The University of Reading

Numerical Weather Prediction

Controlling Imbalance Within VAR by Use of a Weak Constraint

Forecasting Research Technical Report No. 417
Joint Centre for Mesoscale Meteorology Report No. 143

Mark Dixon and Ian Roulstone

email: nwp_publications@metoffice.com

Controlling Imbalance Within VAR by Use of a Weak Constraint

Mark Dixon and Ian Roulstone
JCMM

August 15, 2003

Abstract

The use of a weak constraint to control imbalance in the variational data assimilation system (VAR) has been investigated, the chosen constraint being that of incremental non-linear balance. This method of imposing balance is contrasted with that of the digital-filter technique and theoretical reasons are given as to why a weak constraint may lead to a more accurate analysis than is obtainable via digital filtering. In the experiments performed the algebraic relation embodying non-linear balance is weakly enforced by the constraint at individual grid points. However, it appears that this is achieved in a manner which is non-locally inconsistent, and consequently imbalance is not successfully controlled. It is suggested that this problem may be overcome by the incorporation of a suitable covariance matrix. Such an approach may be especially beneficial in the context of high-resolution forecasting, where gravity modes play a significant role.

1 Introduction

In this report we compare two methods for controlling imbalance in variational data assimilation (VAR). One involves the incorporation of a balance condition defined by a small Rossby number expansion of the primitive equations, via a weak constraint, and the other method is based on the application of the perturbation-forecast (PF) digital filter (DF) [1]. The problem of controlling imbalance in the process of data assimilation — that is, controlling the generation of gravity waves when fitting the model to the observations — is a problem with a long history, which can be traced back to Richardson's first numerical forecast [2]. In the process of data assimilation, experience has shown that the discrepancy between current data, with their random errors, and model first guess, with its errors, can excite a spuriously large amount of inertia-gravity waves in a primitive equation model. Typically these waves are damped-out over a period of 12-24 hours, and have been shown not to affect the 24-48 hour forecasts substantially. However, in an assimilation scheme they can lead to a rejection of data at the next sub-synoptic update time (at 6 hours, for example). Other difficulties arise too if the initial conditions are unbalanced, e.g. verification.

Work is in progress at the JCMM to design and build a convective-scale forecasting system [3]. One of the aims of this project is to develop a capability to provide quantitative precipitation forecasts with short lead times. Nowcasting systems provide information to the forecaster on the 0-6 hour time-scale: if this regime is to be covered by an NWP system, the problems of imbalance alluded to above cannot be tolerated. It is also anticipated that the data assimilation for this model will be designed to make good use of moisture and precipitation data, but this raises the challenge of creating effective ways of assimilating

observations associated with condensed water in such a way that the dynamical information inherent in these observations is used. This is a highly non-linear problem and one in which the coupling between moisture, vertical motion and horizontal momentum balance could lead to the generation of unrealistic gravity modes. The problem of an initial 'spin-up' of precipitation within a model is another well-known difficulty and one that is intimately connected to the problem of imbalance.

With these issues in mind, we are investigating how to control imbalance using various techniques within the data assimilation system. The formulation of the data assimilation system for the high resolution forecasting system has yet to be defined (e.g. interpolation from coarser resolution models [4]; VAR; Ensemble Kalman Filter) and so these investigations are intended to help inform that decision. In this paper we present some preliminary results from a study into the incorporation of a mass-wind coupling within 3D VAR. This initial study has been carried out with the global model, which is appropriate since the scales of motion we wish to examine in the first instance are well represented by this model.

1.1 The problem of spurious gravity waves

In 3D VAR, the best estimate of the current atmospheric state, otherwise known as the *analysis* is defined as the vector $\mathbf{x}$ that minimises the following cost function, which consists of the background penalty, J^b, and the observation penalty, J^o:

$$\begin{aligned} J(\mathbf{x}) &= J^b + J^o \\ &= \frac{1}{2}(\mathbf{x}-\mathbf{x^b})^{\mathbf{T}}\mathbf{B^{-1}}(\mathbf{x}-\mathbf{x^b}) \\ &+ \frac{1}{2}(\mathbf{y}(\mathbf{x})-\mathbf{y^o})^{\mathbf{T}}(\mathbf{E}+\mathbf{F})^{-\mathbf{1}}(\mathbf{y}(\mathbf{x})-\mathbf{y^o}) \end{aligned} \tag{1}$$

Here $\mathbf{x}^b$ is the background state, and $\mathbf{B}$ is the associated background error covariance matrix. $\mathbf{y}^o$ are real observations and $\mathbf{y}(\mathbf{x})$ are model-predicted observations. $\mathbf{E}$ and $\mathbf{F}$ are the instrumental and interpolation error covariance matrices, respectively. See [5] for a detailed description of VAR. We define the increment vector as follows:

$$\mathbf{x}' = \mathbf{x} - \mathbf{x}^b \tag{2}$$

Let $\mathbf{x}'(t)$ denote a set of increments obtained via minimisation of (1), valid at time t. The evolution of the increments from time t to time $t+1$ is given by:

$$\mathbf{x}'(t+1) = \mathbf{M}_t\mathbf{x}'(t) \tag{3}$$

where $\mathbf{M}_t$ is a matrix corresponding to a linearisation of the non-linear forecast model at time t. $\mathbf{x}'(t)$ can be expressed as a linear combination of the normal modes of $\mathbf{M}_t$:

$$\mathbf{x}'(t) = \sum_i^N a_i\mathbf{s}_i + \sum_j^M b_j\mathbf{f}_i \tag{4}$$

where $\mathbf{s}$ are the N 'slow' or balanced modes, $\mathbf{f}$ are the M 'fast', unbalanced modes, and a_i and b_i are the weights of each type of mode at the analysis time. In general the *true* state of the atmosphere will be composed of both fast and slow modes. The increments defined by (3) can contain both *realistic* and *spurious* fast modes. By *realistic* it is meant that a fast mode which is truly present in the atmosphere has been closely approximated by a fast mode in $\mathbf{x}'$. A spurious mode, conversely, represents a case where VAR has erroneously fitted

a fast mode to observations which are characteristic of a slow mode. In the absence of a method by which spurious and realistic modes can be distinguished, it is common practice to post-process the increments via *initialisation* techniques such as the Incremental Analysis Update (IAU) [6], DF methods (see, for example, [7]) or normal-mode initialisation [8]. Each of the procedures are used to filter fast modes, be they accurate or spurious, from increments *after* VAR has been performed. It may be thought that a useful alternative approach would be to preclude the fast modes from the space spanned by $\mathbf{B}$. However, this would lead to the erroneous fitting of slow modes to real fast modes (i.e. it would lead to spurious slow modes) and hence an inaccurate analysis.

