TECHNICAL MEMORANDUM

655

Weak-Constraint and Long-Window 4D-Var

Mike Fisher, Yannick Trémolet, Harri Auvinen[1], David Tan and Paul Poli

Research Department

[1] Lappeenranta University of Technology, Lappeenranta, Finland.

Presented to the SAC 40th Session 3–5 October 2011

December 2011

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

European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, RG2 9AX, England

Abstract

We describe recent progress in weak-constraint and long-window 4D-Var.

Experiments in a two-layer quasi-geostrophic system confirm the results obtained by Fisher et al. (2005) for the Lorenz (1995) model. The experiments show that increasing the length of the analysis window in 4D-Var improves the analysis. Moreover, this benefit is present even if the model error is induced by systematic error in the model parameters, and is imperfectly described by the analysis system.

We present results using the full ECMWF analysis system showing that the analysis window can usefully be extended to 24 hours. It is expected that 24 hour 4D-Var will be implemented in 2012.

We also present results using a degraded observation network comprising surface pressure observations only. We find that a 24-hour analysis window gives better results than a 12-hour analysis window, and the improvement is even more significant when the 24-hour windows are "overlapped" by performing an analysis every 12 hours.

The importance of properly taking into account the mean component of model error is highlighted in both the two-layer quasi-geostrophic context, and in experiments conducted using an Reanalysis configuration of the ECMWF assimilation system in which only historical surface pressure observations are assimilated.

A major potential benefit of weak-constraint 4D-Var is that it allows parallelisation of the inner loops of 4D-Var. We discuss how this benefit might be realised, and present a new parallel algorithm which we call the "saddle point" formulation of 4D-Var. We demonstrate the algorithm using the two-layer quasi-geostrophic system.

1 Introduction

The evolution of the ECMWF data assimilation system over the next few years comprises a number of strands, including increased use of ensemble methods (to provide estimates of analysis and background error and to provide initial perturbations for ensemble prediction), extension of the analysis window, and explicit recognition of model error. Here we concentrate on one aspect of our plans: long-window and weak-constraint 4D-Var. We present results from recent research and experimentation. Since this is an ongoing area of research, the paper should be regarded as a snapshot of the current status of the work.

In section 2 we provide a derivation of the weak constraint 4D-Var equations. This derivation makes a clear distinction between the random and systematic components of model error by explicitly expressing model error as a Gaussian variable with prescribed covariance and (non-zero) mean. We also introduce an operator $\mathcal{L}$ that converts between the alternative "4D-state" and "forcing" formulations. This operator, together with its tangent-linear equivalent $\mathbf{L}$, plays a critical role in determining whether variants of the 4D-Var algorithm are parallelisable.

An incremental formulation of weak constraint 4D-Var is presented in section 3.

Section 4 extends the results obtained by Fisher *et al.* (2005). Results are presented for a much more realistic system than was used by Fisher *et al.*, who conducted "perfect model" assimilation experiments using the Lorenz (1995) model. Our results show that long-window 4D-Var remains beneficial even when model error is imperfectly described by the assimilation system.

Experimentation with the full ECMWF analysis system is described in sections 5 and 6. Weak constraint 4D-Var using the ECMWF system was described by Trémolet (2006) and Trémolet (2007). The results presented here build on this earlier work by introducing a new method for calculating the covariance matrix of model error, by extending and overlapping the analysis window, and by better taking into account the systematic component of model error.

Extending the analysis window to 24 hours appears to give a rather neutral impact on forecast scores. However, it should be stressed that these experiments do not represent long-window 4D-Var in its envisaged form. Model error is assumed to be constant throughout the analysis window, and is restricted to the stratosphere. Nevertheless, we present encouraging evidence that the weak-constraint formulation of 4D-Var is able to diagnose and compensate for an important systematic model error.

Parallelisation of 4D-Var is becoming a critical concern. 4D-Var will not remain a viable data assimilation method on future supercomputers unless significant additional parallelism can be extracted. It is unlikely that this parallelism can be gained from further optimisation of the model. So, algorithmic changes are required. In section 8 we analyse this problem. We characterise variants of the 4D-Var algorithm according to their parallel or sequential natures. We present a new parallel algorithm that we believe will allow 4D-Var to remain a competitive algorithm on next-generation computers.

2 Formulation of Weak-Constraint 4D-Var

Weak constraint 4D-Var estimates a four-dimensional state $\mathbf{x}$, defined as a collection of three-dimensional states x_k ($k = 0, \ldots, N-1$), each of which is valid at the start of a time interval $[t_k, t_{k+1})$. We refer to these intervals as "sub-windows". Together, the sub-windows span the "analysis window" $[t_0, t_N)$.

We write the four-dimensional state as a vector:

$$\mathbf{x} = \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{pmatrix}. \tag{1}$$

(Note: In this paper, we use bold type for four-dimensional vectors and for the matrices that act on them.)

The analysis is defined as the four-dimensional state that maximises the conditional probability, given a prior estimate x_b of the state at the start of the analysis window and observations y_k for each of the sub-windows. That is, the analysis is defined as the state $\mathbf{x}$ that maximises $p(\mathbf{x}|x_b, y_0 \ldots y_{N-1})$.

Using Bayes' theorem, we have

$$p(\mathbf{x}|x_b, y_0 \ldots y_{N-1}) = \frac{p(x_b, y_0 \ldots y_{N-1}|\mathbf{x})p(\mathbf{x})}{p(x_b, y_0 \ldots y_{N-1})}. \tag{2}$$

To simplify this equation, we will assume that the errors in the prior and in observations taken at different times are independent. This allows us to write the conditional probability $p(x_b, y_0 \ldots y_{N-1}|\mathbf{x})$ as a product of independent probabilities. Moreover, we note that the conditional probability of x_b depends only on the three dimensional state at time t_0, and not on the entire four-dimensional state. Likewise, the conditional probability of y_k depends only on x_k. Noting also that the denominator of equation 2 is independent of $\mathbf{x}$, we have

$$p(\mathbf{x}|x_b, y_0 \ldots y_{N-1}) \propto p(x_b|x_0) \left[\prod_{k=0}^{N-1} p(y_k|x_k) \right] p(\mathbf{x}). \tag{3}$$

Taking minus the logarithm gives the 4D-Var cost function:

$$J(\mathbf{x}) = J_b + J_o + J_q + c \tag{4}$$

where c is a constant, and where

$$J_b = -\log\left(p(x_b|x_0)\right) \tag{5}$$

$$J_o = -\sum_{k=0}^{N-1} \log\left(p(y_k|x_k)\right) \tag{6}$$

$$J_q = -\log\left(p(\mathbf{x})\right) \tag{7}$$

The first term of equation 4 represents the cost associated with errors in the prior, x_b. Typically, these errors are assumed to be Gaussian and to be unbiased, in which case the first term of the cost function becomes

$$J_b = \frac{1}{2}\left(x_0 - x_b\right)^{\mathrm{T}} B^{-1}\left(x_0 - x_b\right) + const. \tag{8}$$

where B is the covariance matrix of the error in x_b.

The second term of equation 4 represents the cost associated with errors in the observations. For present purposes, we will assume that these errors are also Gaussian and unbiased, although we note that both assumptions may be relaxed. With these assumptions, we have

$$J_o = \frac{1}{2}\sum_{k=0}^{N-1}\left(\mathcal{H}_k(x_k)-y_k\right)^{\mathrm{T}} R_k^{-1}\left(\mathcal{H}_k(x_k)-y_k\right)+const. \tag{9}$$

Here, $\mathcal{H}_k$ is an operator that maps analysed variables to observed variables. This mapping may include spatial interpolation, as well as conversions from analysed quantities (e.g. temperature, specific humidity, vorticity and divergence) to observed quantities (e.g. radiance or wind components). Note also that if observations are spread throughout the interval $[t_k, t_{k+1})$, then $\mathcal{H}_k$ may also involve an operator to determine the state at intermediate times between t_k and t_{k+1}. The most accurate method available for determining such intermediate states is a numerical forecast model. Thus, $\mathcal{H}_k$ will typically involve integration of the forecast model.

The final term of the cost function, J_q, requires us to assign a probability to a four-dimensional state, $\mathbf{x}$, independently of x_b and of the observations. To assign this probability, we first apply Bayes' theorem recursively to write

$$p(\mathbf{x}) = p(x_{N-1}\cap x_{N-2}\cap\ldots\cap x_0) = \left[\prod_{k=1}^{N-1} p(x_k|x_{k-1},\ldots,x_0)\right] p(x_0) \tag{10}$$

Taking minus the logarithm of this equation, gives us an expression for the J_q component of the cost function:

$$J_q = -\left[\sum_{k=1}^{N-1}\log p(x_k|x_{k-1},\ldots,x_0)\right] - \log p(x_0). \tag{11}$$

The final term in this equation must be evaluated before the prior state or observations are known. In the absence of other information, we assume that all states are equally likely, so that the term may be regarded as an additive constant. We will also assume that the sequence of states forms a Markov chain. That is,

$$p(x_k|x_{k-1},\ldots,x_0) = p(x_k|x_{k-1}) \quad \text{for } k=1,\ldots,N-1 \tag{12}$$

We now use our knowledge of the dynamics and physics of the system, as embodied in the forecast model, to assign the conditional probabilities $p(x_k|x_{k-1})$. Let us denote by $\mathcal{M}_k$ the operator that integrates the forecast model from time t_{k-1} to time t_k. Given the state x_{k-1}, we use this operator to predict the state at time t_k and compare this forecast with the state x_k. We will denote the difference between x_k and its prediction by q_k, and refer to it as "model error":

$$q_k = x_k - \mathcal{M}_k(x_{k-1}). \tag{13}$$

With this definition, we have $p(x_k|x_{k-1}) = p(q_k)$, so that if model error is Gaussian with covariance Q_k and mean $\bar{q}$, we have

$$J_q = -\left[\sum_{k=1}^{N-1}\log p(q_k)\right] - \log p(x_0) \tag{14}$$

$$= \frac{1}{2}\sum_{k=1}^{N-1}\left(q_k-\bar{q}\right)^{\mathrm{T}} Q_k^{-1}\left(q_k-\bar{q}\right)+const. \tag{15}$$

Gathering the three terms of the cost function together, we have (subject to the assumptions outlined above, and with a judicial choice of the additive constant, c):

$$\begin{aligned} J(\mathbf{x}) \quad = \quad & \frac{1}{2}\left(x_0 - x_b\right)^{\mathrm{T}} B^{-1}\left(x_0 - x_b\right) \\ & + \frac{1}{2}\sum_{k=0}^{N-1}\left(\mathcal{H}_k(x_k) - y_k\right)^{\mathrm{T}} R_k^{-1}\left(\mathcal{H}_k(x_k) - y_k\right) \\ & + \frac{1}{2}\sum_{k=1}^{N-1}\left(q_k - \bar{q}\right)^{\mathrm{T}} Q_k^{-1}\left(q_k - \bar{q}\right). \end{aligned} \tag{16}$$

Equation 16 expresses the cost function as a function of the four-dimensional state, $\mathbf{x}$. We will refer to this as the "4D-state formulation". We could equally have expressed the cost function as a function of the vector

$$\mathbf{p} = \begin{pmatrix} x_0 \\ q_1 \\ \vdots \\ q_{N-1} \end{pmatrix}. \tag{17}$$

We will refer to the cost function expressed as a function of $\mathbf{p}$ as the "forcing formulation" of 4D-Var.

The vectors $\mathbf{x}$ and $\mathbf{p}$ are related via equation 13. We will denote this functional relationship as $\mathbf{p} = \mathcal{L}(\mathbf{x})$. We note that $\mathcal{L}$ is invertible, since we can determine $\mathbf{x}$ from $\mathbf{p}$ using

$$x_k = \mathcal{M}_k(x_{k-1}) + q_k. \tag{18}$$

An important observation, which we will return to in section 8, is that the model integrations required to calculate $\mathbf{p}$ given $\mathbf{x}$ can be performed in parallel, since $\mathbf{x}$ contains the initial states for all the integrations. By contrast, the inverse operation (calculating $\mathbf{x}$ given $\mathbf{p}$) requires a sequence of model integrations to determine first x_1, then x_2, then x_3, etc.

3 The Incremental Algorithm for Weak Constraint 4D-Var

The incremental algorithm was introduced for strong-constraint 4D-Var by Courtier *et al.* (1994). Its extension to weak-constraint 4D-Var is fairly straightforward. However, the need to update states at intermediate times during the assimilation window raises some additional considerations that are not of concern in strong-constraint 4D-Var.

The essence of the incremental algorithm is to approximate the cost function (equation 16) by a quadratic function of an increment, $\delta\mathbf{x}^{(n)}$, to a reference (trajectory) state $\mathbf{x}^{(n)}$:

$$\mathbf{x}^{(n)} = \begin{pmatrix} x_0^{(n)} \\ x_1^{(n)} \\ \vdots \\ x_{N-1}^{(n)} \end{pmatrix}. \tag{19}$$

(Note: the superscript (n) represents an integer label that we attach to the states, increments and operators. We will identify this label later in this section as the loop index of an iterative solution algorithm.)

The nonlinear operators in the cost function are approximated to first order:

$$\mathcal{H}_k(x_k) \approx \mathcal{H}_k(x_k^{(n)}) + H_k^{(n)} \delta x_k^{(n)} \tag{20}$$

$$\mathcal{M}_k(x_{k-1}) \approx \mathcal{M}_k(x_{k-1}^{(n)}) + M_k^{(n)} \delta x_{k-1}^{(n)} \tag{21}$$

where $H_k^{(n)}$ and $M_k^{(n)}$ are the tangent linear operators associated with $\mathcal{H}_k$ and $\mathcal{M}_k$, and with the trajectory $\mathbf{x}^{(n)}$.

It will be convenient to introduce the following quantities:

$$b^{(n)} = x_b - x_0^{(n)} \tag{22}$$

$$c_k^{(n)} = \bar{q} - q_k^{(n)} \tag{23}$$

$$d_k^{(n)} = y_k - \mathcal{H}_k(x_k^{(n)}) \tag{24}$$

where $q_k^{(n)} = x_k^{(n)} - \mathcal{M}_k(x_{k-1}^{(n)})$.

With these definitions, we may write the quadratic approximation to the cost function as

$$\begin{aligned} J(\delta\mathbf{x}^{(n)}) = & \frac{1}{2}\left(\delta x_0^{(n)} - b^{(n)}\right)^{\mathrm{T}} B^{-1}\left(\delta x_0^{(n)} - b^{(n)}\right) \\ & + \frac{1}{2}\sum_{k=0}^{N-1}\left(H_k^{(n)}\delta x_k^{(n)} - d_k^{(n)}\right)^{\mathrm{T}} R_k^{-1}\left(H_k^{(n)}\delta x_k^{(n)} - d_k^{(n)}\right) \\ & + \frac{1}{2}\sum_{k=1}^{N-1}\left(\delta q_k^{(n)} - c_k^{(n)}\right)^{\mathrm{T}} Q_k^{-1}\left(\delta q_k^{(n)} - c_k^{(n)}\right). \end{aligned} \tag{25}$$

In section 2, we noted that the cost function could be expressed either as a function of the four-dimensional state, $\mathbf{x}$, or as a function of the vector $\mathbf{p}$ containing the initial state and model errors. Likewise, we can express the incremental cost function, equation 25, as a function of $\delta\mathbf{x}^{(n)}$, or of $\delta\mathbf{p}^{(n)}$, defined as:

$$\delta\mathbf{p}^{(n)} = \begin{pmatrix} \delta x_0^{(n)} \\ \delta q_1^{(n)} \\ \vdots \\ \delta q_{N-1}^{(n)} \end{pmatrix}. \tag{26}$$

The vectors $\delta\mathbf{x}^{(n)}$ and $\delta\mathbf{p}^{(n)}$ are related by an invertible linear function, $\delta\mathbf{p}^{(n)} = \mathbf{L}^{(n)}\delta\mathbf{x}^{(n)}$, where $\mathbf{L}^{(n)}$ can be expressed in matrix form as:

$$\mathbf{L}^{(n)} = \begin{pmatrix} I & & & & \\ -M_1^{(n)} & I & & & \\ & -M_2^{(n)} & I & & \\ & & \ddots & \ddots & \\ & & & -M_{N-1}^{(n)} & I \end{pmatrix}. \tag{27}$$

Note that, as was the case for the operator $\mathcal{L}$, the model integrations required to determine $\delta\mathbf{p}^{(n)}$ from $\delta\mathbf{x}^{(n)}$ can be performed in parallel, whereas calculation of $\delta\mathbf{x}^{(n)}$ given $\delta\mathbf{p}^{(n)}$ must be done by forward substitution, requiring sequential application first of $M_1^{(n)}$, then $M_2^{(n)}$, $M_3^{(n)}$, etc. We return to this point in section 8.

The incremental algorithm is an iterative solution procedure for minimising the cost function (equation 16). Starting with an initial approximate solution $\mathbf{x}^{(0)}$, the algorithm minimises the quadratic approximation (equation 25) to determine an increment which is used to update the solution — i.e. to determine $\mathbf{x}^{(1)}$. The process is repeated to determine $\mathbf{x}^{(2)}$, $\mathbf{x}^{(3)}$, etc. Typically, very few iterations are performed and there is no check for convergence (which is not guaranteed).

There are two main reasons for adopting the incremental algorithm. The first is that, as described by Courtier *et al.* (1994), the quadratic cost function (equation 25) may be further approximated by evaluating all quantities, except $d_k^{(n)}$ and R_k, at a reduced spatial resolution. This reduces considerably the computational cost of 4D-Var, at the expense of introducing approximations that have a detrimental impact on the quality of the analysis.

The second reason for adopting the incremental algorithm is that minimisation of the quadratic cost function is equivalent to solution of the linear equation for its gradient. This allows a range of efficient methods for solving linear systems of equations to be brought to bear on the problem.

The existence of operators $\mathcal{L}$ and $\mathbf{L}^{(n)}$ which convert between $\mathbf{p}$ and $\mathbf{x}$, and between $\delta\mathbf{p}^{(n)}$ and $\delta\mathbf{x}^{(n)}$, provides options for formulating the incremental algorithm that are not available in strong-constraint 4D-Var. The inner (quadratic) cost function may be minimised either as a function of $\delta\mathbf{x}^{(n)}$, or as a function of $\delta\mathbf{p}^{(n)}$. These two possibilities lead to very different optimisation problems that present different possibilities for preconditioning and parallelisation. This point is discussed in more detail in section 8.

Once the quadratic cost function has been minimised, we can (if required) convert $\delta\mathbf{p}^{(n)}$ to $\delta\mathbf{x}^{(n)}$, or *vice versa*, so that we have the choice to update either $\mathbf{p}^{(n)}$ or $\mathbf{x}^{(n)}$ in the outer iteration. (Whichever we choose to update, we can determine the other by applying $\mathcal{L}$ or its inverse.) In practice, however, we have found that updating $\mathbf{x}^{(n)}$ results in better convergence properties for the incremental update. The reason for this may be seen by examining equation 18, which represents the conversion of $\mathbf{p}^{(n)}$ to $\mathbf{x}^{(n)}$. The conversion amounts to a forced integration of the forecast model over the entire analysis window, with the forcing provided by $\mathbf{p}^{(n)}$. For long analysis windows, this forced integration is very sensitive to small changes in the forcing towards the start of the window. There is no guarantee that the nonlinear model will respond similarly to the linearised model used in the inner minimisation, particularly if the latter is run at a lower spatial resolution. Thus, over a long assimilation window, the state at the outer loop is likely to diverge radically from that implied by the inner minimisation, resulting in divergence of the incremental algorithm. By contrast, updating $\mathbf{x}^{(n)}$ directly using $\delta\mathbf{x}^{(n)}$ guarantees that the former remains close to the observations, even if the linear and nonlinear models differ significantly. It also has the advantage that a single integration of the nonlinear model over the entire analysis window is not required. The integration may be split up into a set of N shorter integrations which may be run in parallel.

4 Results from a Simplified System

In this section, we demonstrate the benefits of extending the analysis window in weak constraint 4D-Var, in the context of a two layer quasi-geostrophic model. The experiments confirm the results obtained for the simpler Lorenz 1995 system by Fisher *et al.* (2005). We stress that, whereas Fisher *et al.* (*op. cit.*) worked in the context of a perfect model, the results presented below are obtained for an imperfect model with realistic model error produced by perturbing the parameters of the model.

In the next two subsections, we describe the numerical model and analysis system used for this study. Experimental results are presented in subsection 4.3.

4.1 The OOPS Quasi-Geostrophic Model

The two-layer quasi-geostrophic model implemented in the OOPS data assimilation framework was designed for speed, simplicity and robustness, rather than for realism, accuracy or numerical sophistication. It has, nevertheless, proved a useful tool for the numerical studies reported in this paper and elsewhere. The model represents quasi-geostrophic flow in a cyclic channel.

The equations of the two-level model are given by Fandry and Leslie (1984) (see also Pedlosky, 1979 pp386-393), and are expressed in terms of non-dimensionalised variables:

$$\frac{\mathrm{D}q_1}{\mathrm{D}t} = \frac{\mathrm{D}q_2}{\mathrm{D}t} = 0 \tag{28}$$

where q_1 and q_2 denote the quasi-geostrophic potential vorticity on each of the two layers, with a subscript 1 denoting the upper layer:

$$q_1 = \nabla^2\psi_1 - F_1(\psi_1 - \psi_2) + \beta y \tag{29}$$
$$q_2 = \nabla^2\psi_2 - F_2(\psi_2 - \psi_1) + \beta y + R_s \tag{30}$$

Here, β is the (non-dimensionalised) northward derivative of the Coriolis parameter, and R_s represents orography.

