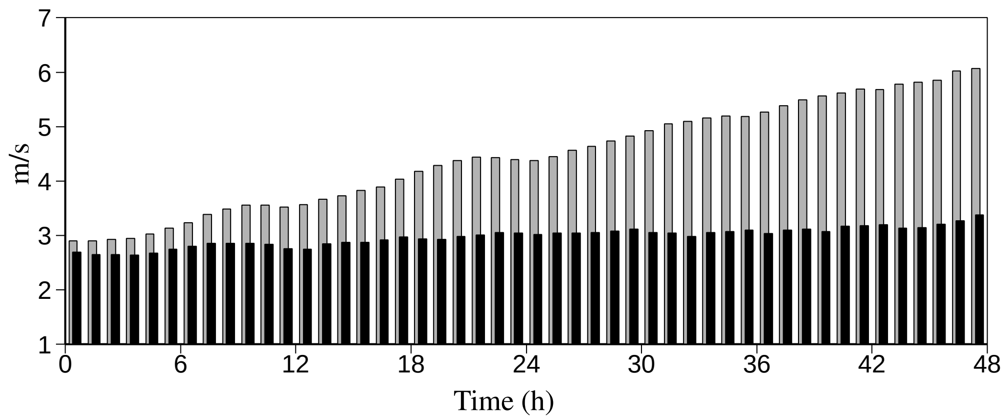

Figure 12: Profiler wind-speed departure standard deviation for North America for each hour bin in the 48-hour assimilation window, at the end of the low resolution minimisation in black and in the ensuing high resolution nonlinear trajectory in grey. This experiment was performed by M. Fisher (personal communication) as part of a long window 4D-Var study.

The large observation departures over Antarctica are associated with oscillations in the temperature profiles which are present in the analysis in the polar winter. The average temperature profiles of the analysis, initial condition increment and model error forcing for Antarctica are shown on figure 9. The oscillations in the analysis are reduced when model error is accounted for and, when the model error covariance matrix defined in section 3.2 is used, the vertical structure of the initial condition increments and of the forcing are different. In strong constraint 4D-Var, the oscillations in the initial condition increments propagate down to the surface. In weak constraint 4D-Var, they are reduced in amplitude and limited to the stratosphere when the model error covariance matrix is not proportional to $\mathbf{B}$. This is confirmed by observation statistics: the fit to radiosonde data over Antarctica show that the vertical oscillations in the bias are reduced and that the standard deviation is reduced above 50 hPa for both the background and analysis. Weak constraint 4D-Var gives the analysis more freedom to fit the data without causing spurious oscillations. This result shows the importance of having different structure functions for the initial condition and model error control variables. Overall, the systematic model error in the winter stratosphere has been captured and corrected by the model error forcing term.

As already mentioned, it is expected from Dee (1995b) that the fit of the analysis to the observations is more uniform over the assimilation window when model error is accounted for. This is confirmed by the fit to frequently reporting observation systems such as wind profilers over North America shown on figure 10 (right panel). The fit of the background to the observations is also improved but mostly in the first part of the assimilation window as shown on the left panel of the figure. This indicates that the forcing is beneficial only in the very short term and raises the question of the validity of the forcing outside the assimilation window. This experiment has also shown that applying the forcing for the entire length of the 10-day forecasts had a small negative impact on forecast performance. The model error determined in this formulation of weak constraint 4D-Var is only valid for a short period of time and would indicate that the decorrelation time for model error is relatively short. From that point of view, using a model error control variable constant for 12 hours might be too long, even if the fit of the analysis to observations is relatively flat. This should be taken into account when choosing a model for model error and when trying to account for time correlation in the model error covariance matrix in the future.

4.3 Model error and observation bias

The average temperature model error forcing at the lowest model level over North America is shown on figure 11. Several spots are visible on the map. Observation statistics show that the bias relative to aircraft low level temperature observations was reduced over North America and removing aircraft data in the box shown eliminates the spot in that box, confirming that aircraft data is driving that phenomenon. The pattern in the forcing was generated by the observations from aircraft near Denver airport. One characteristic of that type of data is that it is very frequent in time as aircraft traffic is very dense around major airports. An independent study (L. Isaksen, personal communication) has shown that aircraft temperature observations are biased in ascending and descending flight periods, i.e. mostly near airports. Although there were a few other observations in this area, they were ignored by the assimilation (as was the background) because of the density in time and space of the biased but consistent aircraft temperature measurements (on average 670 observations per assimilation cycle at 500 hPa and below in the box shown). In strong constraint 4D-Var, the initial condition increment generated by this observation departure pattern will be very diffuse since it is the evolved increment, over a period of time, that has to match the data. With the constant model error forcing term, these observations are fitted by adding a constant forcing at the airport location. This is similar to what was shown in the simplified experiment with 12 observations. However, in this particular case, the model error forcing term has captured observation bias rather than model error.