An alternative approach to the problem is to enforce a weak balance constraint *during* the minimisation process by including a so-called J_c term in the cost function of (1). Since minimisation of (1) leads to observational information being incorrectly projected onto fast modes, this implies that the weightings, a_i, of the slow modes in $\mathbf{x}'$ are not optimal. By contrast, including a J_c term in (1) could lead to a more optimal set of both fast and slow modes. The formulation of the J_c term considered here is intended to impose incremental non-linear balance (INLB). The strategy of imposing INLB as a weak constraint is motivated by the relationship, which exists between INLB and gravity modes.

1.2 Contents of this report

The structure of the report is as follows. The theoretical background, including details both of balanced motion and weak constraints is presented in section 2. Section 3 documents the experiments, and we conclude, in section 4, with a list of recommendations for further work.

2 Theoretical Background

2.1 Balanced models

The balance between the mass and wind fields which exists at large scales, ensures that the flow is dominated by slow, quasi-rotational rather than high-frequency, strongly divergent motion. In linear terms, balanced motion is associated with slow Rossby modes, whilst imbalance is associated with high-frequency gravity and acoustic modes. A balanced model seeks to accurately approximate the Rossby-type flow alone, the high-frequencies having been filtered out. Filtering is achieved by the elimination of time derivatives, leading to diagnostic *balance relations*. The present formulation of VAR is hydrostatic and anelastic (so acoustic modes have been filtered), and hence supports N Rossby and $2N$ gravity modes, where N is the number of grid points. Constructing a balanced model in this context equates with filtering the $2N$ gravity modes which, in turn, implies the neglect of two time derivatives.

2.1.1 The Balance Equations

A number of different balanced models have been derived in the literature. To illustrate the link between balance and the normal modes of a system we concentrate on the models presented in [9]. Consider the relatively simple case of the β-plane shallow-water (SW) equations linearised about a state of rest:

$$\frac{\partial \zeta}{\partial t} + f\delta + \beta v = 0 \tag{5}$$

$$\frac{\partial \delta}{\partial t} + (\nabla^2 \phi - f\zeta + \beta u) = 0 \tag{6}$$

$$\frac{\partial \phi}{\partial t} + \bar{\phi}\delta \tag{7}$$

where ζ is the vorticity, δ is the divergence, ϕ is the geopotential and $\bar{\phi}$ is the geopotential of the linearisation state. The Coriolis parameter is given by $f = f_o + \beta y$, where f_o is a constant and y is the 'latitudinal' coordinate on a Cartesian plane. As shown in [9], if the height variable is geostrophically scaled and all of the variables are expanded in terms of Rossby number, R_o ($= V/2\Omega L$, where V and L are horizontal velocity and length scales respectively, and Ω is the Earth's angular speed of rotation), we can obtain the so-called balanced equation system (BE) correct at order $O(R_o)$. This system consists of equations (5) and (7) along with a truncated version of (6) in which the local time derivative of divergence is omitted, i.e.:

$$\nabla^2 \phi - f\zeta + \beta u = 0 \tag{8}$$

This diagnostic relationship is known as the balance equation. Substituting plane wave solutions into (BE) leads to a quadratic dispersion relation, corresponding to a Rossby and high-frequency spurious mode.[1]

2.1.2 The Slow Equations

Lynch [9] proposes an alternative balanced system referred to as the slow equations (SE). This system is derived by first defining the imbalance, $\epsilon = \nabla^2\Phi - f\zeta + \beta u$. By re-casting the SW system in terms of prognostic equations for ζ, ϵ and δ and performing a Rossby expansion as before, it can be shown that the local time derivative of ϵ is of order $O(R_o^3)$. Hence, the slow equation system, which is valid up to order $O(R_o)$, is given by (5), (8) and:

$$(\bar{\phi}\nabla^2 - f^2)\delta - f\beta v = 0 \tag{9}$$

which is the diagnostic relation formed by omitting $\frac{\partial \epsilon}{\partial t}$. Since the SE system contains only one time derivative it supports only one (balanced) mode.

2.1.3 Definition of INLB to be used in VAR

We have opted to control imbalance within VAR by imposing the form of balance characterised by the BE system, rather than that of the SE system. One aim of using the J_c term might be to constrain the VAR increments so that when they are used as initial conditions of the linearised forecast model, the resulting trajectory closely approximates the trajectory that would be produced by running a linearised BE model using identical initial conditions. This corresponds to ensuring that the incremental divergence tendency, during the period of the trajectory, is of order $O(R_o^2)$ at most, i.e.:

$$\frac{\partial(\nabla.\mathbf{u}')}{\partial t} = \nabla.\left(\mathbf{u}'.\nabla\bar{\mathbf{u}} + \bar{\mathbf{u}}.\nabla\mathbf{u}' + f\mathbf{k}\wedge\mathbf{u}' + \frac{1}{\bar{\rho}}\nabla p' - \frac{\rho'}{\bar{\rho}^2}\nabla\bar{p}\right) \leq O(R_o^2) \tag{10}$$

where ∇ is applied on a surface of constant height. We have neglected terms involving the vertical velocity and assumed a time-independent linearisation state. We refer to (10) as the INLB equation. An alternative aim would be to include a certain degree of *realistic* gravity-wave activity in the increments. In this case, the left-hand side of (10) would be allowed to

[1] Note that this latter mode can be filtered by neglecting βu_χ (where u_χ denotes the divergent part of the velocity) in (6), leading to a linearised version of the so-called Charney balance equation.

take on larger values than of $O(R_o^2)$. We discuss the dynamical implications of these two different aims in section (2.2.2), below. It is important to note that in either of these cases, the balance relation is to be imposed as a *weak* constraint, i.e. the balance relation is imposed an *inequality*. In the case of a *strong* constraint, by contrast, the balance relation is satisfied exactly throughout space. Imposing the left hand side of (10) as a strong constraint would prevent the divergence field from evolving, and hence lead to innacurate increments. [2]

2.2 Weak constraints

The concept of a weak constraint was presented by Sasaki [10]. Here we consider a simple example before moving on to discuss the J_c term itself.