The non-dimensionalisation is standard, but is given here for completeness. We define a typical length scale L, a typical velocity U, the depths of the upper and lower layers D_1 and D_2, the Coriolis parameter at the southern boundary f_0 and its northward derivative β_0, the acceleration due to gravity g, the difference in potential temperature across the layer interface $\Delta\theta$, and the mean potential temperature $\overline{\theta}$.

Denoting dimensional time, spatial coordinates and velocities with tildes, we have:

$$\begin{aligned} x &= \frac{\tilde{x}}{L}, \qquad y = \frac{\tilde{y}}{L}, \qquad u = \frac{\tilde{u}}{\overline{U}}, \qquad v = \frac{\tilde{v}}{\overline{U}}, \\ t &= \tilde{t}\frac{\overline{U}}{L}, \qquad \beta = \beta_0\frac{L^2}{\overline{U}}, \qquad F_1 = \frac{f_0^2L^2}{D_1 g\Delta\theta/\overline{\theta}}, \qquad F_2 = \frac{f_0^2L^2}{D_2 g\Delta\theta/\overline{\theta}}. \end{aligned} \tag{31}$$

The parameters used for this study were as follows:

$$\begin{aligned} &L = 10^6\,\mathrm{m}, \qquad \overline{U} = 10\,\mathrm{ms}^{-1}, \qquad f_0 = 10^{-4}\,\mathrm{s}^{-1}, \qquad \beta_0 = 1.5\times10^{-11}\,\mathrm{s}^{-1}\mathrm{m}^{-1}, \\ &g = 10\,\mathrm{ms}^{-2}, \qquad D_1 = 6000\,\mathrm{m}, \qquad D_2 = 4000\,\mathrm{m}, \qquad \frac{\Delta\theta}{\overline{\theta}} = 0.1 \end{aligned} \tag{32}$$

The Rossby number is $\varepsilon = \overline{U}/f_0L = 0.1$. The parameters listed above are those used to define the true evolution of the system. The data assimilation experiments reported in this paper were conducted using a deliberately imperfect model produced by modifying the upper and lower layer depths to 5500m and 4500m, respectively.

The model variables (streamfunction, potential vorticity and wind components) are defined on an unstaggered rectangular grid of dimension $n_x \times n_y$. All the experiments shown in this paper used $n_x = 40$ and $n_y = 20$, with a dimensional grid spacing of 300km in both the north-south and east-west directions. Centred finite differences are used for all horizontal derivatives.

Values of streamfunction one grid-space to the north and south of the grid are required in order to calculate the zonal component of velocity on the first and last grid row. These values are user-supplied

constants, and determine the mean zonal velocity in each layer, which remains constant throughout the integration. The meridional velocity is assumed to vanish one grid space to the north and south of the domain.

The time-stepping algorithm is somewhat novel. It consists of a semi-Lagrangian advection of potential vorticity, followed by an inversion of the potential vorticity equation to determine streamfunction and velocity components. The time-stepping is only first-order accurate, in the interests of speed and in order to permit the the state of the model to be characterised by the streamfunction at a single time level. This simplifies the implementation and testing of assimilation algorithms, which is the primary purpose of the model. The interpolation to the departure point is bi-cubic. A one-hour timestep was used for all the experiments presented in this paper.

Potential vorticity (equations 29 and 30) is discretised using a standard five-point finite-difference representation of the Laplacian. It is inverted by applying ∇^2 to equation 29 and subtracting F_1 times equation 29 and F_2 times equation 30 to give:

$$\nabla^2 q_1 - F_2 q_1 - F_1 q_2 = \nabla^2 \left(\nabla^2 \psi_1\right) - (F_1 + F_2) \nabla^2 \psi_1 \tag{33}$$

This is a two-dimensional Helmholtz equation, which can be solved for $\nabla^2 \psi_1$. The Laplacian can then be inverted to determine ψ_1. Once ψ_1 and $\nabla^2 \psi_1$ are known, the streamfunction on level 2 can be determined by substitution into equation 29.

Solution of the Helmholtz equation and inversion of the Laplacian are achieved using an FFT-based method. Applying a Fourier transform in the east-west direction to equation 33 gives a set of independent equations for each wavenumber. In the case of the five-point discrete Laplacian, these are tri-diagonal matrix equations, which can be solved using the standard (Thomas) algorithm.

Despite its simplicity, the model provides a surprisingly good analogue of large-scale mid-latitude dynamics. Figure 1 shows the mean growth of perturbation energy, averaged over 100 cases. Here, energy is defined as (see Pedlosky, 1979 p393):

$$E = \frac{1}{D_1 + D_2} \sum D_1 \left(\Delta u_1^2 + \Delta v_1^2\right) + D_2 \left(\Delta u_2^2 + \Delta v_2^2\right) + (F_1 D_1 + F_2 D_2) \left(\Delta \psi_1 - \Delta \psi_2\right)^2 \tag{34}$$

where Δu_i, Δv_i and $\Delta \psi_i$ represent differences of non-dimensional wind and streamfunction on level i, between two integrations of the model with different initial conditions. The summation is over all gridpoints.

The initial states for the two sets of integrations were constructed by taking a sequence of states from an unperturbed "truth run", and adding and subtracting perturbations drawn from a multivariate Gaussian distribution with zero mean and covariance matrix constructed from a large sample of errors in three-hour forecasts made by a version of the model with misspecified layer depths.

The perturbation energy shows an initial decrease followed, after about 4 hours, by a rapid increase. This behaviour is discussed by Snyder *et al.* (2003) (see also Wirth and Ghil, 2000). The initial decrease in energy results from dissipation of the smallest scales, while the exponential growth which follows occurs at a rate determined by the leading Lyapunov exponent.

The perturbation energy increases approximately three-fold between 24 and 48 hours, corresponding to a perturbation doubling time (i.e. an energy quadrupling time) of a little over 30 hours. This is a little shorter than typical error doubling times found in current NWP models (see Simmons and Hollingsworth, 2002).

The growth rate decreases as the integrations progress and the evolution becomes more nonlinear. Between 132 and 240 hours of integration the energy roughly doubles.

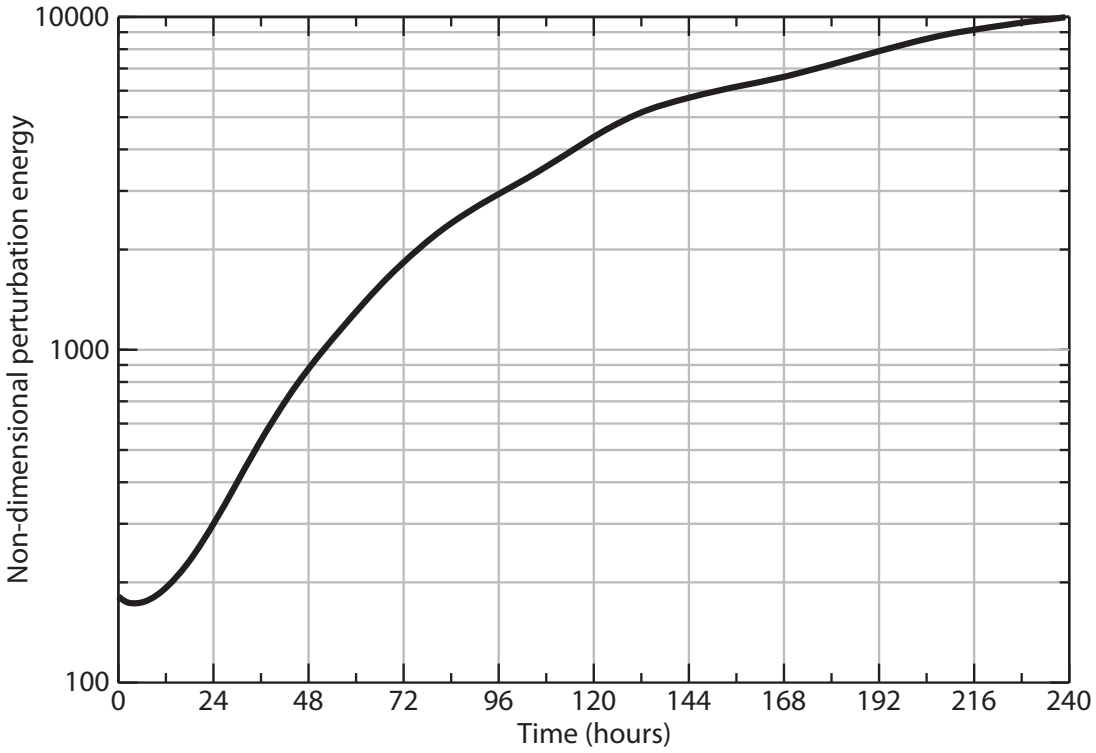

Figure 1: The mean evolution of perturbation energy, averaged over a large sample of initial states and perturbations.

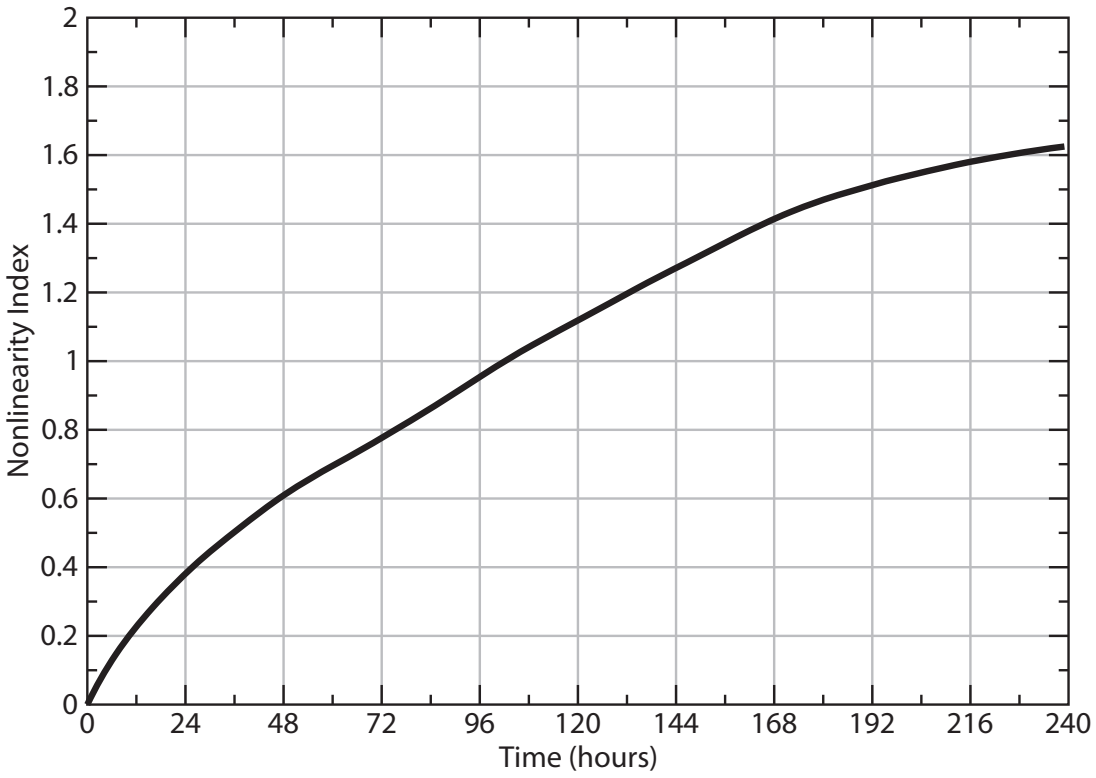

Figure 2: The mean relative nonlinearity for the same sample as shown in Figure 1.

Figure 2 shows the relative nonlinearity measure defined by Gilmour *et al.* (2001). Here, as for Figure 1, we consider integrations initialised by adding and subtracting perturbations to and from an unperturbed state. We also consider a third integration initialised with the unperturbed state. Let us denote by $\delta\psi^+(t)$ the difference at time t between the integration with added initial perturbation and the unperturbed integration. Likewise, the difference between the integration with subtracted initial perturbation and the control integration will be denoted $\delta\psi^-(t)$. The relative nonlinearity is then defined as:

$$\Theta(t) = 2\frac{\|\delta\psi^+(t) + \delta\psi^-(t)\|}{\|\delta\psi^+(t)\| + \|\delta\psi^-(t)\|}. \tag{35}$$

In this paper, the norm is defined as the square-root of perturbation energy.

The relative nonlinearity may be interpreted as the error that would be made if $\delta\psi^+(t)$ were approximated by $-\delta\psi^-(t)$, scaled by the average magnitude of the evolved perturbations. For singular-vector perturbations in a T159 31-level version of the ECMWF forecast model, Gilmour *et al. (op. cit.)* found that the mean relative nonlinearity for 500hpa height reached 0.7 after about 48 hours, at which stage nonlinearity was deemed to dominate the evolution. For the model and initial perturbations used in this paper, the mean relative nonlinearity reaches 0.7 after approximately 60 hours of integration.

4.2 The Analysis System

The experiments presented in the next subsection were conducted using the incremental weak-constraint 4D-Var algorithm described in section 3. In the outer loops of this algorithm, the four-dimensional state $\mathbf{x}^{(n)}$ is updated using the increment $\delta\mathbf{x}^{(n)}$ generated by the inner loop. However, the inner-loop cost function is formulated in terms of $\delta\mathbf{p}^{(n)}$, since this allows standard preconditioning methods to be used (see section 8 for further discussion of this point). Following each inner-loop minimisation, $\delta\mathbf{p}^{(n)}$ is converted to $\delta\mathbf{x}^{(n)}$ by an application of $\mathbf{L}^{-1}$ (i.e. by a forced integration of the tangent-linear model).

The analyses were produced by cycling the analysis system. For each cycle, the analysis window was advanced by 6 hours, regardless of the length of the analysis window. For analysis windows longer than 6 hours, this resulted in an overlap between consecutive analysis windows. We believe that this overlap is an important part of the assimilation algorithm. It allows the initial linearisation trajectory for each analysis cycle to be constructed from the overlapping section of the preceding analysis, with only the final 6 hours of the linearisation trajectory generated from a forecast. In the overlapping portion of the analysis window, the current and preceding analyses are likely to be similar. As a consequence, analysis increments will be small and the linearisation will be accurate throughout the analysis window, even for windows longer than the time-to-nonlinearity implied by figure 2.

The use of the preceding analysis to construct the linearisation trajectory does not violate the statistical correctness of the analysis, since the analysis is not constrained to remain close to the linearisation trajectory. However, it provides an excellent starting point for the minimisation, and helps ensure that the analysis remains close to the global minimum of the cost function.

To avoid correlation between background and observation error, the background state, x_b, must be provided from a non-overlapping analysis. Specifically, we require the total window length to be a multiple of the cycling frequency (6 hours). In this case, there is one analysis window whose end point coincides with the start of the current cycle. The state at the end of that earlier cycle provides the background state for the current cycle.

The first few cycles of a sequence of analyses require special treatment. For the experiments described in this paper, an initial analysis cycle with window length 6 hours is performed. The window length

is increased by 6 hours for each subsequent analysis until the desired window length is achieved, after which the cycling continues as described above. The cycling algorithm is illustrated in figure 3 for the case of an 18 hour analysis window. Note that the first few cycles share the same initial time. This time cannot therefore be used to identify the analysis cycles. Instead, we label each analysis according to the time at the end of the analysis window, as indicated by the black arrows in figure 3.

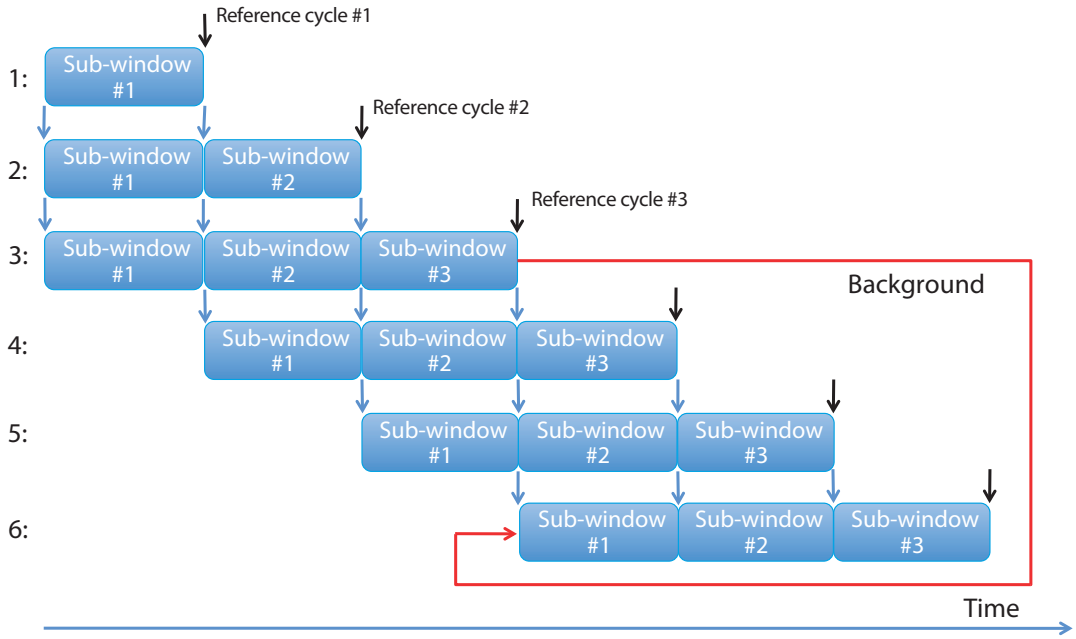

Figure 3: Schematic diagram of the method used to cycle the 4D-Var analysis system. The first six analysis cycles are shown, as labelled on the vertical axis. The horizontal axis represents time. The linearisation trajectory for each analysis is provided by the preceding analysis cycle (as shown by the blue arrows), augmented by a short forecast. The background is provided by the final state of an earlier cycle, as shown by the red arrow.

For all the experiments presented here, observations of non-dimensional streamfunction were taken from a "truth run" of the model at 20 points randomly distributed over both levels. As described above, the truth was generated from a model with layer depths of $D_1 = 6000$m and $D_2 = 4000$m, whereas the assimilating model had layer depths of $D_1 = 5500$m and $D_2 = 4500$m.

The locations of the observations remained fixed for all analysis cycles, and observations were made every six hours. The observations were perturbed by the addition of independent Gaussian noise with unit variance. We note that, even for the longest analysis windows, the number of observations used in an analysis cycle is much smaller than the number of degrees of freedom of the model.

The background error covariance matrix used in these experiments was sub-optimal and had a very simple structure. Background error variance for non-dimensional streamfunction was assumed to be 1.0 at all gridpoints. Vertical correlations were set to 0.2 and horizontal correlations were Gaussian, isotropic and homogeneous, with a length scale of 1000km. Although no great effort was put into tuning these parameters, we believe that they represent a reasonable (but rough) approximation to the climatological covariance structure of background error.

To estimate the statistical characteristics of model error, a set of 6 hour forecasts of the assimilating model was run. At the start of each day of a 50 day period, a set of 50 perturbations from the truth run was generated. Forecasts were run from the perturbed states, and the differences between the forecasts

and the corresponding true states provided a sample of 2500 realisations of model error that were used to estimate covariance statistics.

Figure 4 shows the variance of the model error, plotted for each grid point of the model. The variance is larger in the middle of the domain than at the northern and southern boundaries, where the streamfunction is prescribed and is the same in the truth run and in the assimilating model. The variation of model error variance in the zonal direction (labelled with gridpoint indices 0–40 in figure 4) is a consequence of the zonal variation in orography which induces a stationary component in the climatological flow.

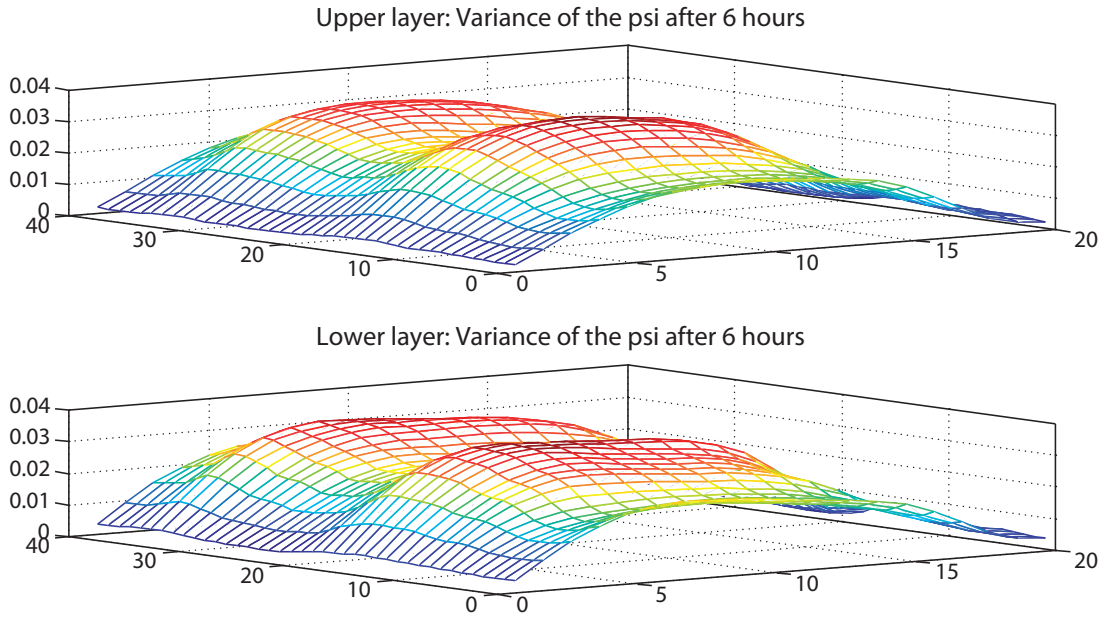

Figure 4: The variance of the model error averaged over a 50 day period.