Weak constraint 4D-Var, with a constant forcing term, is very efficient in capturing a stationary discrepancy between the model and the observations. This can happen when persistent model error is present, but also in case of observation bias, particularly for observations with high frequency.

Weak constraint 4D-Var can only measure the discrepancy between the model and the observations. It is only through the specification of the cost function that this discrepancy can be attributed to the model or to the observations. More research will be necessary to determine the most efficient definition of the cost function to discriminate model error from other sources of error. Observation bias correction might also have an impact in this case and its interaction with weak constraint 4D-Var should be studied and taken into account.

4.4 Longer assimilation window

In strong constraint 4D-Var, the length of the assimilation window is limited by the range up to which model error can be neglected. One benefit of weak constraint 4D-Var is that it should allow the use of longer assimilation windows and better represent the dynamical evolution of the atmosphere. As shown by Fisher *et al.* (2005) another benefit is that when the assimilation window becomes long enough, weak constraint 4D-Var becomes equivalent to a full rank Kalman smoother.

In the incremental approach used in this model error forcing formulation of weak constraint 4D-Var, the tangent linear model is used in the minimisation. After the minimisation, the increment is added to the first guess and a high resolution nonlinear integration is performed. Since the tangent linear approximation is used for the full length of the assimilation window, the nonlinear trajectory might diverge from the linear integration. Experiments have been performed with assimilation windows of up to 48 hours with a model error forcing term constant by 12 hours intervals. Figure 12 shows that the fit to observations is uniform over the full length of the assimilation window at the end of the minimisation, as expected when model error is accounted for. However, this does not hold in the ensuing high resolution trajectory. Although accounting for model error should allow for longer assimilation windows, in this formulation, the length of the window is still constrained by the validity of the tangent linear approximation.

5 Conclusion

Weak constraint 4D-Var with a model error control variable has been implemented and tested within the IFS 4D-Var. The results presented in this study show that weak constraint 4D-Var is affordable in an operational environment for a global forecasting system.

We have shown that the choice of an appropriate model error covariance matrix is important. Using a model error covariance matrice independent from $\mathbf{B}$ gives 4D-Var more freedom to explore various directions to fit the observations. This is better than constraining the algorithm in directions already explored, which would introduce under-determination along those directions, even if the added directions are not optimal. For these reasons, a model error covariance matrix based on statistical properties of tendencies of the forecast model has been proposed and used in this implementation of weak constraint 4D-Var.

As expected from earlier works, weak constraint 4D-Var does fit observations more uniformly over the assimilation window. It captures some known model errors such as the temperature bias in the winter stratosphere and it also prevents some temperature oscillations which are degrading the forecast from occurring.

However, it also captures some observation bias as the example of the temperature bias in ascending and descending aircraft observations has shown. Since 4D-Var only sees the difference between the model and the observations, the distinction between model bias and observation bias can only be achieved through the prior knowledge built into the formulation of 4D-Var. Two tools are available to define the errors to be captured: the model error covariance matrix and the model for model error. The model error covariance matrix, here favouring small scales, can be tuned to selectively capture certain space and time scales representative of certain types of error. The model for model error, here a constant 3D field, can be more or less relevant to various forms of error and other models should be tested in the future, for example a periodic forcing or a formulation with shorter decorrelation time. Interactions with observation bias correction will also have to be studied and taken into account. More research is still needed in these areas to exploit the full potential of weak constraint 4D-Var.

Weak constraint 4D-Var is a generalisation of the more widely developed strong constraint 4D-Var where one simplifying assumption, namely that the forecast model is perfect, has been removed. In addition to lifting a questionable assumption, an estimate of model error is valuable information which can be used in several ways. In addition to providing a better analysis, it should help identify model deficiencies and improve the model. This process should also contribute to increase interactions between model development groups and data assimilation groups and result in better understanding of forecasting systems.