2.2.1 A simple example: damping high spatial frequencies

Consider the following illustrative example (after Daley [8]). A set of observations is given by the continuous 1D function $\phi^o(x)$. We wish to perform a least-squares fit to these observations leading to an analysis, $\phi^a(x)$, subject to a constraint that reduces high spatial frequencies. For the interval $x_0 \leq x \leq x_1$, the cost function is given by:

$$J = \delta \int_{x_0}^{x_1} \alpha_o(\phi^o(x) - \phi^a(x))^2 + f^2 dx \tag{11}$$

where the prespecified weighting factors, α_o and α_c, are assumed to be independent of x. The constraint function, f, is chosen to be:

$$f = \frac{d^2\phi^a}{dx^2} \tag{12}$$

Setting the first variation of J, $\delta J = 0$ leads to the following Euler-Lagrange equation:

$$\alpha_o\phi^a + \alpha_c\frac{d^4\phi^a}{dx^4} = \alpha_o\phi^o \tag{13}$$

If we assume the following form for ϕ^o:

$$\phi^o(x) = \phi_0 - \phi_1 e^{ikx} \tag{14}$$

where the amplitudes ϕ_0 and ϕ_1 are independent of x, then the solution to (13) is given by:

$$\phi^a(x) = \phi_0 + \frac{\alpha_o\phi_1 e^{ikx}}{\alpha_o + \alpha_c k^4} \tag{15}$$

Clearly, the constraint serves its purpose since the degree of damping of the x-dependent part of (15) increases with k. Furthermore, in the limit of large α_c, all spatial dependency is removed. In general, the weighting function can depend on x thereby changing the nature of the solution. Such a procedure would be used if, for example, we wanted to retain more high-frequency components in a sub-region of the total domain.

[2] By way of analogy, we note that in a hydrostatic model the *diagnosed* vertical velocity must be, in general, non-zero.

2.2.2 Controlling temporal frequencies: the J_c term

The J_c term is a weak constraint, analogous to that of $\alpha_c \frac{d^2\phi}{dx^2}$ in the previous example, which is appended to the normal total cost function of (1):

$$J = J^b + J^o + J^c \tag{16}$$

In an attempt to impose INLB on the increments, we define J_c as follows:

$$J_c = w_{J_c} \frac{1}{2} \frac{1}{\sum_i a_i} \sum_i a_i \left(\frac{\partial \delta'(0)}{\partial t}\right)^2 \tag{17}$$

where the sum is over all grid points in model space, (ϕ, λ, η), the scalar w_{J_c} is a weighting factor (assumed to be independent of position), a_i is the mass [3] and area weighting of the ith grid point:

$$a_i = (\cos(\phi) * \Delta p)|_i. \tag{18}$$

The divergence tendency in (17) is evaluated at the initial time only, but for INLB to be imposed effectively, the degree of divergence tendency must not change rapidly with time, i.e. the imbalance tendency must be of order $O(R_o^3)$.

It might be thought that inclusion of this term would lead to a uniform value of incremental divergence tendency throughout the domain. On the other hand, information spread (during minimisation) between grid points via J^b and J^o could lead to spatially non-uniform values of the incremental divergence tendency. We can hypothesize that the inclusion of (17) in the total cost function should generally reduce the degree of incremental divergence tendency throughout the domain. Furthermore, we might expect that (in analogous fashion to the simple case of section 2.2.1) the divergence tendency should decrease with increasing w_{J_c}. Hence, naive reasoning suggests that it may be possible to tune w_{J_c} so that the increments either contain no gravity waves or a realistic degree of such activity. This choice points to a fundamental difference between weak constraints and (post-processing) initialisation methods, which it is vital to bear in mind. Assume, for the sake of simplicity, that we wish to impose SE balance, hence avoiding the complications of the spurious BE modes. As was alluded to in the introduction, it is important that the space spanned by **B** includes all degrees of freedom (i.e. includes both fast and slow modes) in order to prevent observational information pertaining to fast modes from being incorrectly projected onto slow modes. If the weak constraint were used to filter out all of the fast modes, then this would be equivalent to limiting the space spanned to balanced modes only. It is likely that such a procedure would lead to a less accurate set of increments than obtainable via initialisation methods.[4]. On the other hand, if the weak constraint led to the inclusion of realistic gravity waves and the exclusion of spurious gravity waves, then it is likely that the slow-mode component would be a better representation than that obtained via filtering methods. The degree to which such a constraint is practicable is not clear. Ideally, instead of the simple formulation of (17), we would need to impose the chosen balance relation(s) in a way that better represents the way in which they are held in the *true* atmospheric state [5]. Our results suggest that this

[3] It is arguable that a mass weighting should not be included, but this is likely to have little effect on the main results presented in this report

[4] If the true state did not contain any fast modes, spurious fast modes would still contaminate the increments given the current VAR system. Hence a J_c term aimed at imposing balance would be of use even in this case

[5] Note that using the true state of the atmosphere as a set of initial conditions for the UM would not guarantee an accurate forecast, as the underlying equations are of course approximations. Similarly, we would not necessarily want to impose the balance relations *exactly* as they hold in the true state

would require the inclusion of a covariance matrix in the J_c term, of the kind discussed in the final section of this report. [6]

3 Investigations within VAR

3.1 Experimental details

All of the experiments used the New Dynamics test case of 12Z 26/12/99, at a resolution of 2.5^o in the latitudinal direction and 3.75^o in the longitudinal direction, with 38 vertical levels. A constant linearisation state (corresponding to 12Z) was used throughout; we expect the effects of this on the results to be minimal. In what follows, the following terminology is adopted. **Control** experiments correspond to VAR minimisations without any initialisation procedures applied. **DF** corresponds to experiments where the PF digital filter [1] has been applied; **w17**,**w18** and **w19** refer to minimisations where the J_c term has been included in the total cost function and w_{J_c} has been set to 10^{17}, 10^{18} and 10^{19} respectively. The three values of w_{J_c} were chosen to facilitate a comparison with the DF method.