Given the relatively low dimension of the model, it would have been possible to use the sample covariance matrix of model error directly in the analysis system, possibly with some manipulation to avoid problems of sampling error and rank deficiency. However, we felt that this was unrealistic. For this reason, the covariance matrix used in the experiments presented here was simplified to the extent that it could be represented by just three parameters: standard deviation, horizontal length scale and vertical correlation. The standard deviation was set to 0.1 at each gridpoint, the vertical correlation was 0.5 and the horizontal length scale was 2000km. The assumption of isotropic, homogeneous, white-in-time model error is, of course, highly unrealistic. However, we feel that this lack of realism is not unreasonable, given the current status of model error approximation in numerical weather prediction systems.

4.3 Experimental Results

Figure 5 shows root-mean-square (rms) analysis error for streamfunction as a function of time within the analysis window of the 4D-Var system. Also shown is the error in the first-guess linearisation trajectory. The errors are averaged over both model levels and over analysis cycles 101 to 200 (the first 100 cycles were discarded to remove any possible spin-up effects). Curves are shown for windows of various lengths.

A dominant feature of all the analyses is a rapid growth of analysis error towards the end of the analysis window. The rate of error growth is approximately the same for all lengths of assimilation window.

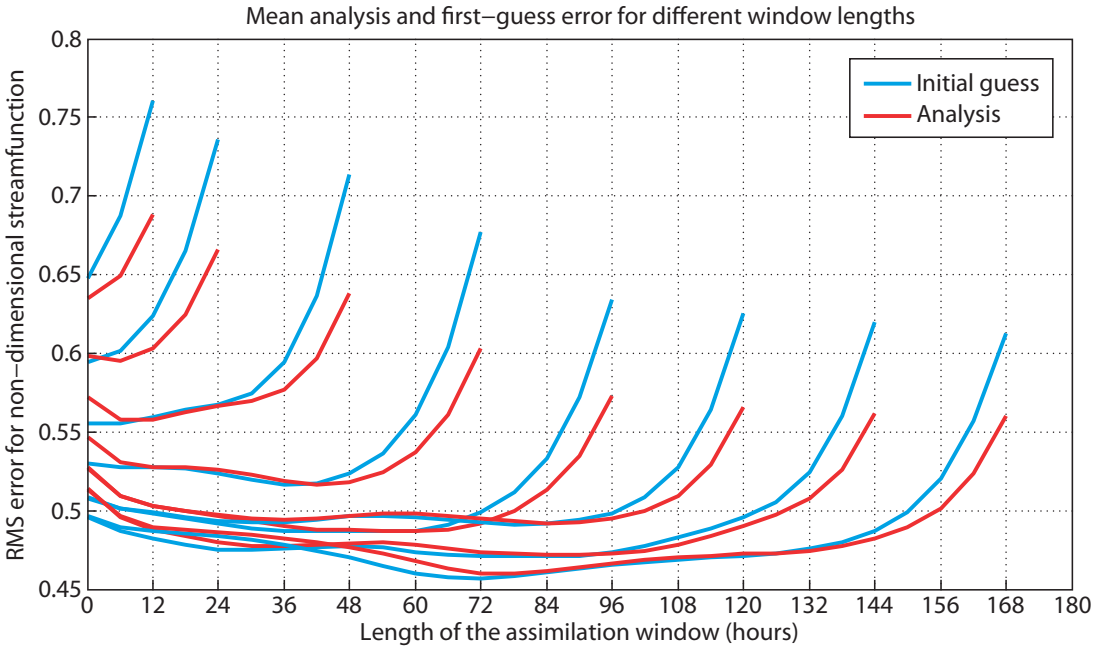

Figure 5: Analysis error for weak-constraint 4D-Var with various window lengths. Each red curve shows the mean rms analysis error as a function of time within the analysis window, for a given window length. Blue lines show the corresponding error in the first guess.

Taking the 48 hour window as typical, we see that the error increases from approximately 0.578 to 0.639 over the final 12 hours of the window, which corresponds to an error doubling time of approximately 83 hours. This is significantly longer than the perturbation doubling time for the model, as diagnosed from figure 1, suggesting that the final few hours of the assimilation window represent a transition between the central part of the window in which the analysis has sufficient information to control error growth, and the free-running forecast model in which error growth is rapid.

Note that for window lengths longer that 12 hours, the first guess error is smaller than the analysis error in the early part of the assimilation window. To explain this, it is important to remember that, with the exception of the final six hours of the window, the first guess for each analysis is equal to the preceding analysis. That is, the first guess is an analysis produced using an analysis window that starts six hours earlier than the window used to produce the curves labelled "analysis". The first guess benefits from observations during the first six hours of the window, that are not used in the current analysis. As a consequence, the first guess is more accurate than the corresponding analysis at the start of the window. By contrast, additional observations at the end of the analysis window, that were not used to produce the first guess, make the analyses more accurate than the first guess in the latter half of the window.

For the analyses shown in figure 5, the mean model error ($\bar{q}$ in equation 16) was assumed to be zero. In practice we found that better results could be obtained by taking into account the mean component of model error. Initially, we set $\bar{q}$ equal to the mean of the sample of 2500 realisations of model error described above. However, this resulted in analyses that were significantly worse than the experiments in which $\bar{q}$ was set to zero. After some experimentation, we found that a positive impact on analysis skill could be achieved by reducing the assumed mean model error. We do not yet understand why this reduction was necessary.

Figure 6, shows analysis and first guess errors for experiments in which $\bar{q}$ was set equal to the computed mean model error divided by the number of sub-windows. For all window lengths the analysis error is

reduced compared with the experiments shown in figure 5.

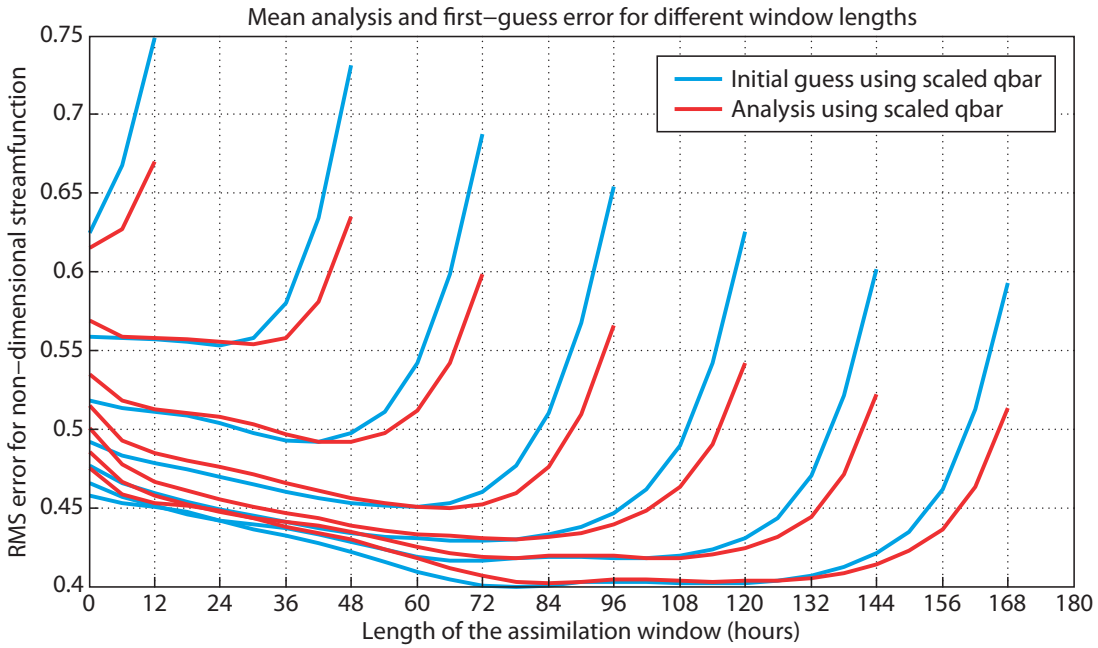

Figure 6: Analysis error for weak-constraint 4D-Var using a scaled mean model error correction with various window lengths. Each red curve shows the mean rms analysis error as a function of time within the analysis window, for a given window length. Blue lines show the corresponding error in the first guess. Please note: the vertical axis differs from that used in figure 5.

It is clear from figures 5 and 6 that increasing the length of the analysis window is beneficial, despite the lack of realism in representing the statistics of model error. We believe that the model error induced by perturbing the layer depths of the model is a reasonable analogue for the type of error that can be expected in a numerical weather prediction model, in so far as the error is flow dependent, time correlated, anisotropic, inhomogeneous, and has a significant mean component. It is encouraging, therefore, that we have been able to demonstrate the potential of long-window, weak constraint 4D-Var in a system with such error characteristics and in which the statistics of model error are poorly described.

Our experimentation also suggests that it is important to take into account the mean component of model error. It is notable that, whereas in figure 5 analysis error at the end of the window showed only minor improvement for windows longer than 96 hours, the corresponding error in figure 6 continues to show significant decrease as the window length increases to 168 hours. This difference in behaviour suggests that it is necessary to properly account for mean error if the full benefit of extending the analysis window is to be gained.

We believe that mean error is a significant component of model error in all numerical weather prediction systems, and that it is incorrect to represent model error as a zero-mean Gaussian process. We do not yet understand why it was necessary, in our system, to reduce $\bar{q}$ from its calculated value, but there is a clear sensitivity to this parameter.

The most accurate estimates occur near the centre of the analysis window. This is expected, as the analysis system can take advantage of observational information several hours before and after the analysis time. By contrast, the estimate at the end of the window is based on information from the past only. This has clear implications for re-analysis, should a long-window analysis system become computationally feasible for a future re-analysis system. It also has interesting ramifications for *post hoc* verification of

analyses and of short-range forecasts.

As already discussed, the first guess comes from the preceding cycle of analysis, which benefits from an additional 6 hours of observations at the start of the window but which did not have access to observations for the final 6 hours of the current window. We therefore expect the minimum in the analysis error to occur 6 hours after the minimum in the first guess error curve. For long windows, we expect these minima to be nearly identical, since the addition and removal of observations at the ends of the window should have only a small influence on the analysis at the central point. This appears to be the case for at least some of the curves in figure 6. However, by the same argument, we would expect the minimum analysis error to saturate for sufficiently long windows. This does not appear to be the case in our experiments. There is a continual improvement in analysis error as the window length is increased to 168 hours.

It is clearly necessary, for statistical consistency, to avoid correlation between the background and the observations used in any analysis cycle. However, we note that this results in a background that is significantly less accurate than the first guess. In principle, provided the errors in the background are correctly described, it is still possible for the background to contribute information to the current analysis. However, it is interesting to speculate whether a more accurate analysis at the start of the window could be obtained by allowing, or constraining, the analysis to remain closer to the the first guess at the start of the window.

5 24h 4D-Var design and results.

5.1 Formulation

A fully four-dimensional control variable implementation of 4D-Var, as described in section 2, is not possible at the moment in the IFS. Several scientific and technical challenges must be addressed before reaching that goal, and simplifications are still necessary. For practical reasons, in particular in implementing the outer loop iterations, the weak constraint 4D-Var implementation in the IFS is limited to the "forcing" formulation, with model error assumed to be constant over the entire analysis window. With this simplification, equations (17) and (18) from section 2 become:

$$\mathbf{p} = \begin{pmatrix} x_0 \\ q \end{pmatrix} \tag{36}$$

and

$$x_k = \mathcal{M}_k(x_{k-1}) + q \tag{37}$$

where q is now independent of k.

The assumption that model error is constant over a certain period of time makes weak constraint 4D-Var practical today. It also means that it is mainly the systematic model error, or model bias, that is estimated. Since there is strong evidence of model bias in the IFS, especially in the stratosphere, it seems reasonable to address this aspect first.

The model error term was defined in equation (16). With the assumption that model error is constant over the assimilation window, it becomes:

$$J_q = \frac{1}{2}(q-\bar{q})^{\mathrm{T}} Q^{-1}(q-\bar{q}) \tag{38}$$

Model bias is expected to vary little from its mean, $\bar{q}$. This is reflected in the covariance matrix Q, which heavily penalises departures of q from $\bar{q}$ at each analysis cycle. As a consequence, the model bias

estimated at one analysis cycle is a good estimate of the mean model bias $\bar{q}$ required by the next cycle. This observation allows us to avoid the need for a separate estimation scheme for $\bar{q}$. By taking $\bar{q}$ to be equal to the model bias estimated by the preceding analysis cycle, we allow $\bar{q}$ to evolve slowly at a rate determined by the specified covariance matrix Q. This scheme was implemented operationally in IFS cycle CY37R3.

The disadvantage of the current cycling method for $\bar{q}$ is that it makes Q play a double role. The matrix describes the statistics of the random component of model bias, but also determines the rate at which $\bar{q}$ varies. We intend to remove this confusion of roles in the future by implementing a separate (probably variational) model bias estimation scheme.

5.1.1 Longer assimilation window

The potential advantages of using longer assimilation windows are discussed in section 4. However, longer windows require that model error is accounted for in the assimilation system. Because of the simplifications described above, it is not expected that very long windows (of the order of a few days) can be used yet in a real system. However, extending the assimilation window beyond 12 hours should be possible.

Experimentation has been conducted with an extended assimilation window, leaving all other parameters unchanged. The system used for these experiments was IFS cycle CY37R3, which includes the model bias term described above, but with model bias restricted to be non-zero only in the stratosphere. In that configuration, the forecast performance is slightly degraded, in particular in the Southern hemisphere (figure 7), but the gap even without any tuning is small enough to justify pursuing that direction.

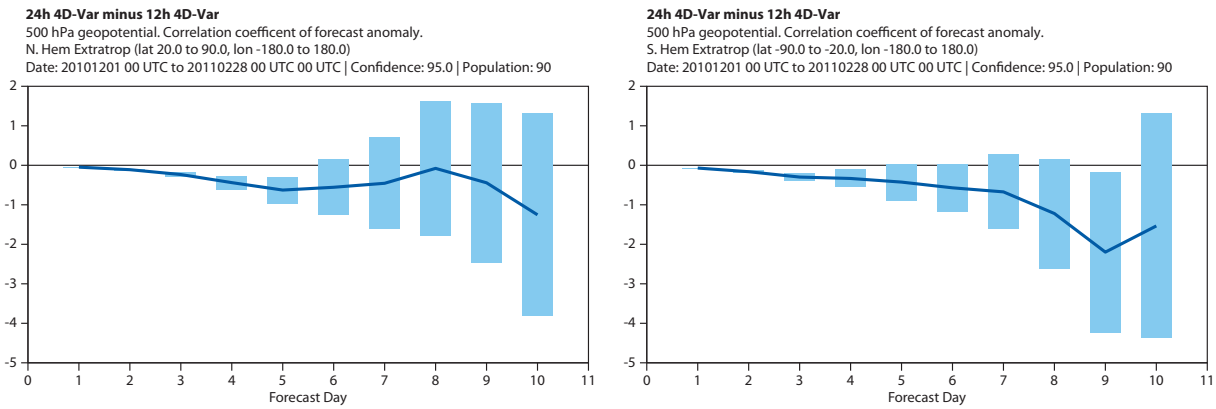

Figure 7: Mean forecast score difference for 500hPa anomaly correlation, averaged over 90 cases. Differences (in percent) are shown between non-overlapping 24h 4D-Var and a control 12h 4D-Var in standard IFS configuration. The left and right panels show differences for the Northern and Southern Hemispheres respectively. Vertical bars show 95% confidence intervals.

The first experiment consisted of running 4D-Var with a 24h assimilation window every 24 hours. The analysis windows followed each other sequentially in time, as is normally the case in 4D-Var. However, in an operational environment, updates are required more frequently than once ever 24 hours. 4D-Var must be run at least twice daily. To address this, the cycling has been modified so that 4D-Var with a 24h assimilation window can be run every 12 hours, resulting in overlapping analysis windows, as described in section 4.

It is already technically possible to run experiments with assimilation windows longer than 24h. With

the simplifying assumptions made in the IFS implementation, the conditioning of the problem is almost independent of the length of the assimilation window and the overall cost of the system is roughly proportional to the length of the assimilation window. For example, 4D-Var with a 48h window run twice daily is approximately four times more expensive than the current 12h 4D-Var. 4D-Var with a 48h window will be evaluated for research purposes in the near future.

5.2 Model error covariance matrix

A lot of research has gone into estimating the background error covariance matrix at all operational centres but almost nothing is known about the model error covariance matrix that appears in the weak constraint 4D-Var cost function and in the Kalman filter equations.

Indirectly related to the data assimilation context, some research has taken place to diagnose forecasting models. For example, some studies have recognised the link between systematic initial tendencies and model error for some time (see, e.g., Klinker and Sardeshmukh; 1992) and used it to assess climate models (Rodwell and Palmer, 2007) or to demonstrate an improvement in model aerosol climatology (Rodwell and Jung, 2008).

Based on short range model tendency statistics or longer range model drift statistics, two types of model error covariance matrices have been used so far in weak constraints 4D-Var experiments in the IFS, overall with similar results. We describe the two covariance models below.

5.2.1 Model tendency covariances

The first model error covariance matrix to be used with the IFS is described by Trémolet (2007). In the current ECMWF data assimilation system, the background error covariance matrix B is estimated from an ensemble of 4D-Var analyses. Since each ensemble member represents a plausible atmospheric state, the model tendencies derived from the states of ensemble members should represent a plausible distribution of the possible evolutions of the atmosphere. 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 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.

To determine the model error covariance matrix, forecasts were run from the 4D-Var ensemble at the spectral resolution of T319 and tendencies were 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 discarded in order to reduce the dependence on the initial state and to avoid 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.

5.2.2 Model drift covariances

The second model error covariance matrix to be used in the IFS is based on samples of model drift and variations in model drift. As explained above, the aim of the current weak constraints 4D-Var formulation is to capture systematic model error. One visible effect of systematic model error is model drift: forecasts at different ranges do not have the same average characteristics. Seasonal averages of differences between day-5 and day-10 forecasts are for example sometimes used by model developers to diagnose systematic problems in the model.

In the formulation proposed here, the goal is also to capture this systematic error. The definition of the model error term in the cost function shows that Q should account for the statistical properties of the variation of model error from one assimilation cycle to the next. The easiest proxy to access for that quantity is the change in model drift from one day to another. Only one realisation of model drift is available for each forecast which does not constitute a proper statistical sample. However, since this is available for every forecast, statistics over a long enough period of time can be used, as was the case for background error in many operational systems until recently with the "NMC" method.

A drawback with this method is the fact that forecast differences depend on model error and initial condition. The assumption being made here is that at long enough ranges, the contribution from model error is larger than the contribution from the initial condition error. In the experiments presented here, the model drift samples were obtained using day-5 minus day-10 forecast differences over a period of one year from June 2009 to May 2010. This sample was used to generate a covariance matrix using the same statistical model as above.

5.2.3 Comparison results

Figure 8 shows the impact of the two model error covariance matrices on the model error estimate. It is clear that tendency-based covariances lead to smaller scale structures in the model error field. This is expected as tendencies are instantaneous small scale responses from the model to a given state and lead to shorter correlation length scales than model drift, which is an accumulated response over several days.

The model error estimates show some similarities, for example a large scale cooling from the top of the model at the South pole down to around 5hPa above the equator, with warming above and below that. In the Northern hemisphere, the two model error estimates are quite different, especially above 2hPa. Despite these differences, the two systems lead to very similar forecast performances and it is very difficult to determine which covariance matrix is better. A smoother estimate might be preferable and more realistic if longer term systematic model error is sought.

5.3 Experimental results

All results presented in this section are for experiments run at T255 with two minimisations at T95 and T159, using the full operational observing system. The results of section 5.3.1 were obtained with the tendency-based model error covariance, all other results were obtained with the drift-based model error covariance matrix.