However, the formulation presented in this study, using a model error forcing term, is not sufficient to allow for the use of long windows in 4D-Var because it still relies on the tangent linear approximation over the full length of the assimilation window. Using a four dimensional model state control variable seems necessary if the length of the assimilation window is to be increased, at least in the incremental approach. This paper is part of an on-going research effort in the area of weak constraint 4D-Var. Other formulations proposed by Trémolet (2006) are being developed at ECMWF.

The most appropriate formulation of weak constraint 4D-Var in terms of the control variable will have to be determined. A choice will have to be made depending on the type of model error (random or bias) to be estimated and may depend on the context in which it is used. The best formulation might differ, for example in an operational context and in a reanalysis context where future observations are available to the data assimilation scheme.

References

E. Andersson, C. Cardinali, M. Fisher, E. Hólm, L. Isaksen, Y. Trémolet, and A. Hollingsworth. Developments in ECMWF's 4D-Var System. In *AMS Symposium on Forecasting the Weather and Climate of the Atmosphere and Ocean*, January 2004. Available from the American Meteorological Society, http://ams.confex.com/ams/84Annual/20WAF16NW/program.htm.

A. Bennett, B. Chua, and L. Leslie. Generalized Inversion of a Global Numerical Weather Prediction Model. *Meteorol. Atmos. Phys.*, 60:165–178, 1996.

P. Courtier. Dual formulation of four-dimensional variational assimilation. *Q. J. R. Meteorol. Soc.*, 123:2449–2461, 1997.

P. Courtier, E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, F. Rabier, A. Hollingsworth, and M. Fisher. The ECMWF implementation of three dimensional variational assimilation (3D-Var). Part I: Formulation. *Q. J. R. Meteorol. Soc.*, 124:1783–1807, 1998.

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

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

D. Dee. Testing the perfect model assumption in variational data assimilation. In *WMO Symposium on Assimilation of Observations in Meteorology and Oceanography*, Tokyo, March 1995b.

D. Dee. Bias and data assimilation. *Q. J. R. Meteorol. Soc.*, 131:3323–3343, 2005.

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

J. Derber. A variational continuous assimilation technique. *Mon. Wea. Rev.*, 117:2437–2446, 1989.

M. Fisher. Background error covariance modelling. In *Workshop on Recent developments in data assimilation for atmosphere and ocean*, ECMWF, pages 45–63, September 2003.

M. Fisher, M. Leutbecher, and G. Kelly. On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.*, 131:3235–3246, 2005.

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 335–343, 1998.

A.H. Jazwinski. *Stochastic processes and fi ltering theory*. Academic Press, 1970.

E. Klinker, F. Rabier, G. Kelly, and J.-F. Mahfouf. The ECMWF operational implementation of four dimensional variational assimilation. Part III: Experimental results and diagnostics with operational configuration. *Q. J. R. Meteorol. Soc.*, 126:1191–1215, 2000.

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

J.-F. Mahfouf and F. Rabier. The ECMWF operational implementation of four dimensional variational assimilation. Part II: Experimental results with improved physics. *Q. J. R. Meteorol. Soc.*, 126:1171–1190, 2000.

A. McNally. The assimilation of stratospheric satellite data at ECMWF. In *ECMWF/ SPARC Workshop on modelling and assimilation for the stratosphere and tropopause*, pages 103–106, June 2003.

F. Rabier, H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons. The ECMWF operational implementation of four dimensional variational assimilation. Part I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.*, 126:1143–1170, 2000.

C. Rodgers. *Inverse Methods for Atmospheric Sounding*. World Scientific, 2000.

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

O. Talagrand. A posteriori evaluation and verification of analysis and assimilation algorithms. In *Workshop on Diagnosis of Data Assimilation Systems*, ECMWF, November 1998.

O. Talagrand and F. Bouttier. Internal diagnostics of data assimilation systems. In *Workshop on Diagnosis of Models and Data Assimilation Systems*, ECMWF, September 1999.

Y. Trémolet. Accounting for an imperfect model in 4D-Var. *Q. J. R. Meteorol. Soc.*, 132:2483–2504, 2006.

P. A. Vidard, A. Piacentini, and F.-X. Le Dimet. Variational data analysis with control of the forecast bias. *Tellus*, 56A:177–188, 2004.

D. Zupanski. A general weak constraint applicable to operational 4DVAR data assimilation systems. *Mon. Wea. Rev.*, 125:2274–2292, 1997.

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