3.2 The effect of DF on the increments

First we consider the effect of DF on the incremental balance. These results indicate the degree to which $\frac{\partial \delta}{\partial t}$ and $\frac{\partial \epsilon}{\partial t}$ are reduced by the filtering of fast modes, and will be of use in interpreting the effect of the J_c term (see next section).

Figure (1) shows vertical cross-sections running along 50^oN of the divergence tendency associated with the control and DF-initialised increments at the analysis time, T+0. Application of the DF clearly reduces $\frac{\partial \delta}{\partial t}$. Figure (2) shows estimates of the first time derivative of imbalance, i.e. $\frac{\partial \epsilon}{\partial t}$, formed by differencing the T+0 state and a PF model T+15min state, for the case of the control and DF experiments. The DF leads to a significant reduction in the magnitude of $\frac{\partial \epsilon}{\partial t}$. Note, for future reference, that these values imply that the degree of INLB obtained via the DF at the initial time will be maintained on a time-scale longer than that of an hour.

3.3 The effect of J_c on the increments

Figure (3) shows the total cost function and its components as a function of VAR iteration number for the case of the control experiment. Given the default convergence criterion (a relative change in the norm of cost function gradient of around 0.005), the minimisation would be stopped after around 60 iterations. Also shown is the divergence-tendency penalty, which was passive in this case, corresponding to $w_{J_c} = 10^{17}$.

Figures (4)-(6) show the behaviour of the cost functions in the cases of **w17**,**w18** and **w19** respectively. All lead to a reduction in the converged value of J_c. In each case convergence is obtained, as in the control experiment, after 60 iterations. A measure of the global divergence tendency in each case may be obtained by dividing the converged value of J_c by w_{J_c}. As hypothesised, the divergence tendency decreases as w_{J_c} increases. The converged value of J_o is larger than in the control case and increases with w_{J_c}. In other words, the J_c worsens the fit to observations. Presumably this is because the J_c term is erroneously projecting observational information which characterises fast modes onto slow modes. Note that the converged value of J_b decreases with increasing w_{J_c} suggesting that, to some degree, the

[6] The covariance matrix in (17) is the (real-space) identity matrix

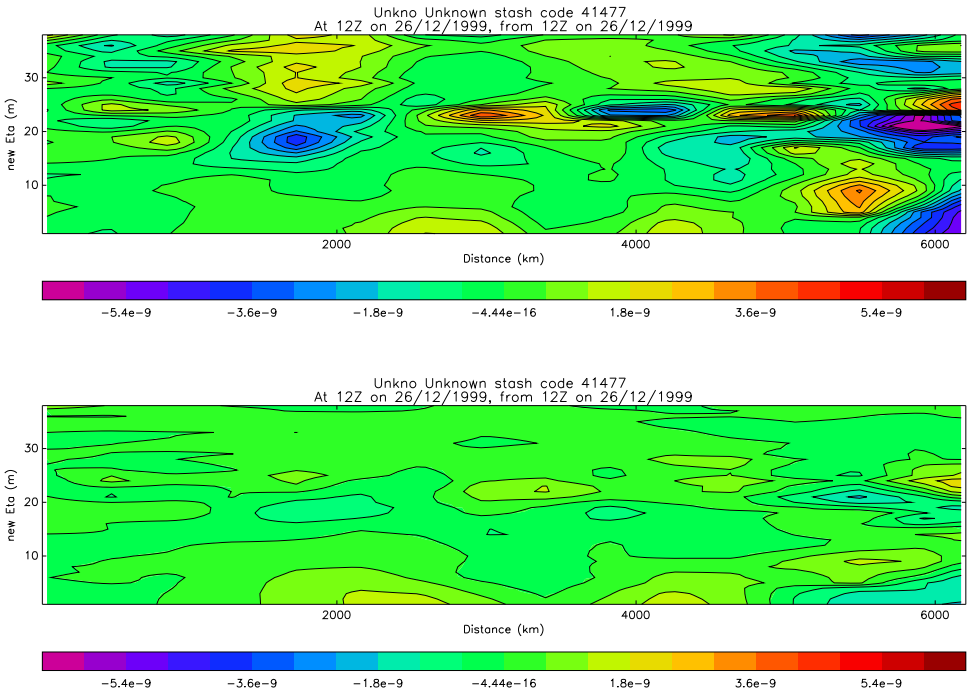

Figure 1: Vertical cross-section of the divergence tendency of control increments (upper) and DF-initialised increments (lower). The section runs along 50^oN and the vertical coordinate is the model level number, with 38 corresponding to the top level. Units = s^{-2}