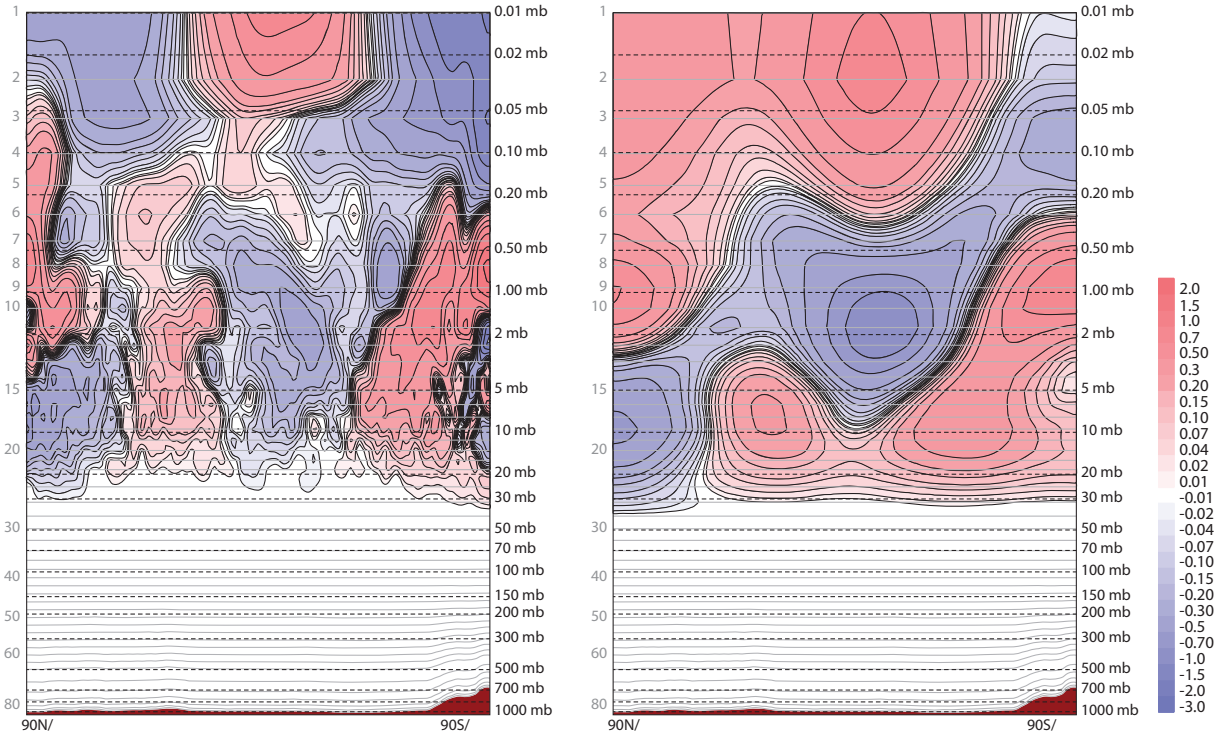

Figure 8: Zonal mean of average temperature model error (in K per 12 hours) for the month of August 2010 with the tendency-based model error covariance matrix (left) and the drift-based model error covariance matrix (right).

5.3.1 Model error cycling

Up to now, the mean model error has been estimated in the IFS cycle by cycle, with no transfer of information from one cycle to the next. That is, $\bar{q}$ in equation 38 was set to zero every cycle. Experiments have shown that there are benefits in cycling the model error information, by setting $\bar{q}$ equal to q from the preceding cycle. Figure 9 shows a time-series of AMSU-A channel 13 statistics for the Northern hemisphere and analysis increment and model error statistics for the model level where this channel is most sensitive. Without cycling, the mean background departure (red) and analysis increment (green) stay negative throughout the period. When model error is cycled, the average background departure stays around zero and the average analysis increment is also reduced, showing that weak constraint 4D-Var can in that case compensate for some of the model systematic error. The mean increment is still not zero even with model error cycling because of the presence of AMSU-A channel 14 which is not bias corrected and to which the level shown has some sensitivity. This configuration will become the default in the next IFS implementation (CY37R3). However, as model error is used only in the stratosphere, the impact on forecast performance is limited.

5.3.2 Observation statistics

The main modification of the assimilation system for running with a 24h assimilation window was to make it possible to run 4D-Var twice daily. This involved many technical modifications throughout the system and more are necessary before such a configuration can be run operationally (for example to allow proper archiving in the overlapping part of the window). However, research experiments can be performed with some limitations regarding the diagnostics that can be produced because part of the data

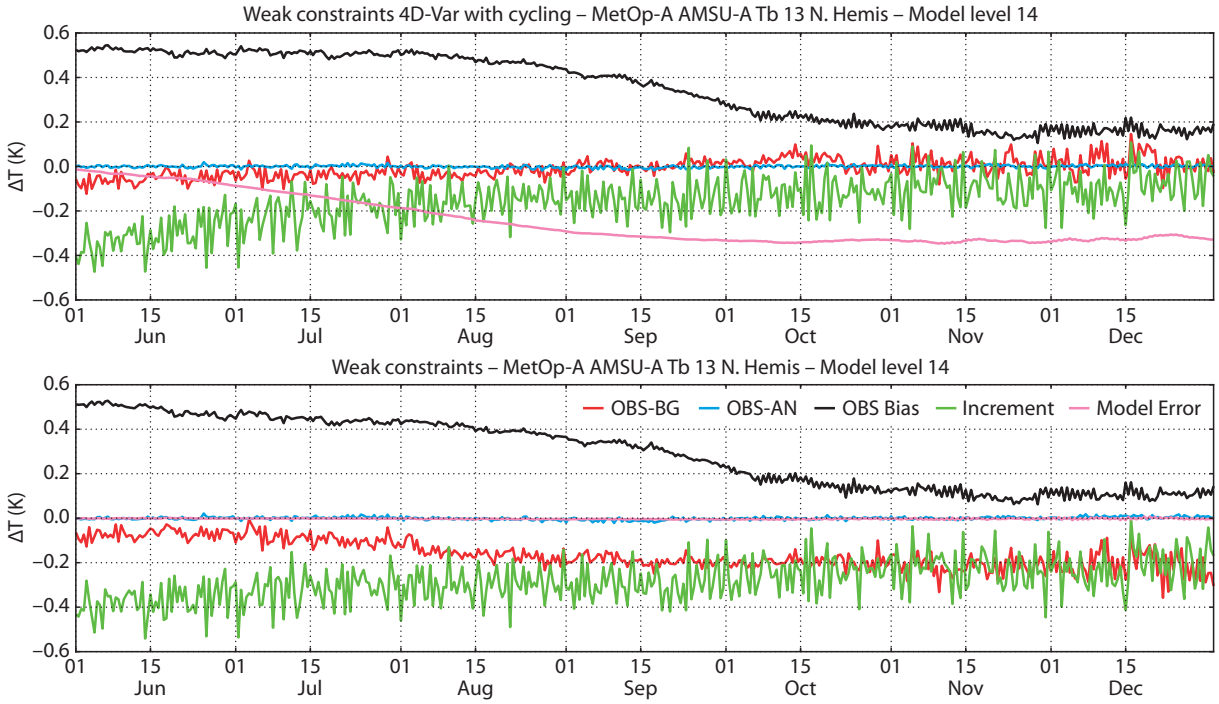

Figure 9: Time-series of mean background and analysis fit to observations (red and blue respectively), observation bias correction (black), analysis increment (green) and model error (magenta) without (top) and with (bottom) model error cycling. The analysis increment and model error statistics are computed at model level 14 where AMSU-A channel 13 has the highest sensitivity.

is over-written when the next overlapping analysis cycle runs. For example, forecasts issued from the two analyses valid at the same time but produced by two successive 4D-Var cycles cannot be compared since only one analysis can be saved.

Figures 10, 11 and 12 show statistics for wind profiler, airep and radiosonde observations. The control experiment, shown in red, used a 12h assimilation window. For consistency, the statistics for the 24h assimilation window (shown in black) only account for the observations used in the last 12 hours of the window. The first two figures show a very positive impact of the 24h assimilation window. More observations are used for these two observing systems. The background departures have less bias at all levels and have a slightly reduced standard deviation for the wind profilers. Analysis departure biases are unchanged but standard deviations are larger. This is a consequence of the fact that, compared with a 12h system, 24h 4D-Var has the extra constraint that it must fit observations during the first half of the window.

However, figure 12 shows that the impact on background-departure biases can also be negative. This is particularly the case here between 400 and 200 hPa. Comparing with figure 11, we see that the 24h system draws more to aircraft temperature measurements, negatively affecting the fit to radiosondes. The bias in GPSRO background departures (not shown) is shifted, which is indicative of a warming of the troposphere and/or cooling of the stratosphere, consistent with the change in radiosonde background departures. This is unfortunate since aircraft measurements are known to be biased. It is a consequence of the constant model error approximation. A formulation with a constant forcing term makes it possible to fit biased observations that are frequent in time at low cost, and this is favoured by the optimisation algorithm. This behaviour has already been observed by Trémolet (2007).

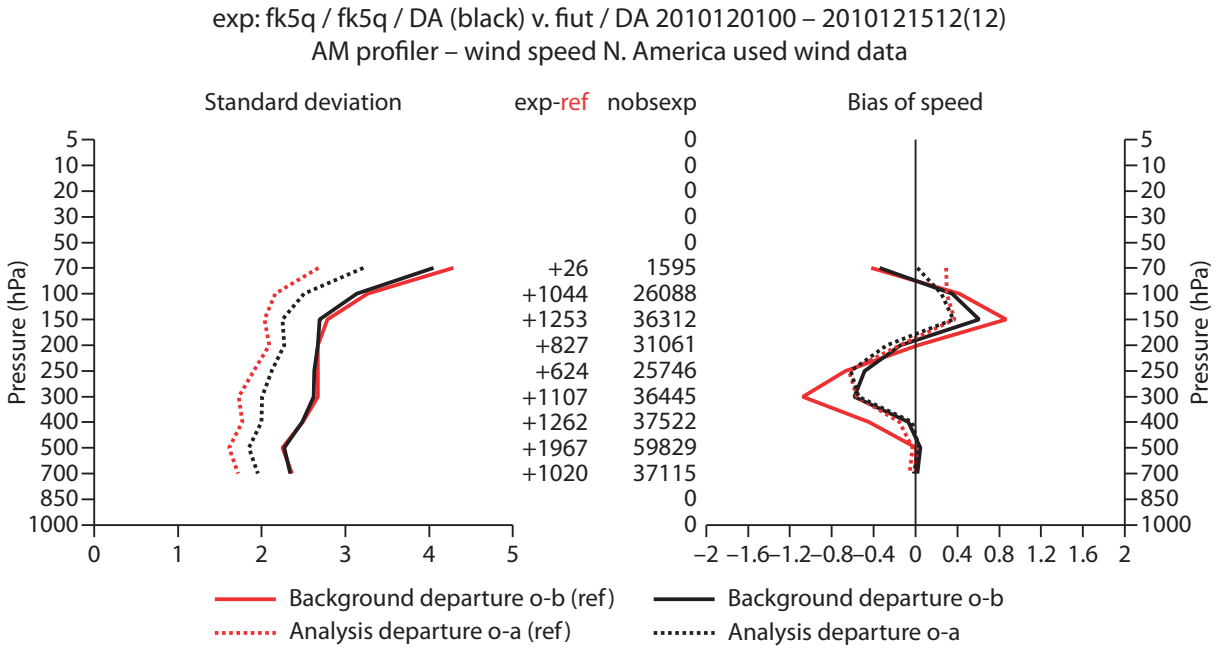

Figure 10: Wind profiler observation statistics for 24h 4D-Var (black) and 12h 4D-Var (red).

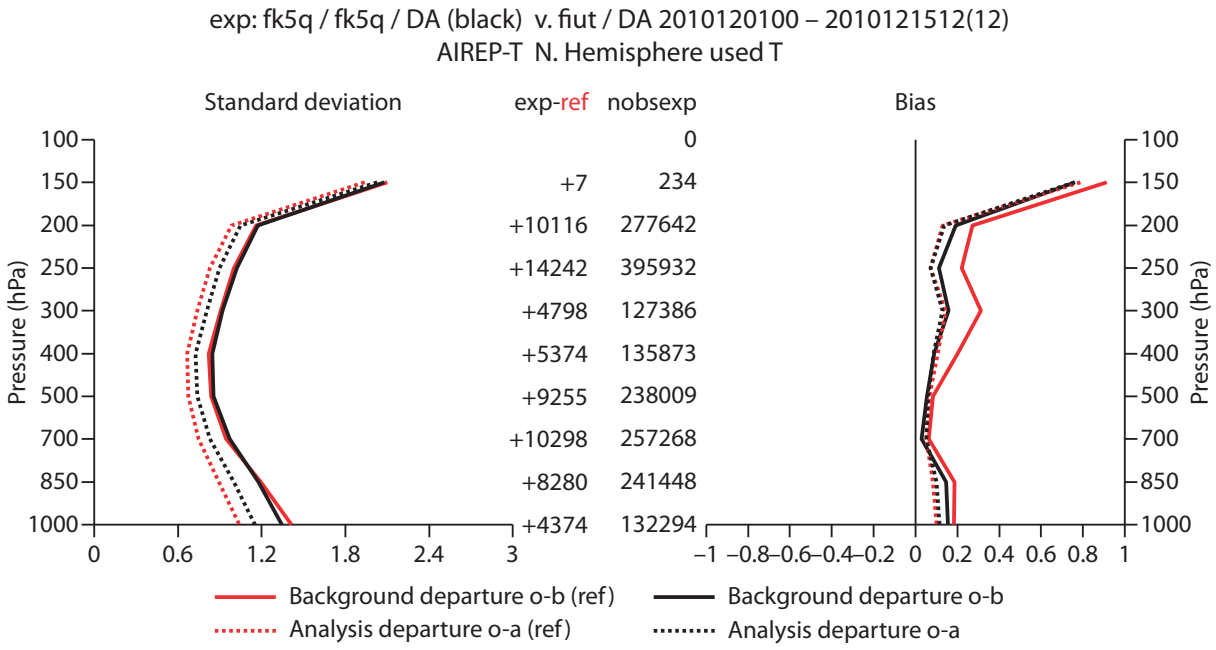

Figure 11: Airep observation statistics for 24h 4D-Var (black) and 12h 4D-Var (red).

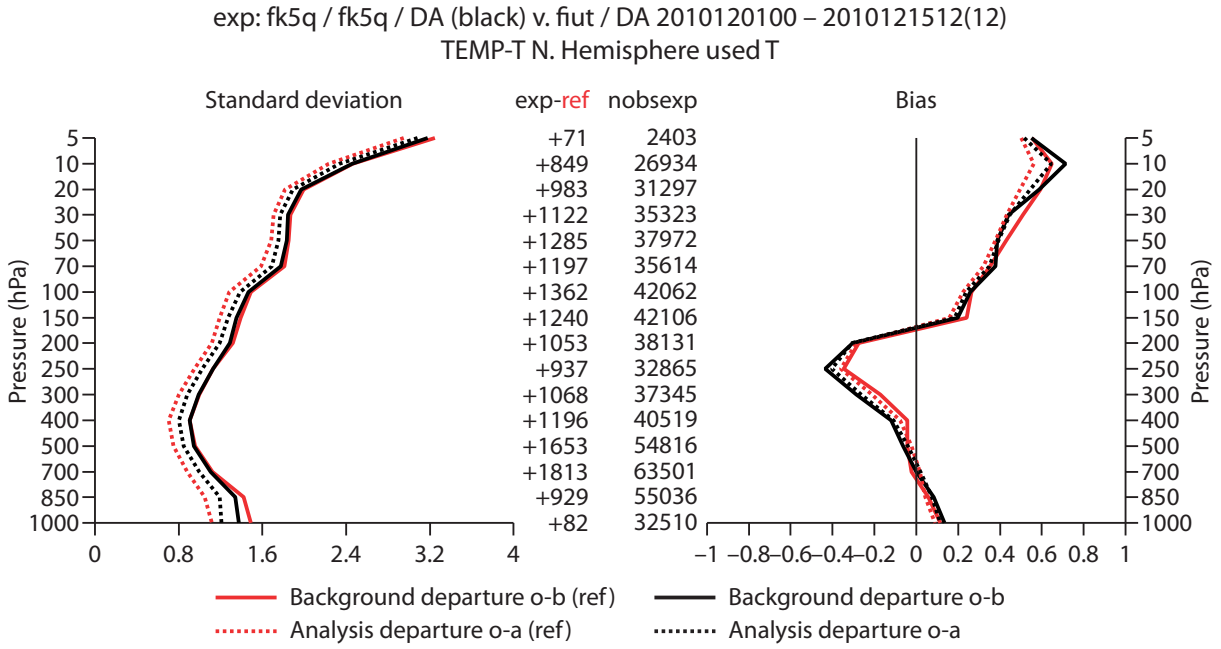

Figure 12: TEMP observation statistics for 24h 4D-Var (black) and 12h 4D-Var (red).

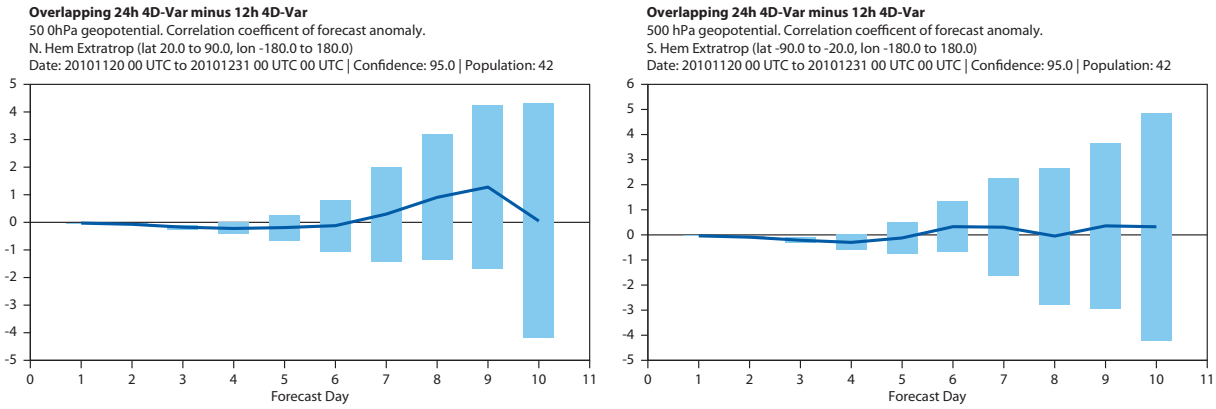

Figure 13: Mean difference in forecast anomaly correlation between overlapping 24h 4D-Var with and a control 12h 4D-Var in standard IFS configuration. Scores are for 500hPa anomaly correlation averaged over 42 cases for the Northern Hemisphere (left panel) and Southern Hemisphere (right panel). Vertical bars show 95% confidence intervals.

5.3.3 Forecast skill

At this stage of experimentation, the forecast scores for the 24h overlapping window system suggest a small degradation up to day 4, and an improvement later in the forecast for both hemispheres (figure 13). However, both the apparent improvement and the degradation could be subject to sampling error and will have to be confirmed by longer experiments and experiments at full operational resolution.

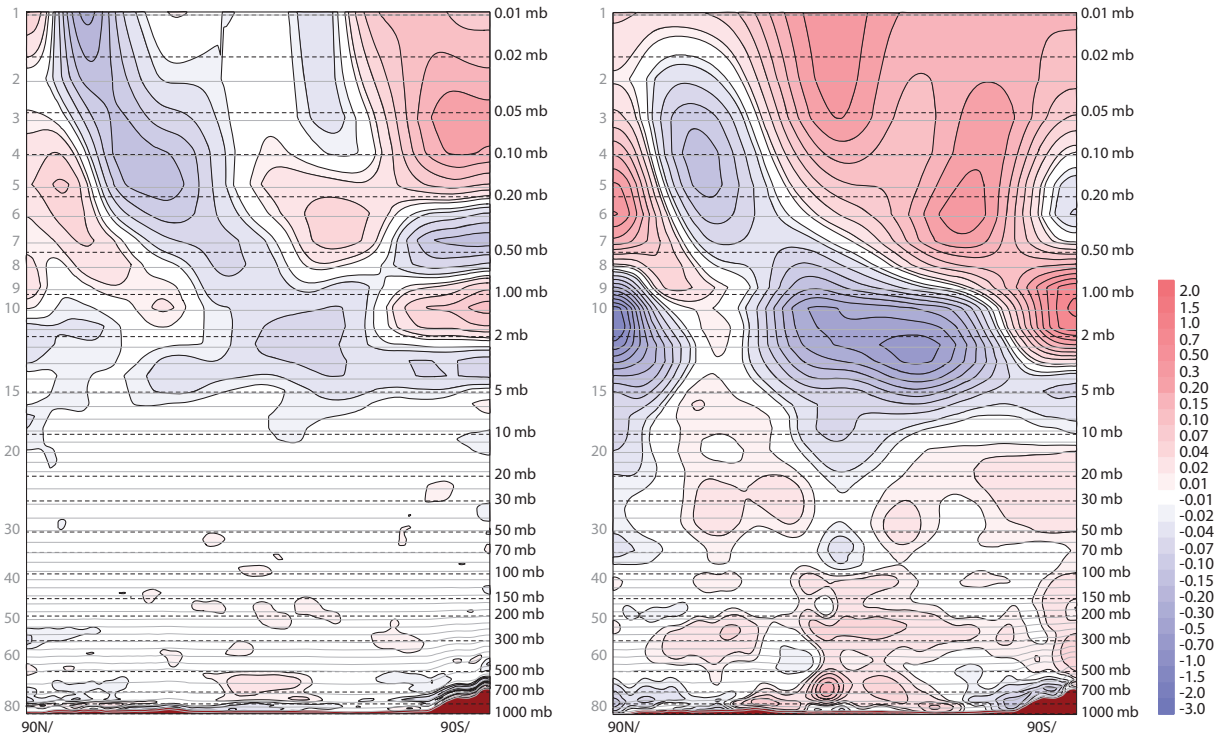

Figure 14: Zonal mean of the average temperature initial condition increment (left, in K) and model error (right, in K/day) for the month of December 2010 with the 24h 4D-Var overlapping window.

5.3.4 Model error estimates

Figure 14 gives an overview of the temperature model error and initial condition increment estimates as zonal mean cross sections for one month (December 2010). At first glance, both model error and initial condition increment have the same general pattern. This indicates that there is still a relatively large component of the systematic error that is attributed to the initial condition by 4D-Var. That seems to indicate that the variances assigned to model error in the cost function are too small. Larger values have been tested but showed a degradation in the fit to observations (not shown). The model error also shows a systematic warming at around 200 hPa, most likely related to biased aircraft observations.

Figure 15 shows the standard deviations corresponding to the averages shown on figure 14. Both the initial condition increment and the model error have large variability above 10 hPa over the winter pole. In both cases, the standard deviation is as large as or larger than the mean value. Since the initial condition error is assumed unbiased in 4D-Var, this is as expected. However, if the model error estimate is supposed to capture the systematic component of model error, its variability should be smaller compared to its mean. In particular, this figure shows that in the winter stratosphere, the estimates have not converged to a mean value. The same is true at around 200 hPa where the standard deviation of the model error estimate is as large as the mean.

In all experiments presented here, model error is estimated within each assimilation cycle, and is used as a first guess and possibly as a "background" ($\bar{q}$) value for the next cycle, but it is not applied as a forcing term during the forecasts. Such experiments have been performed but always lead to very poor performance. In principle, if the estimate was an accurate representation of the mean slowly varying model error, applying it during the forecast should improve the forecast. Figure 15 shows that there is a large variability in the model error estimate which makes it very unlikely that model error determined on

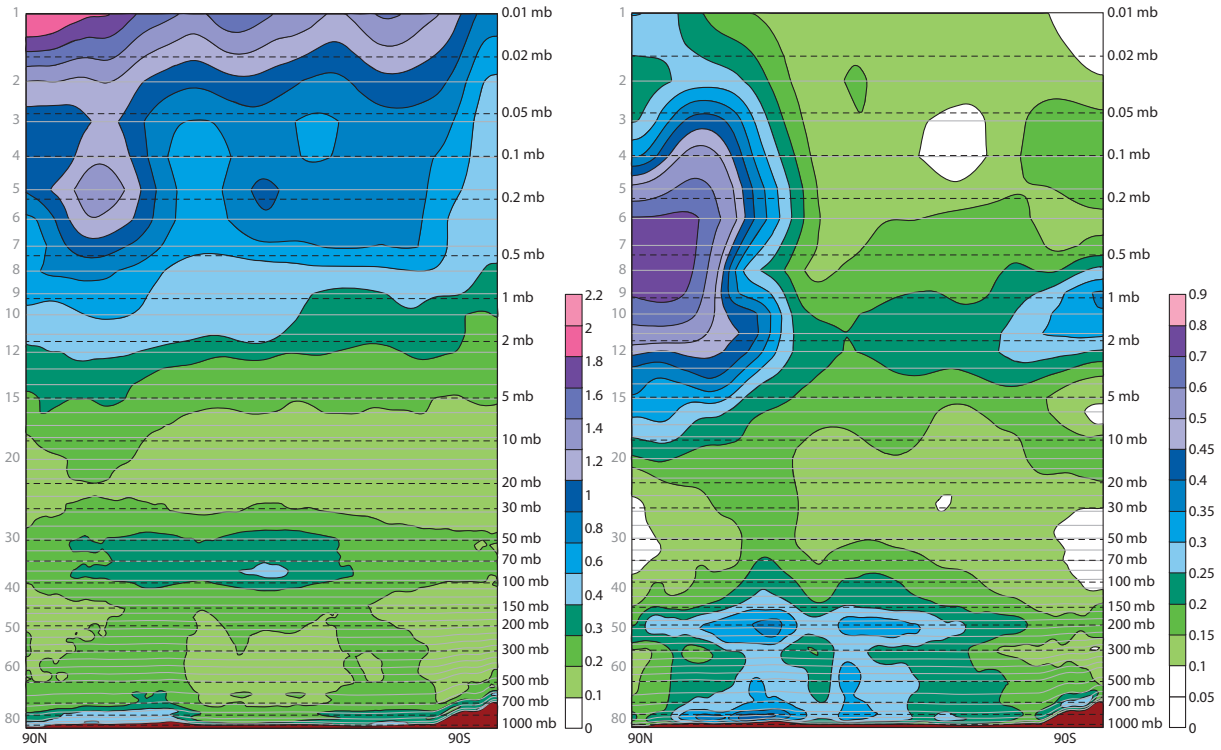

Figure 15: Zonal mean of the standard deviation of the temperature initial condition increment (left, in K) and model error (right, in K/day) for the month of December 2010 with the 24h 4D-Var overlapping window.

a particular day will be meaningful a few days later in the forecast. This most likely explains the poor performance of the forced forecasts.

5.4 Discussion

All the results obtained with the weak constraint 4D-Var implemented in the IFS show some encouraging signs but also point to weaknesses in the system. There are two main areas where the system should be improved.

The first area for improvement is the specification of the model error covariance matrix. Two approaches have been tested at ECMWF, based on model tendencies statistics and on model drift statistics. Both of these methods could be improved upon by using the EDA and the EPS to gather more samples of model tendencies and model drift, possibly leading to the use of a flow dependent component in the Q matrix.

Another on-going direction for research in this area aims to determine a projection of the model error covariance matrix into observation space (collaboration with R. Todling, GMAO). This could potentially allow comparison of the model error covariance matrices used at ECMWF with a more rigorous statistical estimation. Unfortunately, this projection into observation space is limited as it requires a fixed observing network which is the case only for a very limited subset of the global observing network. This projection cannot directly provide a model error covariance in model space but can be used for diagnostics. We are not aware of any other attempt to estimate model error statistics although this is clearly an area where much more research is needed.

The second main area for improvement is the formulation of the weak constraint 4D-Var problem. Cur-

rently, the assumption of a constant model error forcing is a strong limitation. A proper representation of the random component of model error is also needed in order to fit the observations as well as or better than with the 12h assimilation window. This should improve the short range forecast performance by allowing the analysis in the later part of the window to draw closer to observations. A better separation between the methods used to estimate the systematic and random components of model error will be important.

Removing the constant model error assumption poses both scientific and technical challenges. Implementing another formulation of weak constraint 4D-Var in the current IFS code would be extremely difficult and re-structuring the code is a prerequisite to further work in this area. The Object Oriented Prediction System (OOPS) project is currently under way in order to address this issue and, more generally, to make the future code more flexible, reliable and efficient. It will make all formulations of weak constraint 4D-Var technically achievable and will be the base for future scientific developments.

Within the current IFS implementation, it is however possible to run in a mixed mode where the inner loop of 4D-Var uses a forcing formulation while the outer loop is split and uses a four dimensional state formulation. In practice, and among other difficulties, attempts at using this implementation have so far been hindered by instabilities of the tangent linear model in the top few levels. A side benefit of this implementation is that the screening is performed in two 12h tasks, reducing the necessary amount of memory required. In turn, the minimisation can then also be run with fewer processors since memory is not a limiting factor, leading to a more efficient use of the supercomputers.

Finally, although independent of the work on weak constraint 4D-Var, it is important that observation biases are properly accounted for. The current forcing formulation of weak constraint 4D-Var is very sensitive to frequent biased observations. It will be very interesting to study the impact of aircraft variational bias correction in the next IFS cycle on model error estimates and on weak constraint 4D-Var performance in general.

6 Impact of Cycling Model Error in Surface-Pressure-Only Reanalysis Experiments

6.1 Context of a Surface-Pressure-Only Reanalysis System

The treatment of model error described in this paper is being developed primarily for ECMWF's operational forecast system, but it is also relevant in a reanalysis context and is especially timely in the light of current reanalysis activities.

Production of a next-generation global atmospheric reanalysis spanning the entire 20th Century is planned to begin in 2014 and preparatory work is already underway via the ERA-CLIM project funded by the European Commission's seventh Framework Programme. In addition to many well-established applications, reanalyses are being used increasingly by the international research community as a reference atmospheric dataset for climate monitoring. Emerging applications for ECMWF reanalyses include the evaluation of climate data records derived from long-term satellite-based data products, e.g. in ESA's Climate Change Initiative. Such developments provide further reasons to improve reanalysis quality.

Maximising the consistency of reanalysis products continues to be an important goal. The use of a fixed data assimilation system for the entire reanalysis period does much to impart consistency in reanalysis products, and assimilation developments such as Variational Bias Correction (VarBC) counteract some limitations of the observing system (e.g. imperfections in instrument calibration procedures). However,

major changes in the observing system still have the potential to introduce undesirable artefacts. The much-reduced observing system in early decades of the reanalysis period is a significant consideration when attempting century-long reanalyses. Whitaker *et al.* (2009) demonstrated that useful analyses could be produced given a network of sparse surface pressure observations similar to what is available for the 1930s, by using modern advanced four-dimensional assimilation schemes such as 4D-Var and the Ensemble Kalman Filter, which both provide flow-dependent covariance information. Nonetheless, the much-reduced observing system makes the assimilation problem more under-constrained and allows systematic model errors to dominate away from the Earth's surface. A reanalysis system is thus a demanding context for demonstrating the effectiveness of the weak-constraint 4D-Var approach. As a first step, we obtain model error estimates by cycling weak-constraint 4D-Var given a modern observing system, and then investigate whether these estimates can counteract the systematic model errors that dominate for a much-reduced observing system.

6.2 Estimate of model error covariance matrix

For the experiments described in this section, we use the "model drift" covariance matrix described in section 5.2.2. The forecast differences required to calculate the covariance matrix were taken from ECMWF Operations for the period June 2009 to May 2010, retaining all 91 model levels but applying spectral truncation at T255 in the horizontal to emphasise the broader horizontal scales. By taking forecast differences for a full 12 month period, the covariance estimate covers a complete annual cycle and is used in the "cycled model error" experiments below (but not in the Control experiments).

6.3 Impact of Cycling Model Error

Experiments to cycle model error were conducted using modifications to IFS cycle CY36R4 (the operational cycle from late 2010). Our experiments were performed at reduced horizontal resolution (T255) and used two scenarios for the observing system (Table 1). The first scenario used a modern observing system consisting of all observations used operationally from November 2009 onwards. The second scenario used a much-reduced observing system, covering the same period but restricted to only the surface-pressure observations. For comparison purposes, control experiments were run for both observing system scenarios. In the surface-pressure-only assimilations, withheld observations (e.g. satellite radiances and GPSRO) have been retained in passive mode to provide additional diagnostics.

Observing system	Assimilation method	Comments
Full-obs	Control (CY36R4).	Full, modern observing system as at November 2009.
Full-obs	Cycled model error, background from previous assimilation cycle of this experiment.	Provides background model error for Ps-only.
Ps-only	Control (CY36R4).	
Ps-only	Cycled model error, background from Full-obs.	

Table 1: Experiments to test model error cycling in a reanalysis system. For more details, see text in Section 6.3.

As described in sub-section 6.1 above, an important first step is to obtain model error estimates by cycling weak-constraint 4D-Var given a modern observing system. After seven months of cycling, the resulting estimate of model error is shown in figure 16a (zonal mean for July 2010). Note that model error estimation is only attempted in model levels 1 to 28, corresponding approximately to the stratosphere above

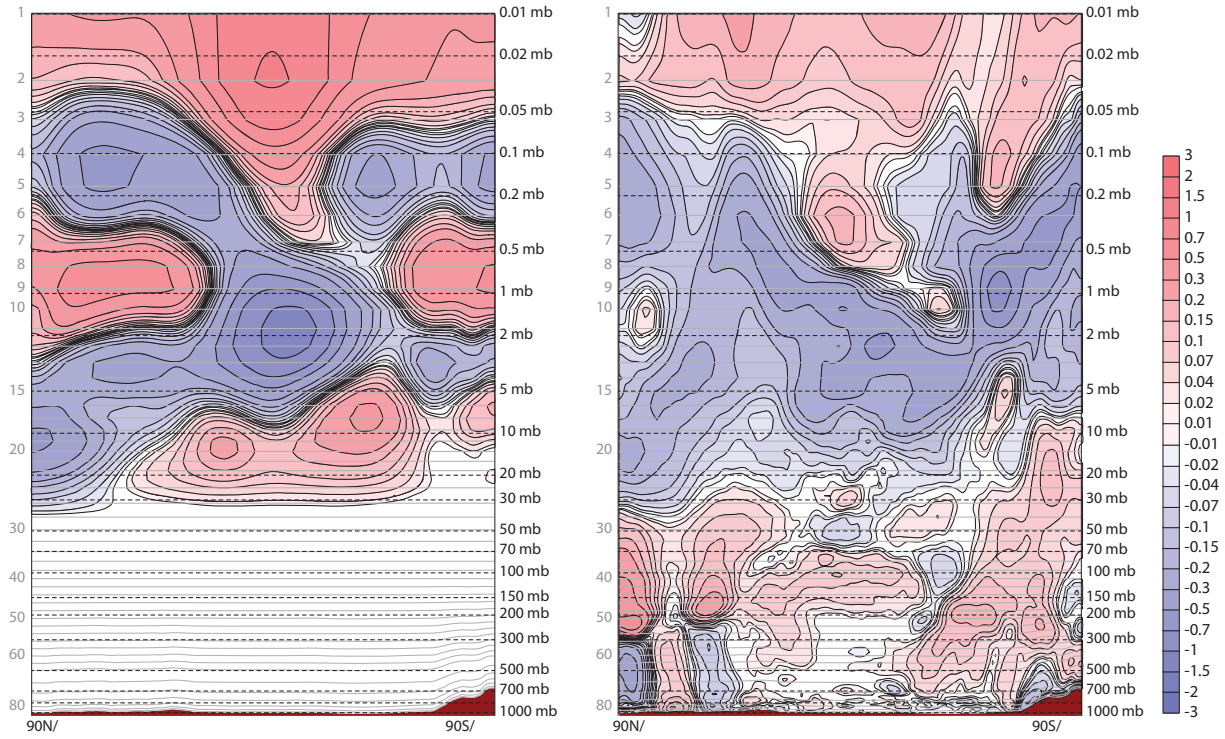

Figure 16: a: Zonal mean model error for July 2010 (temperature error in units of K per 12h) estimated by cycling with a modern (2009/2010) observing system. Estimation only attempted in model levels 1 to 28. b: Zonal mean forecast drift for July 2010 (Day-10 minus Day-5 temperature in units of K per day) from the corresponding Control (Full-Obs) experiment.

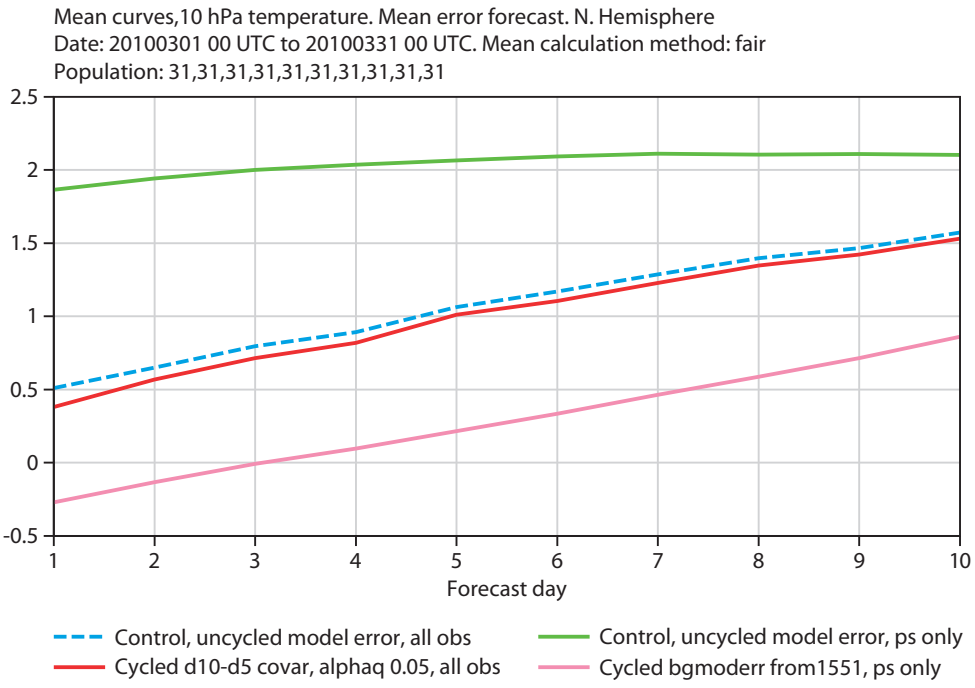

Figure 17: Mean forecast error averaged over 31 cases for temperature at 10 hPa, Northern hemisphere, March 2010.

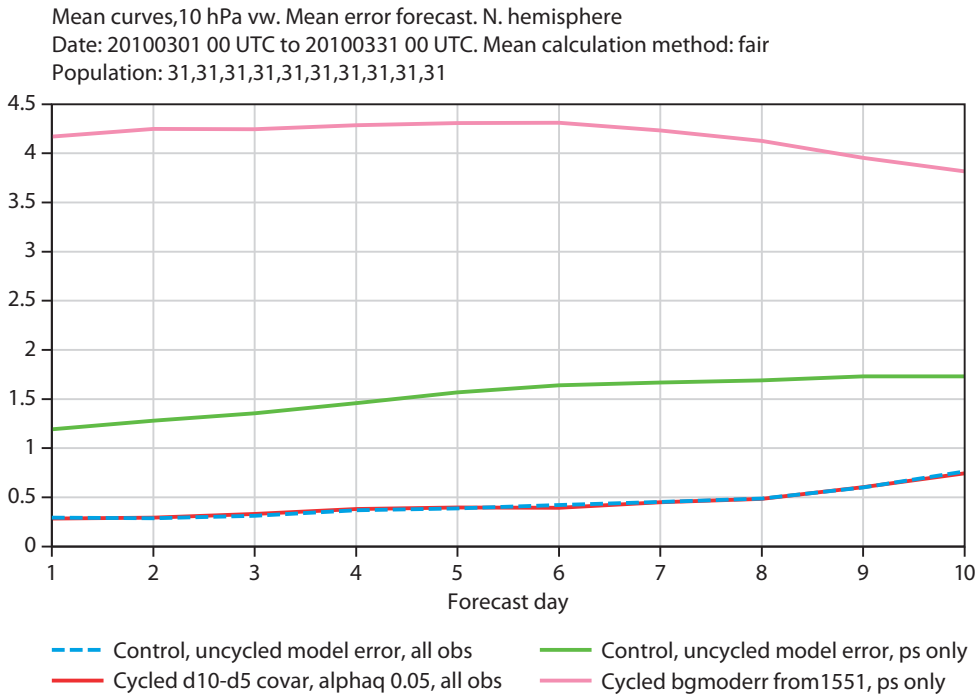

Figure 18: Mean forecast error averaged over 31 cases for vector wind at 10 hPa, Northern hemisphere, March 2010.

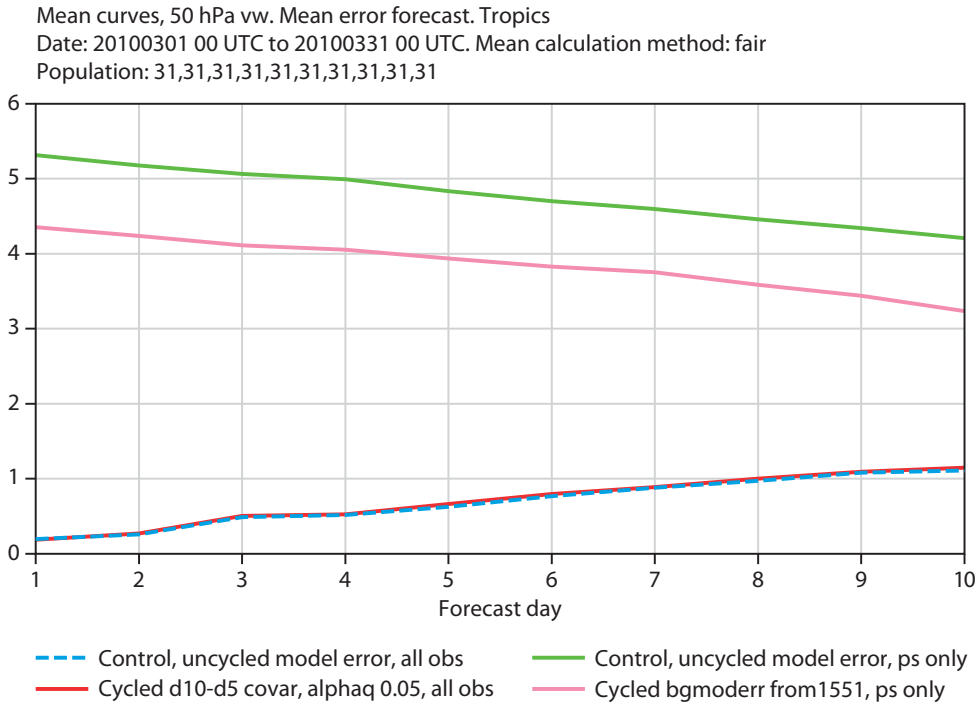

Figure 19: Mean forecast error averaged over 31 cases for vector wind at 50 hPa, Tropics, March 2010.