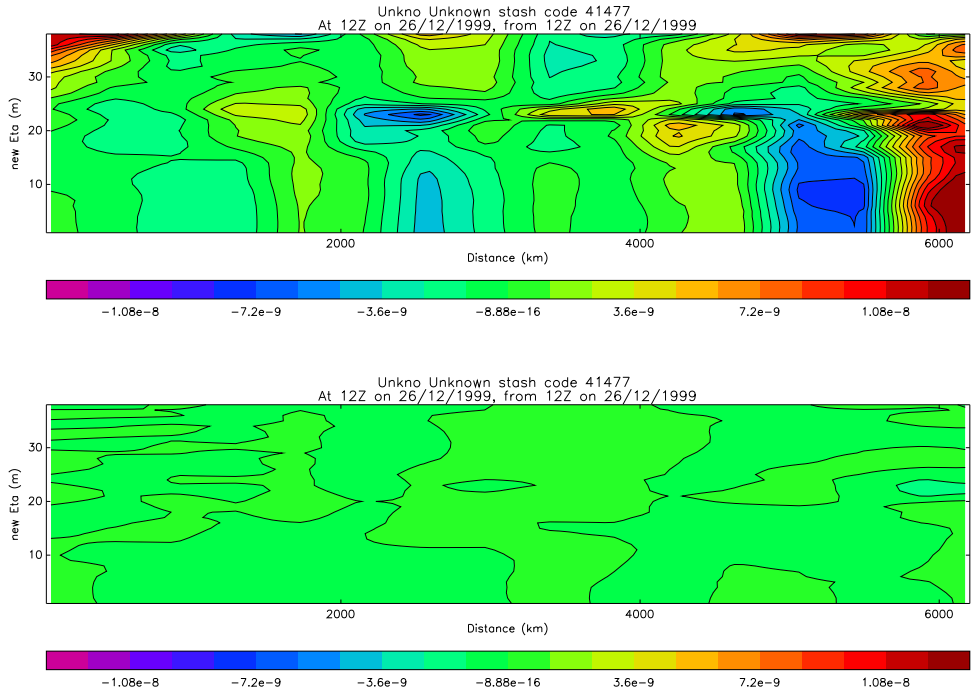

Figure 2: Cross-sections (as in Figure 1) of $\frac{\partial \epsilon}{\partial t}$ from the control experiment at T+0 (upper) and T+15min (lower). Units = $s^{-2}h^{-1}$

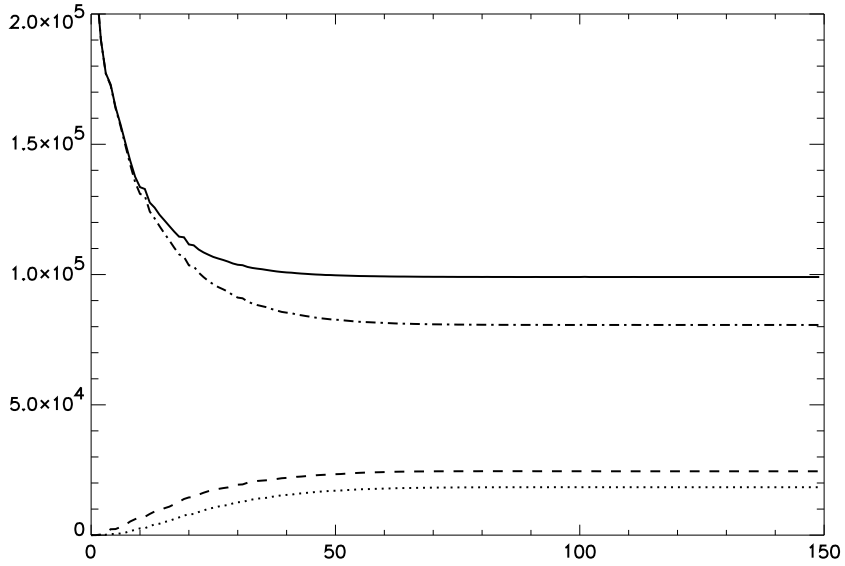

Figure 3: Total and component cost functions as functions of iteration number in the case of the control experiment. Total (full line), J_o (dot-dashes),J_b (dots) and (passive) J_c (dashes)

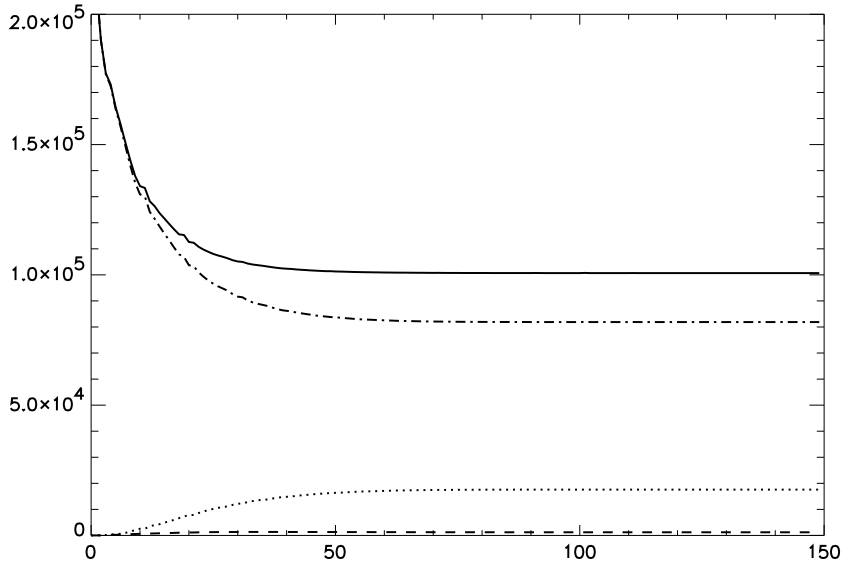

Figure 4: Total and component cost functions as functions of iteration number in the case of the $w17$. Total (full line), J_o (dot-dashes),J_b (dots) and J_c (dashes)

divergence tendency is reduced via the reduction in increment size rather than by a change in morphology of the mass and wind fields. Figure (7) shows cross-sections of the divergence tendency at T+0 in the case of the control and the three J_c experiments. It is clear that, in general, the magnitude of divergence tendency decreases with increasing w_{J_c}. Figures (1) and (7) suggest that the weighting factor must be set to a value in the range of 10^{17} and 10^{18} to obtain a similar degree of divergence tendency as is present in the DF-initialised increments. [7] Furthermore, this implies that the constraint has been applied too strongly in the case of w19. That the J_c term leads to a reduction in the overall degree of divergence tendency does not guarantee that INLB has been successfully imposed. Rather, this reduction must be achieved in a manner which is (spatially) non-locally consistent. The following results imply that this latter condition has *not* been satisfied. Figure (8) shows the imbalance tendencies produced in each experiment. In all of the cases $\frac{\partial \epsilon}{\partial t}$ exceeds that obtained via the DF procedure. Assuming that the constraint has been applied too strongly in the case of w19, we concentrate on w17 and w18. Whilst the J_c term has significantly reduced $\frac{\partial \epsilon}{\partial t}$ in some regions (such as close to the tropopause), elsewhere the values are equal to, or larger than, those associated with the control. Assuming that these growth-rates are maintained, $\frac{\partial \delta}{\partial t}$ (i.e. the departure from INLB), will attain values in places which exceed those imposed at the initial time.