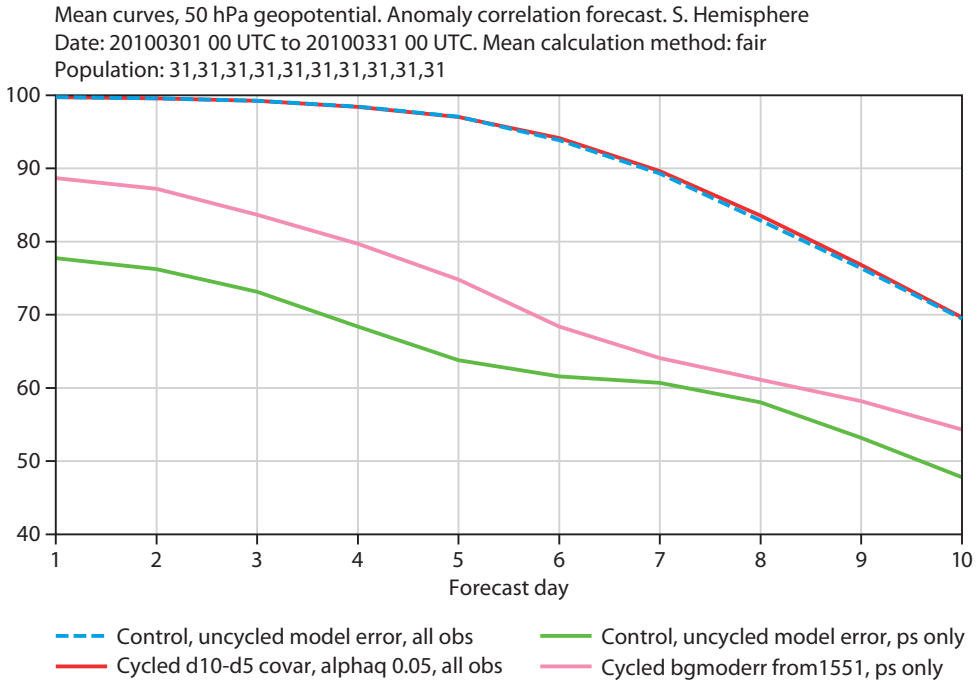

Figure 20: Anomaly correlation averaged over 31 cases for 50 hPa geopotential height, Southern hemisphere, March 2010.

40 hPa. (In other experiments not described here, we found a promising level of consistency when the model error estimation is extended to include the upper troposphere down as far as model level 60, i.e. around 400 hPa.) The dominant features are a cooling tendency (shaded blue) throughout much of the stratosphere, punctuated by warming tendencies at both Poles between 0.5 and 2 hPa. A broadly similar pattern of cooling is evident in the Day-5 minus Day-10 forecast differences taken from the Full-obs Control experiment (figure 16b) but the polar warming tendencies are not as pronounced. Taking into account forecast differences for different lead times (Day-1 minus Day-6), the relative magnitudes shown in figures 16a and 16b are broadly consistent with an exponential relaxation of model bias towards the climatology of the model with an *e*-folding time of around 10 days. The Day-1 minus Day-6 forecast differences (figures not shown) also show polar warming tendencies more in keeping with the cycled model error estimate (i.e. figure 16a), so it is conceivable that the less-pronounced polar warming tendencies in the Day-5 minus Day-10 differences are due to saturation of the model error growth.

Overall the correspondence between figures 16a and 16b has recognisable imperfections but is nonetheless encouraging and suggests that further research is warranted. For example, we expect this type of model error estimation would benefit from further tuning of the Q matrix, including consideration of flow-dependency, and this could perhaps be guided by further diagnostics derived from the assimilation system (e.g. systematic contributions to departure statistics and analysis increments) coupled with insight into the formulation of the model physics. It is worth noting that for pragmatic reasons our Q matrix estimate was derived from IFS cycles CY35R2-CY36R1 (namely those in Operations between June 2009 and May 2010), and hence may not be well-tuned for our weak-constraint experiments (CY36R4).

A comparison of Figures 8b, 14b and 16a reveals that the dominant spatial scales are commensurate (consistent with the use of a common model error covariance matrix). We attribute other differences to a combination of factors including differences in the IFS cycles used for the experiments, and to monthly variations in model error (i.e. meteorological flow-dependence).

Taking the Full-obs Cycling experiment as our best available estimate of model error, we then proceeded to use this estimate in a Ps-only experiment. Compared to the Ps-only control, we look for signs that the use of model error is able to make an impact on systematic aspects of forecast errors (verified against ECMWF Operations). Although we found statistically-significant beneficial impact for a number of parameters, we also found detrimental impact in others, as shown in figures 17 through 20.

To illustrate impact in the Northern hemisphere, figure 17 shows mean errors for temperature at 10 hPa. The Full-obs experiments show typical behaviour in that errors increase with lead time, but note that model-error cycling gives reduced error. The Ps-only control (without model error cycling) is the poorest with a rather constant mean error at all lead times, indicating that surface pressure observations alone are not able to counteract a systematic model error in the upper stratosphere. The Ps-only experiment with model error cycling has the lowest mean error at all lead times. In contrast, figure 18 shows the analogous plot for vector wind and here the impact of model error cycling in the Ps-only experiment degrades the mean error. However, the corresponding forecast error standard deviation, which is dominant, is reduced for Ps-only with cycling (figure not shown), suggesting that the additional degrees of freedom introduced by weak-constraint 4D-Var have actually corrected random errors rather than systematic errors in 10 hPa wind. This was not the intention of the model error representation and highlights the challenge of correctly apportioning discrepancies between observations and model background into contributions from random observation error, random model error, systematic observation error and systematic model error. We note that improvements here need not come solely from terms directly related to the model error specification, e.g. the Q matrix, but could also come from re-tuning of the assimilation scheme's specification of random errors at this altitude.

Shifting attention to the Tropics, figure 19 shows mean errors for vector wind at 50 hPa. For this parameter, the Ps-only experiments have larger mean errors than the Full-obs experiments but again Ps-only with model error cycling improves on Ps-only without. Although we have no convincing explanation for why the Ps-only experiments show decreasing error at longer lead times, we note that 50 hPa is below the region where the model error estimate is explicitly applied, and so impacts here are probably related to the interaction between a modified background and the tails of the analysis increments induced by surface pressure observations.

Shifting now to the Southern hemisphere, the impact of model error cycling on anomaly correlation for geopotential height appears to be neutral for the troposphere and beneficial for the stratosphere. The benefit is more apparent at 50 hPa than at 10 hPa, figure 20. Note, however, that these conclusions must be regarded as tentative as they are based on a sample of just 31 cases.

6.4 Discussion

In summary, there are encouraging signs that weak-constraint 4D-Var is able to produce meaningful estimates of systematic model error given a modern observing system, and that such estimates can compensate for a much-reduced observing system which is typically encountered in early decades of century-long reanalyses. Through weak-constraint 4D-Var, improvements were found in a range of forecast scores, but we recommend further research to widen the extent of the improvement and enhance the robustness of the results.

Estimation of the covariance matrix for systematic model error (Q) is a relatively unexplored area. Based on experience with covariance estimation for short-range background error (B), we expect substantial gains to be possible with further tuning and representation of flow-dependence in Q. We see this as a promising direction for improving performance in strongly perturbed situations, e.g. sudden stratospheric

warmings.

Experimentation with the surface-pressure-only reanalysis system described here provides an excellent framework for evaluating the performance of weak-constraint 4D-Var. The Ps-only assimilations contain a wealth of other diagnostic information including departure statistics from withheld passive observations (satellite radiances, GPSRO, etc.), offering more opportunities to identify strengths and weaknesses of the weak-constraint approach.

7 Impact of Analysis Window Configuration in Surface-Pressure-Only Reanalysis Experiments

Following the experiments described in the preceding section investigating the impact of cycling model error in a surface-pressure-only reanalysis configuration, we conducted a set of experiments to investigate the impact of different analysis window configurations.

Reanalysis differs from numerical weather prediction in that it has access to observations in the future with respect to the analysis time. This additional wealth of observations has always been proposed to represent a unique asset for reanalysis. Yet, its potential has so far remained untapped, because algorithms to exploit this information in an optimal manner were not practically available. However, by extending the 4D-Var analysis window, so as to include both past and future observations beyond the 6- or 12-hour norm of NWP, it is possible to start experimenting with the idea of a "retrospective analysis" that differs fundementally from analysis for NWP.

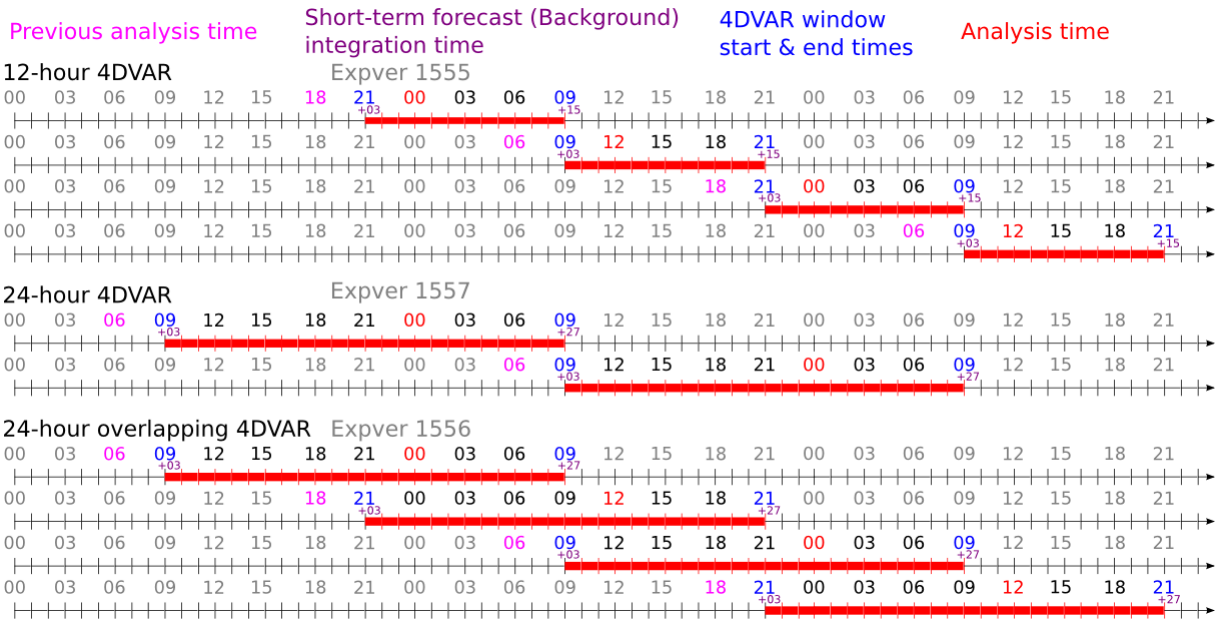

Figure 21: Schematic illustration of various analysis windows: 12-hour 4D-Var (top), non-overlapping 24-hour 4D-Var (middle), overlapping 4D-Var (bottom).

7.1 Analysis window configurations

Figure 21 shows in schematic form the various cycling methods explored here for reanalysis. Red bars indicate the analysis windows for successive cycles of 4D-Var. In all cases, analyses are produced every 6 hours at 00, 06, 12 and 18 UTC. However, it is convenient to define a "main analysis time", shown in hours UTC by a red number, useful to label the particular analysis cycle.

In all experiments presented here, the background for each analysis cycle is computed as a 3-hour forecast from the most recent analysis of the preceding cycle. The analysis which provides the initial state for the background forecast is indicated in hours UTC by a magenta number.

The first two configurations in figure 21 represent 12-hour cycling 4D-Var and 24-hour cycling 4D-Var.

The third configuration in figure 21 represents 24-hour 4D-Var overlapping-window, with 12-hour cycling. There are two obvious approaches for constructing such analyses, depending on the choice of the background state. The first approach enforces strict statistical correctness by requiring that the background should be taken from an analysis that has not made use of any of the observations that are to be assimilated in the current cycle. The background could, for example, be defined as a 3-hour forecast from the last-but-one analysis cycle.

The alternative approach is to use the background produced by the last analysis. This is the third configuration illustrated in figure 21. One can immediately see that this configuration violates one of the cornerstone assumptions in data assimilation, namely that background and observations should be independent from each other.

To justify taking the background state from the immediately-preceding analysis (rather than the statistically independent "older" background), let us consider what happens if we extend a 12-hour analysis window by moving the final time forward by 12 hours. The additional observations assimilated at the end of the window will clearly change the analysis later in the window. However, we may suppose that they have a relatively smaller impact towards the start of the window. In the extreme case of a very long window, the additional observations would have no impact on the analysis towards the beginning of the window. In this case, since the state towards the start of the window does not change from cycle to cycle, we are justified in removing the initial 12 hours of the current window. We can then replace the corresponding contributions to the cost function with a constraint that the state at the start of the window should not deviate far from the state determined by the previous analysis cycle. In other words, we can regard an analysis that takes its background from preceding analysis cycle as an approximation to an analysis with a longer window.

Clearly, a 24-hour window may not sufficiently long for this justification to be very convincing. However, we have found in practice that the analysis quality benefits from the more accurate background provided by the preceding analysis. This early practical result provides sufficient justification for taking a few liberties with statistical correctness.

Before discussing the results in detail, it is important to note that, in the overlapping 24-hour configuration, each observation set is analyzed twice: the first time as part of the second 12-hours of a 24-hour window, in conjunction with a background that is a 15-27 hour forecast; the second time as part of the first 12-hours of a 24-hour window, in conjunction with a background that is a 3-15 hour forecast. It is important to realise that this second forecast was in fact initialized from an analysis that was already influenced by observations also seen by the prior analysis. As such, that second forecast already contains some information about the set of observations in the first half of the window.

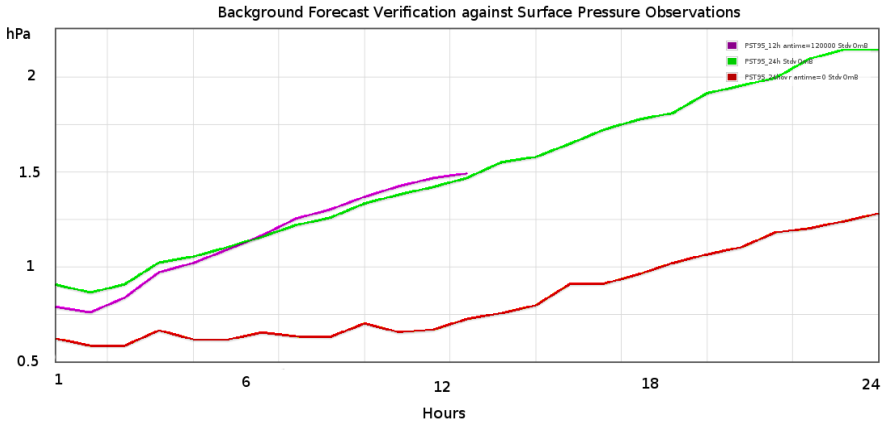

Figure 22: RMS forecast scores for surface pressure for experiments in which only surface pressure observations were assimilated. Blue: 12-hour 4D-Var. Green: 24-hour non-overlapping 4D-Var. Red: 24-hour overlapping 4D-Var in which the background state is taken from the immediately preceding analysis cycle.

7.2 Comparison to observations within the assimilation window

Each of the experiments considered in this section was run for several months, from 1 June 2004 until April 2005. Figure 22 shows RMS forecast errors with respect to surface pressure observations for the background forecasts for three experiments, as a function of background time step.

To guarantee a fair comparison to the approximate same set of observations between the various experiments, we only considered in our evaluation the observations assimilated between 9 UTC and 21 UTC. Looking at figure 21, it is easy to see that this is the only common set of observations seen by all three experiments and for which the background is a forecast at the same time-steps. Omitting this distinction, for example by considering all observations assimilated in a 24-hour interval, results in observations being counted twice in the overlapping analyses, and in comparing first-guess departures at different background time-steps.

In the first few hours (up to about 6 hours), the 12-hour 4D-Var background fits the observations better than the 24-hour 4D-Var background. This is expected because the analysis from which the 24-hour 4D-Var background was drawn fitted the observations better than the 12-hour experiment. However, the 12-hour 4D-Var background errors grow faster than those of the 24-hour 4D-Var and at some point the curve for 12-hour experiment crosses that of the 24-hour experiment: the 24-hour 4D-Var background fits observations better at forecast lead times 6-12 hours than the 12-hour 4D-Varbackground. One probable explanation is that fitting observations over a longer window (24 hours instead of 12 hours) may constrain the large-scale circulation in the analysisn more than the small-scale circulation (which has a shorter memory). This could also explain why the curves cross between 1 and 12 hours. We stress that this result is obtained at all times in the two experiments; we believe is a stable feature, and not a result of insufficient representativeness. However, using a dedicated background error covariance matrix computed specifically for the 24-hour window analysis scheme could result in a change in the relative weight given to small and large wave-number increments.

The comparison with 24-hour overlapping 4D-Var at background time-steps of 1-12 hours shows, as expected, much reduced departures compared with the other two experiments. This is because, in the overlapping configuration, information from the observations in that part of the window was already extracted in the earlier analysis, and is thus contained in the background. In other words, the background

departures shown here are more like analysis departures from the first of the two overlapping analyses. However, for background time-steps of 12-24 hour, the overlapping 24-hour 4D-Var background fits the observations far better than the other two experiments do for background time-steps of 1-12 hour. This would seem to suggest a clear superioriy of the overlapping configuration over the other two configurations.

7.3 Forecast scores

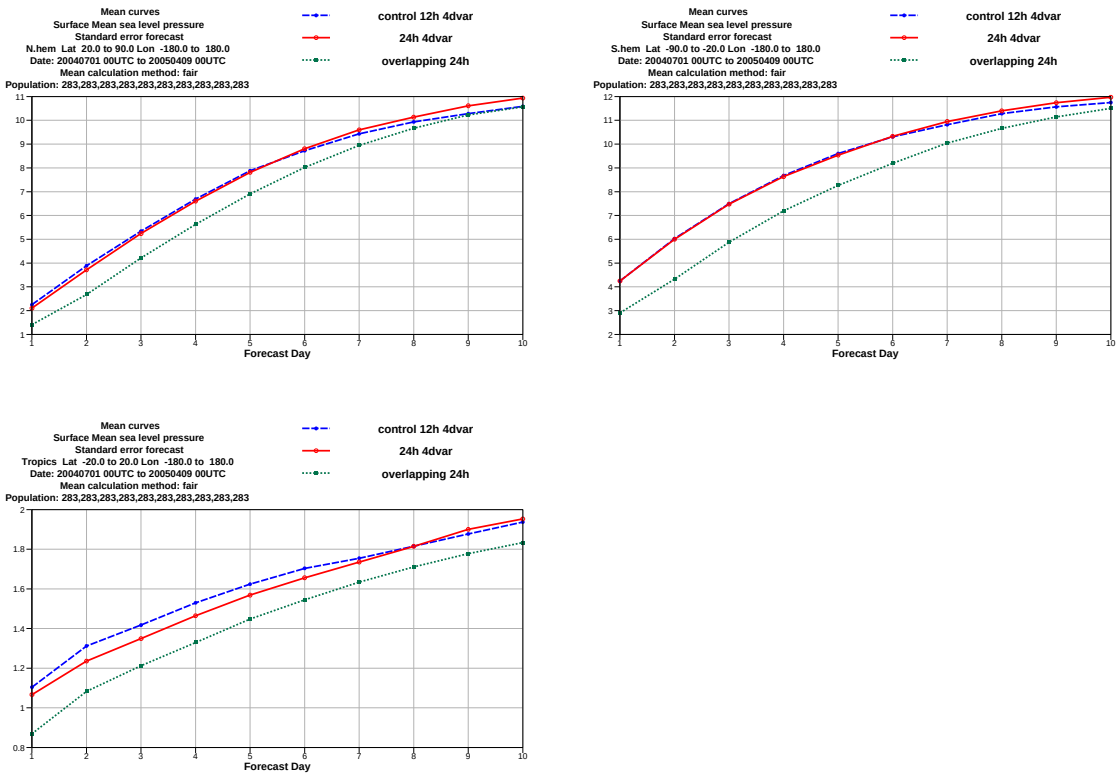

Figure 23: RMS Forecast scores verified against the ECMWF operational analysis for 12-hour, 24-hour non-overlapping and 24-hour overlapping analyses. Note that the 24-hour forecasts benefit from the assimilation of an additional 12-hours of observations.