The J_c term is not expected to limit the size of $\frac{\partial \epsilon}{\partial t}$ to the extent to which the DF method does, since the imposition of INLB does not remove the spurious modes associated with the BE system. However, ideally the weak constraint should maintain INLB on the same time-scale as achieved via the DF. Although the initial state may contain growing (as opposed to neutral) fast modes, the DF method shows that it is possible to damp the initial amplitudes

[7] Attempting to set the degree of INLB to that obtained via the DF is not, as explained earlier, advisable from the viewpoint of accuracy

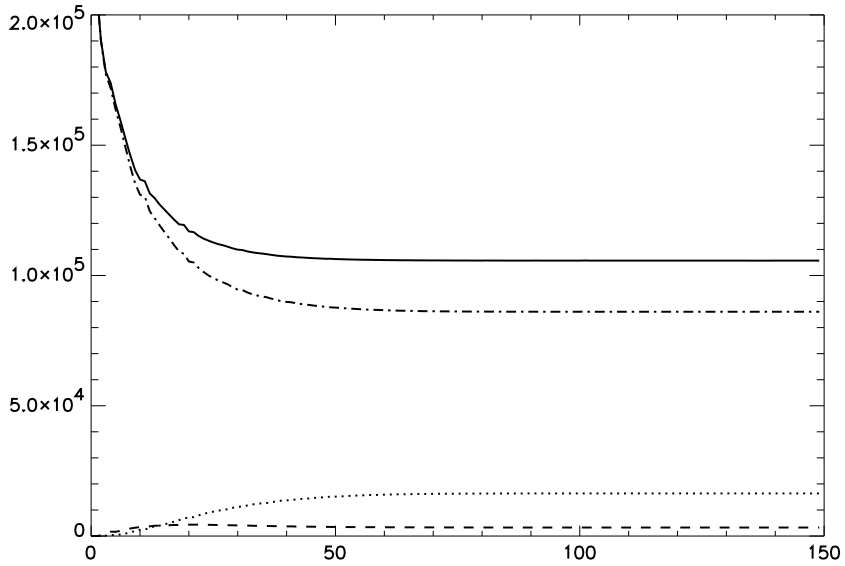

Figure 5: Total and component cost functions as functions of iteration number in the case of the $w18$. Total (full line), J_o (dot-dashes),J_b (dots) and J_c (dashes)

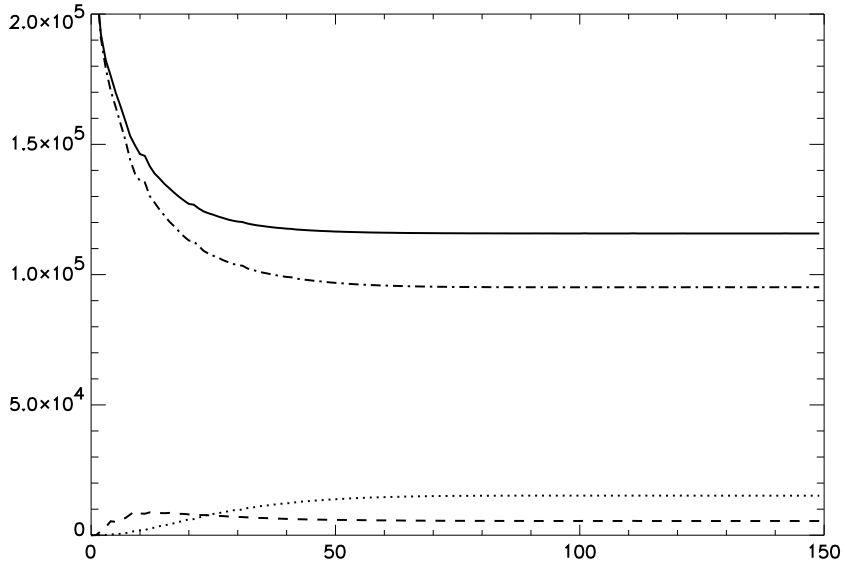

Figure 6: Total and component cost functions as functions of iteration number in the case of the $w19$. Total (full line), J_o (dot-dashes),J_b (dots) and J_c (dashes)

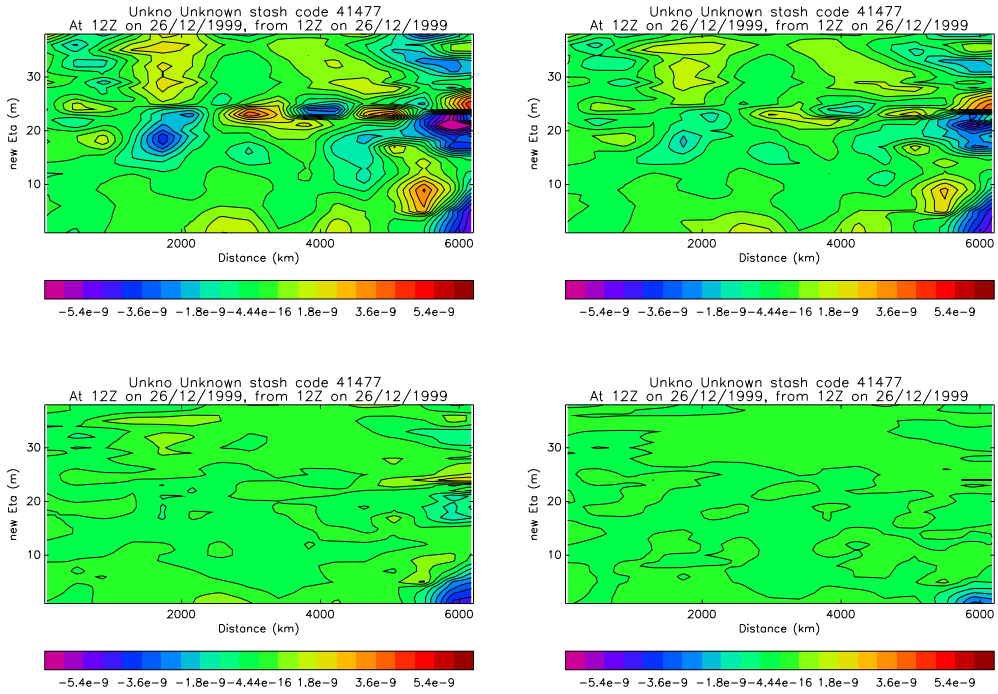

Figure 7: Cross-sections (as in Figure 1) of divergence tendencies at T+0. Control (top left), w17 (top right), w18 (bottom left) and w19 (bottom right). Units = s^{-2}

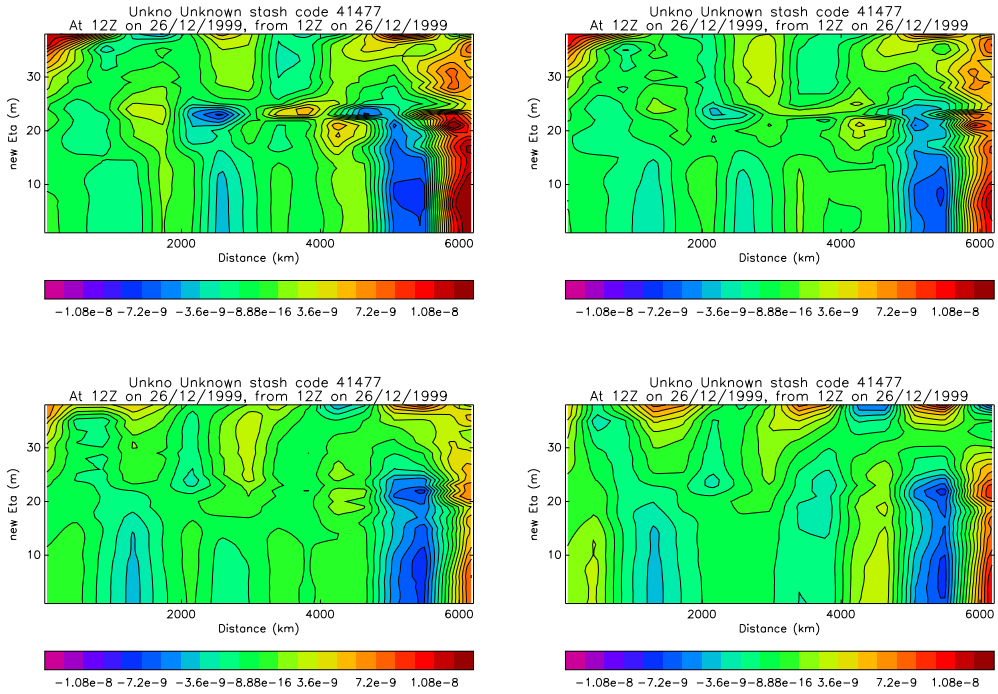

Figure 8: Cross-sections (as in Figure 1) of imbalance tendencies at T+0. Control (top left), w17 (top right), w18 (bottom left) and w19 (bottom right). Units = $s^{-2}h^{-1}$

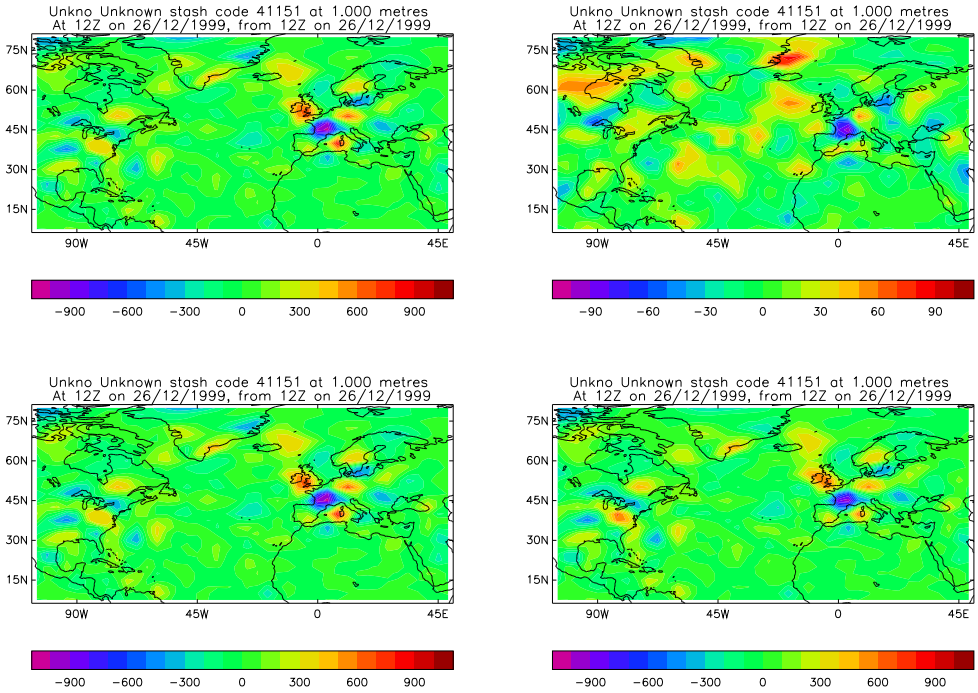

Figure 9: Plots of pressure tendencies at T+0 on model level 1. Control (top left), DF (top right), w17 (bottom left) and w18 (bottom right). Units = Pah^{-1}. Note that the colour scale in the case of the DF experiment is ten times smaller than in scale used in the other three plots. Similar behaviour is exhibited throughout the entire global domain

of such modes sufficiently for balance to hold on a longer time-scale than an hour (see Figure 2).