Figure 23 shows forecast scores from the 12-hour, 24-hour non-overlapping and 24-hour overlapping experiments. The 24-hour overlapping configuration issued two sets of 10-day forecasts from 00 UTC. The first set of forecasts comes from the first analysis, when 00 UTC is towards the end of the 24-hour analysis window and benefits mostly from past observations. The second set of forecasts comes from the second analysis, when 00 UTC is at the beginning of the 24-hour analysis window and benefits from both past and future observations. Unfortunately, when these experiments were conducted, the archiving system could not discriminate between two forecasts issued from the same analysis time, and the first set of forecast fields were overwritten by later forecasts. As a consequence, for the 24-hour overlapping configuration, the initial states for the forecasts are 21 hours before the end of the analysis window, whereas for the non-overlapping 24-hour and 12-hour systems, the initial time of the forecast is 9 hours before the end of the window. The forecasts for the 24-hour overlapping system therefore benefit from an additional 12 hours of observations. This advantage clearly contributes to the much better forecast

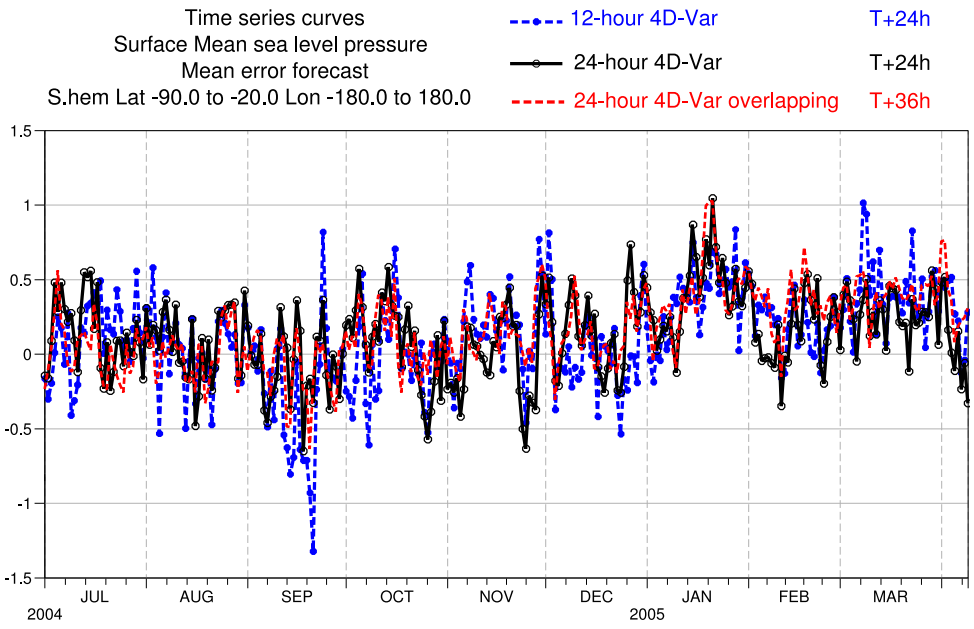

Figure 24: Mean foreast error for mean sea-level pressure in the southern hemisphere extra-tropics at 24-hour for the 12-hour and 24-hour 4D-Var experiments, and at 36-hour for the overlapping 24-hour 4D-Var experiment.

scores of the 24-hour overlapping system.

However, even if this 12-hour advantage is taken into account, the 24-hour overlapping configuration is significantly better than the others. For example, the advantage is more than 24-hours in the southern hemisphere and Tropics. Furthermore, we note that the unfairness of the comparison only applies in the NWP context. In the reanalysis system, the ability to take into account more observations in the future (relative to the analysis time) is a significant advantage. The results suggest that, for reanalysis purposes, the analysis should be taken from a time towards the beginning of the window.

Another metric to consider is the "forecast jumpiness". Figure 24 shows the mean sea level pressure mean forecast error in the Southern Hemisphere extratropics, at 24-hour for the 12-hour and 24-hour 4D-Var experiments, and at 36-hour for the overlapping 24-hour 4D-Var experiment (to offset the fact that this latter experiment has seen 12 hours worth more of observations). The red curve appears visually smoother, suggesting less forecast jumpiness for overlapping window configuration. In an operational NWP context, this could represent significant benefits for forecast users.

One important caveat in all the results presented here is that the background error covariance matrix was not adjusted, but taken as used today in Operations. Ideally one should have recomputed of J_b for each configuration.

8 Parallelisation of 4D-Var

There is a pressing need to find new levels of parallelism in 4D-Var. The relatively low spatial resolution used in the inner minimisation means that the model can be spread efficiently over only a relatively small number of processors. Furthermore, many of the computations are inherently sequential: iterations of the minimisation algorithm proceed sequentially, as do the timesteps of the model integrations.

Fortunately, weak-constraint 4D-Var opens opportunities for parallelisation that are not available in strong-constraint 4D-Var. One such opportunity occurs at the outer level of the incremental algorithm. Here, the state $\mathbf{x}^{(n)}$ is updated by the addition of an increment, $\delta\mathbf{x}^{(n)}$ determined by the inner minimisation. It is then necessary to integrate the model over each of the sub-windows $[t_k, t_{k+1})$ $(k = 0, \ldots, N-1)$, to allow calculation of the model errors and to compare the model with observations that are distributed throughout the sub-windows. Clearly, since the updated four-dimensional state contains the initial conditions for each of these sub-window integrations, the integrations may be performed in parallel.

The possibility to parallelise over sub-windows at the outer level of the incremental algorithm is welcome. However, it is arguably more important to parallelise the inner minimisation, where spatial resolutions are lower and where most of the computational resources are spent. This is considerably more challenging.

8.1 Parallel Minimisation Algorithms

One obvious idea is to parallelise the iterations of the minimisation algorithm, by simultaneously calculating more than one cost function gradient at each iteration. We believe this is unlikely to prove beneficial.

The inner cost function is quadratic, so that the equation for the gradient is linear. Minimisation is therefore equivalent to solving the linear equation $\nabla J = 0$. Using linear solution methods for the gradient equation is significantly more efficient than applying general-purpose minimisation methods to the quadratic cost function.

Modern iterative solution methods for linear equations (e.g. conjugate gradients) are based on Krylov methods. Consider a linear system:

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{39}$$

A Krylov method seeks an approximate solution which, after K iterations, is a linear combination of the vectors:

$$\mathbf{b}, \quad \mathbf{A}\mathbf{b}, \quad \mathbf{A}^2\mathbf{b}, \quad \ldots \quad \mathbf{A}^K\mathbf{b} \tag{40}$$

Note that the vectors that define the Krylov space are constructed from the right hand side vector, $\mathbf{b}$. This choice is important. A Krylov space constructed by repeated application of $\mathbf{A}$ to a different initial vector will not, in general, contain an accurate solution of the problem $\mathbf{A}\mathbf{x} = \mathbf{b}$.

The reason for this may be partially understood by noting that, as a consequence of the Cayley-Hamilton theorem, the inverse of $\mathbf{A}$ may be expressed as a polynomial in $\mathbf{A}$:

$$\mathbf{A}^{-1} = \sum_{n=0}^{N} a_n \mathbf{A}^n. \tag{41}$$

The solution to the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ can therefore be written as:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \sum_{n=0}^{N} a_n \mathbf{A}^n \mathbf{b}. \tag{42}$$

That is, the solution is a linear combination of the vectors $\mathbf{b}$, $\mathbf{A}\mathbf{b}$, $\mathbf{A}^2\mathbf{b}$ etc. that constitute the Krylov space generated by $\mathbf{b}$. This space has a special significance for solution of the problem $\mathbf{A}\mathbf{x} = \mathbf{b}$ that is not shared by the Krylov space generated by an arbitrary starting vector.

The most obvious way in which the inner loop of 4D-Var could be parallelised is to calculate more than one gradient at each iteration. This is equivalent to choosing a set of initial vectors $\mathbf{b}_1 \ldots \mathbf{b}_m$ and generating the space:

$$\mathbf{b}_1, \ldots, \mathbf{b}_m, \quad \mathbf{A}\mathbf{b}_1, \ldots, \mathbf{A}\mathbf{b}_m, \quad \ldots \quad \mathbf{A}^K\mathbf{b}_1, \ldots, \mathbf{A}^K\mathbf{b}_m, \tag{43}$$

It is clear that such an approach might be appropriate if the aim were to solve a set of linear equations with different right hand sides, $\mathbf{b}_1 \ldots \mathbf{b}_m$. However, the spaces generated by the vectors $\mathbf{b}_2 \ldots \mathbf{b}_m$ contribute little to the solution of the problem for $\mathbf{b}_1$, so that this type of block Krylov approach does not significantly reduce the number of iterations (and hence the number of sequential applications of $\mathbf{A}$) that are required to produce a solution. (It is worth noting that block Krylov methods can produce a small reduction in the number of iterations if the extreme eigenvalues of $\mathbf{A}$ are clustered. However, this is not the case in 4D-Var at least with the usual preconditioning. In any case, the number of iterations saved is unlikely to warrant the extra computational cost involved.)

8.2 Parallelisation Within An Iteration

Since the iterations of the inner loop cannot usefully be parallelised, it makes sense to try to parallelise the computations within a single iteration.

Let us consider the inner cost function, equation 25. For convenience of notation, we will drop the superscript (n). We will also introduce the following matrices and vectors:

$$\mathbf{R} = \begin{pmatrix} R_0 & & & \\ & R_1 & & \\ & & \ddots & \\ & & & R_{N-1} \end{pmatrix}, \quad \mathbf{D} = \begin{pmatrix} B & & & \\ & Q_1 & & \\ & & \ddots & \\ & & & Q_{N-1} \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} H_0 & & & \\ & H_1 & & \\ & & \ddots & \\ & & & H_{N-1} \end{pmatrix},$$

$$\mathbf{b} = \begin{pmatrix} b \\ c_1 \\ \vdots \\ c_{N-1} \end{pmatrix} \quad \text{and} \quad \mathbf{d} = \begin{pmatrix} d_0 \\ d_1 \\ \vdots \\ d_{N-1} \end{pmatrix}. \tag{44}$$

With these definitions, and using the operator $\mathbf{L}$ defined in equation 27, we can write the quadratic cost function (equation 25) as:

$$J(\delta\mathbf{x}) = (\mathbf{L}\delta\mathbf{x} - \mathbf{b})^{\mathrm{T}}\mathbf{D}^{-1}(\mathbf{L}\delta\mathbf{x} - \mathbf{b}) + (\mathbf{H}\delta\mathbf{x} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x} - \mathbf{d}). \tag{45}$$

An alternative expression for the cost function can be derived by substituting $\delta\mathbf{p} = \mathbf{L}\delta\mathbf{x}$ into equation 45:

$$J(\delta\mathbf{p}) = (\delta\mathbf{p} - \mathbf{b})^{\mathrm{T}}\mathbf{D}^{-1}(\delta\mathbf{p} - \mathbf{b}) + (\mathbf{H}\mathbf{L}^{-1}\delta\mathbf{p} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\mathbf{L}^{-1}\delta\mathbf{p} - \mathbf{d}) \tag{46}$$

In this case, the operator $\mathbf{L}^{-1}$ appears in the second term of the cost function, where it is needed to transform $\delta\mathbf{p}$ to $\delta\mathbf{x}$ prior to application of the observation operators. In both versions of the cost function, the first term represents the sum of J_b and J_q, whereas the second term represents J_o.

The usual method of preconditioning the minimisation in 4D-Var employs a change of variable that reduces the first term of the cost function to the identity matrix. In the second version of the cost function (equation 46), this change of variable is defined by

$$\delta\mathbf{p} = \mathbf{D}^{1/2}\delta\chi \tag{47}$$

For the first version of the cost function (equation 45), the change of variable is

$$\delta\mathbf{x} = \mathbf{L}^{-1}\mathbf{D}^{1/2}\delta\chi \tag{48}$$

The two approaches to preconditioning are clearly equivalent. They lead to the same solution and any choice of minimisation algorithm will have identical convergence properties in the two cases.

In both cases, the Hessian matrix of the cost function (with respect to $\delta\chi$) is:

$$J''_{\delta\chi} = \mathbf{I} + \mathbf{D}^{\mathrm{T}/2}\mathbf{L}^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{L}^{-1}\mathbf{D}^{1/2} \tag{49}$$

The presence of the identity matrix in equation 49 ensures that the eigenvalues of the Hessian are bounded away from zero, ensuring a well-conditioned system. Except for the presence of $\mathbf{L}^{-1}$ in equation 49 and the fact that the matrices are four-dimensional, the Hessian resembles that of strong-constraint 4D-Var. We expect, and have found in practice, that the convergence properties of the minimisation in weak-constraint 4D-Var are similar to those of strong-constraint 4D-Var (at least for moderate window lengths).

Note that the operator $\mathbf{L}^{-1}$ appears in the cost function in equation 46. Since application of $\mathbf{L}^{-1}$ requires sequential integration of the linear model over the entire analysis window, the problem is inherently sequential. There is no possibility to parallelise over sub-windows.

By contrast, 45 contains $\mathbf{L}$, rather than its inverse. The model integrations required to evaluate the cost function (and its gradient) may be performed in parallel. With this formulation, $\mathbf{L}^{-1}$ appears in the preconditioner, which is therefore sequential. However, we are free to modify the preconditioner, without altering the problem, provided we do not destroy the numerical conditioning. There is hope, therefore, that we may be able to replace $\mathbf{L}^{-1}$ in equation 48 by an operator that is cheaper or more parallel.

The most obvious approximation is

$$\hat{\mathbf{L}} = \begin{pmatrix} I & & & & \\ -I & I & & & \\ & -I & I & & \\ & & \ddots & \ddots & \\ & & & -I & I \end{pmatrix}. \tag{50}$$

This seems reasonable, since the model integrations in $\mathbf{L}$ are for relatively short periods of time (a few hours) during which the state does not change much.

Unfortunately, this approximation does not provide effective preconditioning. In tests using the simple two-layer quasi-geostrophic system described in section 4, the condition number of the Hessian matrix increased from around 10^4 to 10^9 as a consequence of the approximation, and the minimisation failed to converge. A number of alternative approximations to $\mathbf{L}$ have been tried, but we have so far not found a parallelisable preconditioner for equation 45 that results in acceptable convergence properties.

It would appear from our preliminary investigations that minimisation of equation 45 is highly sensitive to the choice of preconditioner. We do not yet fully understand why this is the case. However, it may be a consequence of the large condition number of the covariance matrix, $\mathbf{D}$, which assigns very small variances to the smallest spatial scales. Using $\hat{\mathbf{L}}$ to define the change of variable transforms the Hessian matrix of the first term of the cost function from the identity matrix to:

$$\mathbf{D}^{\mathrm{T}/2}\hat{\mathbf{L}}^{-\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{L}\hat{\mathbf{L}}^{-1}\mathbf{D}^{1/2} \tag{51}$$

The effectiveness of the preconditioning relies on a cancellation between $\mathbf{D}^{-1}$, the initial $\mathbf{D}^{\mathrm{T}/2}$ and the final $\mathbf{D}^{1/2}$. We suspect that the large condition number of $\mathbf{D}$ makes this cancellation highly sensitive to the accuracy with which $\hat{\mathbf{L}}^{-1}$ approximates $\mathbf{L}^{-1}$. For this reason, we have concentrated on developing algorithms that do not rely on this cancellation. The most promising approach is described in the next section.

8.3 Saddle Point Formulation of Weak-Constraint 4D-Var

Saddle point problems have received a lot of attention in the optimisation community in recent years. In part, this is because a very wide range of problems can be expressed in saddle point form. This has allowed a unification of solution and preconditioning methods across a broad range of fields. There is a large and growing literature on their numerical solution (see Benzi, Golub and Liesen, 2005, and references therein).

It is informative to cast the 4D-Var minimisation problem in saddle point form. To do this, let us consider equation 45. Setting the gradient of the cost function to zero, we have:

$$\nabla J = \mathbf{L}^{\mathrm{T}}\mathbf{D}^{-1}(\mathbf{L}\delta\mathbf{x}-\mathbf{b})+\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x}-\mathbf{d})=\mathbf{0}. \tag{52}$$

Let us now define

$$\lambda = \mathbf{D}^{-1}(\mathbf{b}-\mathbf{L}\delta\mathbf{x}) \tag{53}$$
$$\mu = \mathbf{R}^{-1}(\mathbf{d}-\mathbf{H}\delta\mathbf{x}) \tag{54}$$

which we can rearrange to:

$$\mathbf{D}\lambda+\mathbf{L}\delta\mathbf{x} = \mathbf{b} \tag{55}$$
$$\mathbf{R}\mu+\mathbf{H}\delta\mathbf{x} = \mathbf{d} \tag{56}$$

while from equation 52, we have:

$$\mathbf{L}^{\mathrm{T}}\lambda+\mathbf{H}^{\mathrm{T}}\mu=\mathbf{0} \tag{57}$$

Writing equations 55, 56 and 57 as a single matrix equation, we arrive at the saddle-point formulation:

$$\begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{L} \\ \mathbf{0} & \mathbf{R} & \mathbf{H} \\ \mathbf{L}^{\mathrm{T}} & \mathbf{H}^{\mathrm{T}} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \\ \delta\mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{d} \\ \mathbf{0} \end{pmatrix} \tag{58}$$

We will denote the block 3×3 matrix in equation 58 by $\mathcal{A}$. The matrix is symmetric and indefinite.

The saddle point formulation of the 4D-Var is particularly elegant. Each of the matrices involved in the cost function appears as a separate block of $\mathcal{A}$. (We note in passing that strong-constraint 4D-Var can also be expressed in this form, in which case the $(1,1)$ block becomes B, and $\mathbf{L}$ is replaced by the identity matrix.)

None of the blocks of $\mathcal{A}$ contains an inverse matrix. In particular, since the inverse of $\mathbf{L}$ does not appear, the application of $\mathcal{A}$ can be parallelised. An additional advantage of this formulation is that it does not require the inverse of $\mathbf{R}$. This could make accounting for correlated observation error considerably easier than it is in the usual 4D-Var formulation.

The matrix $\mathcal{A}$ can be expected to have a large condition number. Thus, preconditioning is required for efficient solution. A range of preconditioning methods for saddle point systems is discussed by Benzi and Wathen (2008). However, most preconditioners described in the literature require application of $\mathbf{L}^{-1}$.

One promising approach that avoids the need to apply $\mathbf{L}^{-1}$ is the so-called Inexact Constraint Preconditioner (Bergamaschi, *et al.*, 2011). In this approach, the preconditioner is constructed by replacing the sub-matrices $\mathbf{L}$ and $\mathbf{H}$ in $\mathcal{A}$ by approximations that allow the inverse matrix to be applied at low computational cost. A particularly attractive approximation from this point of view is to replace $\mathbf{L}$ by an easily inverted matrix $\tilde{\mathbf{L}}$, and to replace $\mathbf{H}$ by a zero matrix;

$$\tilde{\mathcal{P}} = \begin{pmatrix} \mathbf{D} & \mathbf{0} & \tilde{\mathbf{L}} \\ \mathbf{0} & \mathbf{R} & \mathbf{0} \\ \tilde{\mathbf{L}}^{\mathrm{T}} & \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{59}$$

It is readily shown that the inverse of $\tilde{\mathcal{P}}$ is:

$$\tilde{\mathcal{P}}^{-1} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \tilde{\mathbf{L}}^{-\mathrm{T}} \\ \mathbf{0} & \mathbf{R}^{-1} & \mathbf{0} \\ \tilde{\mathbf{L}}^{-1} & \mathbf{0} & -\tilde{\mathbf{L}}^{-1}\mathbf{D}\tilde{\mathbf{L}}^{-\mathrm{T}} \end{pmatrix} \tag{60}$$

Note that $\tilde{\mathcal{P}}^{-1}$ does not contain $\mathbf{D}^{-1}$. Thus, we avoid dividing by tiny variances, and the preconditioning does not rely on a cancellation between $\mathbf{D}^{-1}$ and $\mathbf{D}$

Bergamaschi, *et al.*, (*op. cit.*) show (for a more general approximation) that the preconditioned saddle point matrix $\tilde{\mathcal{P}}^{-1}\mathcal{A}$ can be expected to have a large number of eigenvalues τ exactly equal to one, with the remaining eigenvalues contained within a disc centred on the point $\tau = 1$ in the complex plane. We prove in appendix A, and have verified experimentally, that for the case $\tilde{\mathbf{L}} = \mathbf{L}$, the eigenvalues of the preconditioned system all lie on the line $\Re(\tau) = 1$, and are therefore bounded away from zero. For $\tilde{\mathbf{L}} \neq \mathbf{L}$, our experiments show that many eigenvalues also lie on this line, with others lying in a cloud centred on the point $\tau = 1$.

The efficacy of the preconditioner can be expected to depend on the accuracy with which $\tilde{\mathbf{L}}$ approximates $\mathbf{L}$. We found little success with the approximation suggested in equation 50, and therefore seek alternatives. Let us define

$$\mathbf{M} = \mathbf{I} - \mathbf{L} = \begin{pmatrix} 0 & & & & \\ M_1 & 0 & & & \\ & M_2 & 0 & & \\ & & \ddots & \ddots & \\ & & & M_{N-1} & 0 \end{pmatrix} \tag{61}$$

Now, it is easy to show that

$$\mathbf{M}^2 = \begin{pmatrix} 0 & & & & & \\ 0 & 0 & & & & \\ M_2M_1 & 0 & 0 & & & \\ & M_3M_2 & 0 & 0 & & \\ & & \ddots & \ddots & \ddots & \\ & & & M_{N-1}M_{N-2} & 0 & 0 \end{pmatrix} \tag{62}$$

Note that, whereas the non-zero blocks of $\mathbf{M}$ are along the first sub-diagonal and represent tangent-linear model integrations over one sub-window, the non-zero blocks of $\mathbf{M}^2$ are along the second sub-diagonal, and represent tangent-linear model integrations over two sub-windows. More generally, it can be shown that for $n > 0$, the non-zero blocks of $\mathbf{M}^n$ lie along the n^{th} sub-diagonal, and represent tangent-linear integrations of length n sub-windows. Moreover, for $n \geq N$, where N is equal to the total number of sub-windows, the non-zero blocks are effectively pushed off the edge of the matrix, so that $\mathbf{M}^N = \mathbf{0}$. (Mathematically, this shows that $\mathbf{M}$ is nilpotent with degree N.)

Now consider the following expression:

$$\left(\mathbf{I}+\mathbf{M}+\mathbf{M}^2+\ldots+\mathbf{M}^{N-1}\right)(\mathbf{I}-\mathbf{M}) = \mathbf{I}-\mathbf{M}^N = \mathbf{I} \tag{63}$$

The matrix $(\mathbf{I}-\mathbf{M})$ in the left hand side of this equation is simply $\mathbf{L}$, so that $(\mathbf{I}+\mathbf{M}+\mathbf{M}^2+\ldots+\mathbf{M}^{N-1})$ must be equal to $\mathbf{L}^{-1}$. Writing this in terms of $\mathbf{L}$ rather than $\mathbf{M}$, we have:

$$\mathbf{L}^{-1} = \mathbf{I}+(\mathbf{I}-\mathbf{L})+(\mathbf{I}-\mathbf{L})^2+\ldots+(\mathbf{I}-\mathbf{L})^{N-1} \tag{64}$$

Clearly, application of the right hand side of equation 64 to a vector is expensive. However, if we recognise that equation 64 is a Neumann series representation of the inverse of $\mathbf{L}$, then it is natural to consider approximating $\mathbf{L}^{-1}$ by truncating this series (although, it is not clear that the terms in the series decay). An alternative, which we have yet to try, is to invert $\mathbf{L}$ approximately using a few iterations of a Krylov method. Equation 64 shows that such a method would converge in at most N iterations. We also note, in passing, that the matrix $\mathbf{L}$ plays an important role in the "Parareal" time-parallel integration algorithm (Lions *et al.*, 2001. See also Gander and Vandewalle, 2007). It is possible that exploiting this connection could lead to efficient and highly parallel preconditioners for weak constraint 4D-Var.

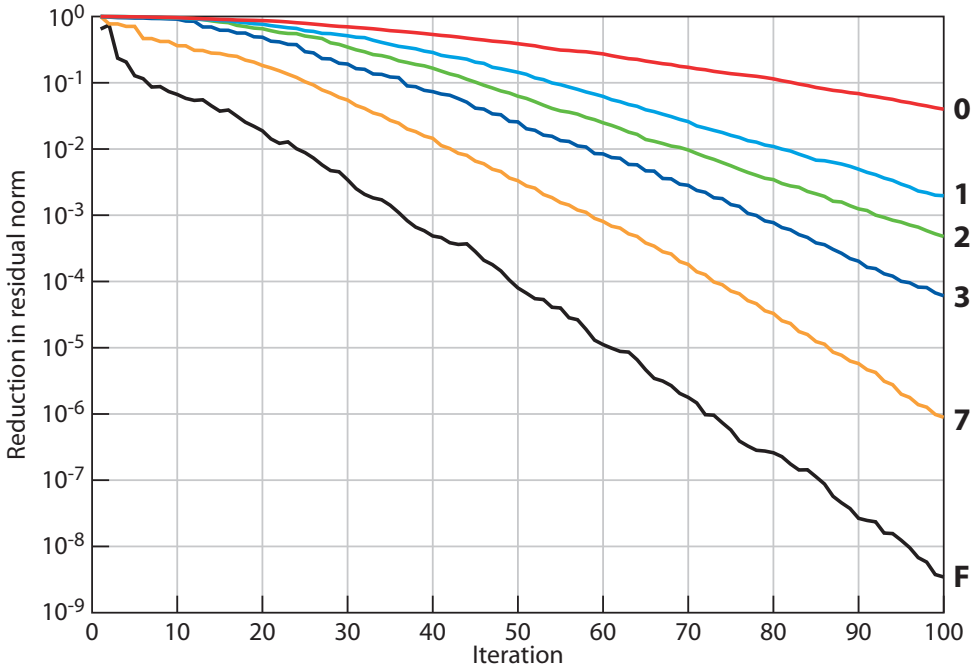

Figure 25: Convergence of the saddle point formulation of weak constraint 4D-Var with different truncations of the series approximation of $\mathbf{L}^{-1}$. The lines are labelled with numbers indicating the truncation order (largest power of $\mathbf{L}$). The convergence of the standard formulation (equations 46, 47 and 49) is shown by the line labelled "F".

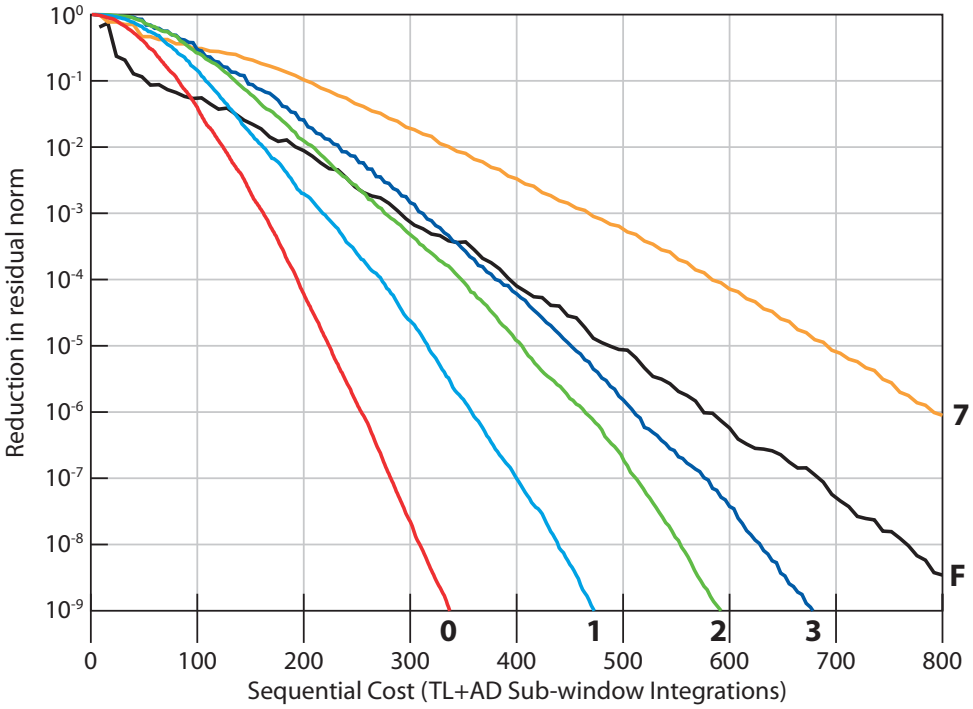

Figure 26: Convergence of the saddle point formulation of weak constraint 4D-Var as a function of sequential cost.

Figure 25 illustrates the convergence of the saddle point algorithm using the preconditioner described in this section for various truncations of the series expansion of $\mathbf{L}^{-1}$ defined in equation 64. The analysis system in this case is a two-level quasi-geostrophic model described in section 4. The analysis window is of length 24 hours with 8 sub-windows of length 3 hours, and the solution algorithm for the saddle point system was GMRES.

In general, higher order truncation of the series results in improved preconditioning, although truncations at order 4 and 5 were found to produce slightly slower convergence in this case than truncation at order 3, and truncation at order 6 produced slower convergence than zeroth-order truncation. Note also that the approximation of $\mathbf{L}^{-1}$ is exact for truncation order 7. No further improvement can be achieved by truncating the series at higher orders.

Despite these encouraging results, it should be noted that the sequential (wallclock) cost of each iteration increases as the truncation order is increased. A single application of $\mathbf{L}$ can be performed in the wallclock time required for one integration over the sub-window, if the sub-windows are processed in parallel. The total wallclock cost per iteration of the saddle point formulation is equivalent to $n+1$ applications of $\mathbf{L}$ and $n+1$ applications of $\mathbf{L}^{\mathrm{T}}$, where n is the truncation order used in the preconditioner. That is, the cost is equal to $n+1$ sub-window integrations of the tangent-linear model and $n+1$ sub-window integrations of the adjoint model. For comparison, the standard "forcing" formulation (equation 46) requires one integration of the tangent linear model and one of the adjoint over the entire analysis window (i.e. over $N = 8$ sub-windows).

To compare the sequential (wallclock) costs of the proposed algorithms, we re-plot figure 25 with the abscissa measured in sub-window integrations. That is, we have multiplied the abscissa for each curve by $n+1$, where n is the truncation order, or by 8 in the case of the curve labelled "F". This scaling allows a fair comparison of the curves in terms of sequential cost, under the assumption that the model integrations required to apply $\mathbf{L}$ are performed in parallel. The result is shown in figure 26. Note that

the saddle point formulation preconditioned using a zeroth-order approximation to $\mathbf{L}^{-1}$ (i.e. $\tilde{\mathbf{L}} = \mathbf{I}$ in equation 59) requires less than half the wallclock time of the "forcing" formulation.

It would appear from figure 26 that the improved conditioning that results from higher order approximation of $\mathbf{L}^{-1}$ is not sufficient to outweigh its computational cost. However, this result may depend on the number and length of the sub-windows.

Finally, we conducted an additional experiment in which the analysis window was increased in length from one to two days, and the number of sub-windows was doubled from 8 to 16. In this case (not shown) the number of iterations required to minimise the cost function remained approximately four times larger than for the corresponding "forcing" formulation. This would imply a wallclock cost for the saddle point formulation around four times smaller than for the "forcing" formulation. This preliminary result suggests that the saddle point formulation favours large numbers of short sub-windows, and that the conditioning of the problem does not deteriorate as the window length is increased. Experiments to evaluate the cost of the saddle point formulation for a range of window and sub-window lengths will be conducted shortly.

9 Summary

The results presented above represent a snapshot of the current status of the development of a weak-constraint, long-window data assimilation system at ECMWF. We are convinced that such a system will lie at the heart of a future operational data assimilation system, although it is clear that much work remains to be done.

Implementation of a weak-constraint 4D-Var at ECMWF has proceeded cautiously. Nevertheless, a weak-constraint 4D-Var system is now running operationally, and being tested for use in re-analysis. Properly accounting for systematic model error is a vital counterpart to observation bias correction.

Experiments with an analysis window extended to 24 hours now show an essentially neutral impact in a fully observed system. When a reduced observing system is considered (assimilating only surface pressure observations), use of longer (24-hour, rather than 12-hour) analysis window results in slower error growth during the first few hours of the forecast. Also, in the same reduced observing system context, an overlapping 24-hour analysis window, cycled every 12 hours, proves to be clearly superior to the other two schemes, leading to reduced forecast jumpiness. These results are encouraging and will allow a 24-hour analysis window to be implemented operationally next year. Improvements in forecast skill are expected as we relax the restriction of model error to the stratosphere, improve the estimation of mean model error, and take better account of the random component through improved covariance statistics.

The main scientific challenges we face revolve around the statistical description of model error. We are acutely aware that the covariance matrices we have used to date are poor representations of the true covariance of model error. Work is under way, in collaboration with R. Todling at GMAO to diagnose model errors from innovation statistics. This approach has the potential to provide objective information about model error. At the same time, we are working to incorporate model error more fully into the EDA (Ensemble Data Assimilation) system. We believe that it should be possible to use ensemble methods to diagnose model error in much the same way as they are currently used to determine covariances of background error. A crucial component of such an approach will be the availability of realistic stochastic physics and backscatter parameterisations.

In addition to improving the representation of model error covariance, further work is required to improve

our estimate of the mean component. The current approach appears to work well, but the time scales over which the mean error adjusts are determined by the covariance matrix for the random component of the error. This is wrong, and has resulted in the choice of covariance matrix being largely determined by its impact on the estimation of systematic error. It will be important to remove this confusion of purpose if we are to make progress in properly representing the covariance statistics of model error.

The technical challenges of weak-constraint and long-window 4D-Var should not be underestimated. There is a clear need to reduce the complexity and inflexibility of the current IFS code and scripts, if we are to continue to develop and test new assimilation algorithms, The work described in section 4 was conducted using the Object-Oriented Prediction System (OOPS). The OOPS project aims to develop a flexible data assimilation framework which allows new and existing algorithms to be expressed easily, and which allows these algorithms to be transferred directly from simplified systems with small models to the full numerical weather prediction system. The experience gained in implementing the algorithms described in section 8 is particularly encouraging in this respect. The algorithms were quickly and easily implemented in OOPS, and are therefore available for any model that is implemented in the OOPS framework. We consider that continued support for the OOPS project, and an effort to incorporate the IFS into OOPS is an important part of our strategy for further progress in weak-constraint and long-window 4D-Var at ECMWF.

So far, we have concentrated on addressing the systematic component of model error, since this is the major source of error in our system. A major milestone in our development of a weak-constraint 4D-Var system will occur when we also take proper account of the random, non-stationary component of model error. It is only at this stage that we will be able to exploit the full potential of a long-window analysis.

10 References

Bergamaschi L., J. Gondzio, M. Venturin and G. Zilli, 2011: Erratum to: Inexact constraint preconditioners for linear systems arising in interior point methods. *Computational Optimization and Applications*, **49**, 401-406

Benzi M., G. H. Golub, and J. Liesen, 2005: Numerical Solution of Saddle Point Problems, *Acta Numerical*, **14**, 1–137.

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

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

Fandry, C.B. and L.M. Leslie, 1984: A Two-Layer Quasi-Geostrophic Model of Summer Trough Formation in the Australian Subtropical Easterlies. *Journal of Atmos. Sci.*, **41**, pp807-817.

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

Gilmour I, L.A. Smith and R. Buizza, 2001: On the Duration of the Linear Regime: Is 24 hours a Long Time in Weather Forecasting? *Journal of Atmos. Sci.*, **58**, 3525–3539.

Klinker, E. and P. D. Sardeshmukh, 1992: The diagnosis of mechanical dissipation in the atmosphere from large-scale balance requirements. *Journal of Atmos. Sci.*, **49**, 608–627.

Lions J-L, Y. Maday and G. Turinici, 2001: A "parareal" in time discretization of PDE's. *C. R. Acad. Sci. Paris Sér. I Math.*, **332**, 661–668.

Li Z. and I. M. Navon, 2001: Optimality of 4D-Var and its relationship with the Kalman filter and Kalman smoother. *Q. J. R. Meteorol. Soc.*, **127** 661–684.

Lorenz, E.N., 1995: Predictability: a problem partly solved, ECMWF Seminar on Predictability, 4-8 September 1995, 1–18.

Martin J. and S. Vandewalle, 2007: Analysis of the parareal time-parallel time-integration method. *SIAM J. Sci. Comput.*, **29**, 556–578.

Ménard, R. and R. Daley, 1996: The application of Kalman smoother theory to the estimation of 4D-Var error statistics. *Tellus*, **48A**, 221–237.

Pedlosky, J., 1979: Geophysical Fluid Dynamics. *Springer-Verlag*.

Rodwell, M. J. and T. Jung, 2008: Understanding the local and global impacts of model physics changes: An aerosol example. *Q. J. R. Meteorol. Soc.*, **134** 1479–1497.

Rodwell, M. J. and T. N. Palmer, 2007: Using numerical weather prediction to assess climate models. *Q. J. R. Meteorol. Soc.*,**133**, 129–146.

Simmons A. and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Q. J. R. Meteorol. Soc.*,**128**, 647–677.

Snyder C., T.H. Hammil and S.B. Trier, 2003: Linear Evolution of Covariances in a Quasigeostrophic Model. *Monthly Weather Review*, **131**, 189–205.

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

Trémolet, Y., 2007: Model error estimation in 4D-Var. *Q. J. R. Meteorol. Soc.*, **133**, 1267–1280.

Whitaker J.S. G.P. Compo and J-N. Thépaut, 2009: A Comparison of Variational and Ensemble-Based Data Assimilation Systems for Reanalysis of Sparse Observations. *Monthly Weather Review*, **137**, 1991–1999.

Wirth A. and M. Ghil, 2000: Error Evolution in the Dynamics of an Ocean General Circulation Model. *Dyn. Atmos. Oceans*, **32**, 419–431.

A Eigenvalues of the Preconditioned Saddle Point System

We consider the eigenvalues of the preconditioned saddle point matrix for the inexact constraint preconditioner described in section 8, and for the case $\tilde{\mathbf{L}} = \mathbf{L}$. We have:

$$\tilde{\mathcal{P}}^{-1}\mathcal{A} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{L}^{-\mathrm{T}} \\ \mathbf{0} & \mathbf{R}^{-1} & \mathbf{0} \\ \mathbf{L}^{-1} & \mathbf{0} & -\mathbf{L}^{-1}\mathbf{D}\mathbf{L}^{-\mathrm{T}} \end{pmatrix} \begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{L} \\ \mathbf{0} & \mathbf{R} & \mathbf{H} \\ \mathbf{L}^{\mathrm{T}} & \mathbf{H}^{\mathrm{T}} & \mathbf{0} \end{pmatrix} \tag{65}$$

$$= \begin{pmatrix} \mathbf{I} & \mathbf{L}^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{R}^{-1}\mathbf{H} \\ \mathbf{0} & -\mathbf{L}^{-1}\mathbf{D}\mathbf{L}^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}} & \mathbf{I} \end{pmatrix} \tag{66}$$

$$= \mathbf{I} + \begin{pmatrix} \mathbf{0} & \mathbf{L}^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}^{-1}\mathbf{H} \\ \mathbf{0} & -\mathbf{L}^{-1}\mathbf{D}\mathbf{L}^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}} & \mathbf{0} \end{pmatrix} \tag{67}$$

Let σ denote an eigenvalue of the second matrix on the right hand side of equation 67. We see immediately that $\tau = \sigma + 1$ is an eigenvalue of $\tilde{\mathcal{P}}^{-1}\mathcal{A}$, since the addition of the identity matrix in this equation acts to shift the eigenvalues by one.

The eigenvalue σ satisfies:

$$\begin{pmatrix} \mathbf{0} & \mathbf{L}^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{R}^{-1}\mathbf{H} \\ \mathbf{0} & -\mathbf{L}^{-1}\mathbf{D}\mathbf{L}^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \end{pmatrix} = \sigma \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \end{pmatrix} \tag{68}$$

where $(u_1^{\mathrm{T}}, u_2^{\mathrm{T}}, u_3^{\mathrm{T}})^{\mathrm{T}}$ is the corresponding eigenvector.

The second and third rows of equation 68 give a pair of coupled equations for $\mathbf{u}_2$ and $\mathbf{u}_3$:

$$\mathbf{R}^{-1}\mathbf{H}\mathbf{u}_3 = \sigma\mathbf{u}_2 \tag{69}$$

$$-\mathbf{L}^{-1}\mathbf{D}\mathbf{L}^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{u}_2 = \sigma\mathbf{u}_3 \tag{70}$$

Multiplying the second equation by $\mathbf{H}$ and the first by $\mathbf{R}$ allows us to eliminate $\mathbf{u}_3$ to give the following generalised eigenvalue problem:

$$\mathbf{H}\mathbf{L}^{-1}\mathbf{D}\mathbf{L}^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{u}_2 = -\sigma^2\mathbf{R}\mathbf{u}_2. \tag{71}$$

Hence, we either have $\mathbf{u}_2 = 0$ (in which case it is easy to show from equation 68 that $\sigma = 0$), or the eigenpairs satisfy equation 71. In the latter case, we note that $\sigma^2 \leq 0$, since $\mathbf{H}\mathbf{L}^{-1}\mathbf{D}\mathbf{L}^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}}$ is positive semi-definite and $\mathbf{R}$ is positive definite. Hence, the eigenvalues σ lie on the imaginary axis of the complex plane. The corresponding eigenvalues τ of the preconditioned saddle point matrix lie on the line $\Re(\tau) = 1$, and are therefore bounded away from zero.

Moreover, for each non-zero eigenvalue σ we have:

$$\sigma = \pm\sqrt{\frac{\mathbf{u}_2^{\mathrm{T}}\mathbf{H}\mathbf{L}^{-1}\mathbf{D}\mathbf{L}^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{u}_2}{\mathbf{u}_2^{\mathrm{T}}\mathbf{R}\mathbf{u}_2}}\; i \tag{72}$$

The matrix $\mathbf{H}\mathbf{L}^{-1}\mathbf{D}\mathbf{L}^{-\mathrm{T}}\mathbf{H}^{\mathrm{T}}$ can be interpreted as a transformation of the matrix $\mathbf{D}$ to observation space. Thus, the numerator of equation 72 represents the sum of the background and model-error variance associated with $\mathbf{u}_2$, expressed in terms of the observed quantities. The denominator is the corresponding observation error variance. It would appear that, as is the case in strong constraint 4D-Var and in the forcing formulation of weak constraint 4D-Var, the conditioning of the preconditioned saddle point problem is controlled by the ratio of background error and model error variance to observation error variance. We can expect the problem to be well conditioned if this ratio is small, and poorly conditioned in the presence of accurate observations and large background or model error. It is is interesting to note that the conditioning of the saddle point problem depends on the square-root of the variance ratio, whereas the condition numbers of the forcing formulation of weak constraint 4D-Var and of strong constraint 4D-Var are directly proportional to the ratio of variances.

Finally, we note that the analysis presented above is for the case $\tilde{\mathbf{L}} = \mathbf{L}$. If $\mathbf{L}$ is approximated, then the eigenvalues of the preconditioned saddle point system are not constrained to lie on the line $\Re(\tau) = 1$. We have found in practice that many of the eigenvalues do lie on this line, with the remainder in a cloud surrounding the point $\tau = 1$. The size of this cloud presumably depends on the accuracy with which $\tilde{\mathbf{L}}$ approximates $\mathbf{L}$.