The relatively large values of imbalance tendency exhibited in Figure 8 are symptomatic of a fundamental problem with the J_C term, which can be seen more clearly by considering the resulting behaviour of other dynamical fields. Figure 9 shows the pressure tendencies (formed via differences of T+0 and T+15 fields) on model level 1. The control increments are contaminated with erroneously high values, attaining a maximum of around $10mbh^{-1}$. The DF reduces the magnitude of the pressure-tendency field by a factor of ~ 10, while the J_c fails to reduce the magnitude of this field significantly. Moreover, the fields in the case of w17, w18 and the control are of a very similar morphology, implying that the J_c term has had little impact on the PF trajectory. As alluded to above, we propose that this behaviour is indicative of a failure to constrain the divergence tendency in a non-locally consistent manner. The J_c term works by imposing an algebraic relationship between mass, wind and density fields at each grid-point. Whilst the value of $\frac{\partial \delta}{\partial t}$ at each grid point may be decreased by the inclusion of J_c, this does guarantee that this has occurred in a consistent non-local fashion; some means of spreading between grid points information pertaining to the balance relation (10) is required. As suggested in section 2, it is possible that such information could be implicitly spread via J^b and J^o. That they are failing to do this, might explain why INLB

is not maintained. It should be noted that these arguments apply equally in the case where the J_c term is intended to filter out all gravity modes, and in the case where realistic gravity modes are to be retained, since both cases are non-local problems.

4 Discussion and recommendations for future work

In summary, the J_c term leads to a convergent solution, and the degree of divergence tendency decreases as the J_c weighting factor is increased. However, the behaviour of the imbalance and pressure tendencies imply that INLB is not successfully imposed. The most likely cause of this is that the divergence tendency is altered in a non-locally inconsistent manner. Below, we consider possible ways of improving the formulation of J_c.

4.1 Covariance matrix

A method of explicitly including non-local information into the weak constraint is to include a covariance matrix in the J_c term as follows:

$$J_c = \left(\frac{\partial(\nabla.\mathbf{u}'(0))}{\partial t}\right)^T \mathbf{G}^{-1} \left(\frac{\partial(\nabla.\mathbf{u}'(0))}{\partial t}\right) \tag{19}$$

where $\mathbf{G}$ is the covariance error matrix of the background divergence tendency. One possible way of constructing $\mathbf{G}^{-1}$ would be to adopt the NMC approach, i.e. form a model-climatological estimate of $\mathbf{G}^{-1}$ using an average calculated by taking differences between T+48 and T+24 forecasts. Whether such a estimate would lead to the inclusion of realistic fast modes and the exclusion of spurious modes is not clear. To do this, it may be necessary to include further flow-dependent information in $\mathbf{G}^{-1}$, which would be a complex task. [8]

Practical implementation of this method would involve a significant simplification since the size of covariance matrix in real space is too large. This problem could be dealt with by including, for each grid-point, the inverse covariances associated with a relatively small number of surrounding points. Investigations of how many surrounding points are required for adequate convergence would need to be performed.

It is likely that even an improved formulation of the J_c term would not be able to prevent *all* erroneous projection of observational information onto fast modes, but it may reduce such contamination to an acceptable level and hence could be used in place of the DF. On the other hand, if the contamination is too great, then it may be worthwhile implementing the J_c term and then applying the DF; the rationale being that the J_c term would lead to a more accurate estimate of the slow modes which are retained by the DF process (see section 1.1, above).

4.2 A further J_c term

As discussed in section 2, the INLB system supports a spurious high-frequency mode. Even if we could successfully control the size of the divergence tendency using a J_c term, i.e. so that it is of the order $O(R_o^2)$, the rate of change of divergence tendency of the resulting

[8] To implement a J_c term intended to remove *all* gravity waves during VAR (which, as already stated, is inadvisable), the relevant $\mathbf{G}$ should only span the space defined by modes of the primitive equations which approximate those of the BE system.

increments may still be too large (larger than $O(R_o^3)$). We could attempt to control this mode by introducing a second J_c term of the form:

$$J_c^{(2)} = \left(\frac{\partial \epsilon'(0)}{\partial t}\right)^T \mathbf{G}_\epsilon^{-1} \left(\frac{\partial \epsilon'(0)}{\partial t}\right) \tag{20}$$

where $\mathbf{G}_\epsilon$ is the relevant covariance matrix.

4.3 High-resolution modelling

At the horizontal resolutions of the convective-scale forecasting system being developed at JCMM [3] (i.e. 4 km to 1 km), an accurate model solution will, in general, contain a significantly larger degree of gravity modes than in the global model runs considered in this report. The question arises whether the data assimilation procedure should be designed to capture these important fast modes along with the slow modes, or whether we should aim to capture the slow mode component alone. The approach being adopted presently [4] aims to do the latter, using reconfigured low-resolution (i.e. 12 km or 4 km) VAR increments introduced to a high-resolution (i.e. 4 km or 1 km) background via the IAU. However, it is certainly worth investigating the use of a J_c term in order to incorporate gravity modes into the increments (or to obtain a better characterisation of the slow modes).

Acknowledgments

Thanks to Ross Bannister, Adam Clayton, Olaf Stiller and Andy White for useful discussions.

References

[1] Clayton, A., http:/www-nwp/~frca/ND_init/options/options.html

[2] Richardson, L. F., 1965, Weather prediction by numerical process. CUP, Cambridge, reprinted Dover, New York, 236 pp.

[3] High resolution trial model homepage, http://mm0100/~apma/hrtm/hrtm.html

[4] Dixon, M., http://vulcan/~apmd/meso/high_res_main.html

[5] Lorenc, A. C., Ballard, S. P., Bell, S., Ingleby, N. B., Andrews, P. L. F., Barker, D. M., Bray, J.R.,Clayton, A. M. ,Barker, D. M., Bray, J. R., Clayton, A. C., Dalby, T., Li, D., Payne, T. J. and Saunders, F. W., 2000, The Met Office Global 3-Dimensional Variational Data Assimilation Scheme *Q. J. R. Meteorol. Soc.* **126** pp. 2991-3012.

[6] Clayton, A., 2003, UMDP 31: Incremental Analysis Update (IAU) Scheme

[7] Lynch, P.,Digital filters for numerical weather prediction, HIRLAM Technical Report No. 10 [9]

[8] Daley, R., Atmospheric Data Analysis. *Cambridge University Press*

[9] Lynch, P., 1989 *Q. J. Roy. Met. Soc.* **114** pp.201-219. The Slow Equations.

[10] Sasaki, Y., 1970, Some Basic Formalisms In Numerical Variational Analysis. *Mon. Wea. Rev.* bf 12 pp.875-883

[9] available via http://www.maths.tcd.ie/~plynch/Publications/Publications.html