A Regime-dependent balanced control variable based on potential vorticity

Ross N. Bannister(1) and Mike J. P. Cullen(2)

(1) Data Assimilation Research Centre, Dept. of Meteorology, Univ. of Reading, Reading, RG6 6BB, UK
(2) Meteorological Office, Fitzroy Road, Exeter, EX1 3PB, UK

ABSTRACT

In this paper it is argued that rotational wind is not the best choice of leading control variable for variational data assimilation, and an alternative is suggested and tested. A rotational wind parameter is used in most global variational assimilation systems as a pragmatic way of approximately representing the balanced component of the assimilation increments. In effect, rotational wind is treated as a proxy for potential vorticity, but one that it is potentially not a good choice in flow regimes characterised by small Burger number.

This paper reports on an alternative set of control variables which are based around potential vorticity. This gives rise to a new formulation of the background error covariances for the Met Office's variational assimilation system, which leads to flow dependency. It uses similar balance relationships to traditional schemes, but recognises the existence of unbalanced rotational wind which is used with a new anti-balance relationship. The new scheme is described and its performance is evaluated and compared to a traditional scheme using a sample of diagnostics.

1 Introduction

The background state used in three and four dimensional variational data assimilation (VAR) constrains the assimilation towards a numerical forecast of the atmosphere. The manner in which this is done is governed by the background error covariance matrix, **B**, which describes the (Gaussian) probability distribution function (PDF) of forecast errors. In VAR systems of operational scale, the size of the problem renders it prohibitive to be able to generate complete information about the structure of **B**, or to represent it as an explicit matrix. To allow VAR to work, **B** is represented in a compact form which makes assumptions about how elements of the forecast vector are correlated with other elements. Control variable transforms are used for this purpose (see Sec. 2). A usual assumption is that the wind field is correlated with the mass field in a way that is consistent with diagnostic balance relationships, e.g. by geostrophic balance.

The formulation of the **B**-matrix is thought to be extremely important to the ability of VAR to give a realistic state of the atmosphere, as, e.g. analysis increments are parallel to the **B**-matrix. The **B**-matrix is particularly important in poorly observed regions and for poorly observed variables. An important concern is that the **B**-matrix has inadequate flow dependence, as it is common practice to consider **B** to be static (or quasi-static) allowing it to describe only the average characteristics of the PDF. There are strong arguments for reviewing the way that **B** is treated in VAR to allow it to capture flow dependence. For instance it is well known that background error covariance structures (as predicted by the Kalman filter equations in simpler systems) distort with the flow, and can deviate greatly from the static case, especially in areas of strong instability. The Kalman filter equations cannot be applied fully to problems of operational scale, so instead flow-dependent 'fixes' to VAR are sought that hopefully mimic the most important properties of the Kalman filter. Apart from the 4d-

VAR method itself, which implicitly propagates an otherwise static **B**-matrix through the assimilation window, there have been many attempts to modify the **B**-matrix formulation to allow flow dependent aspects of the error statistics to be utilized in VAR. Important examples are listed under the following headings.

- Variance updating: The grid-point variances of variables can be dynamically updated from cycle-to-cycle, e.g. as used with some variables by a cycling algorithm in the ECMWF system (Fisher and Courtier 1995).

- Control variable transform modifications: Flow dependent relations (rather than static relations) can be introduced to relate variables in a way that adapts with the background state. For instance wind and mass variables can be related with a non-linear balance equation (Fisher 2003), and the spatial structure of point-to-point correlation functions can be adjusted to fit the flow, e.g. by using a geostrophic co-ordinate transform (Desroziers 1997).

- Special treatment of fast-growing features: The most active modes of the dynamical system can be diagnosed (e.g. by the breeding method or by determination of the fastest growing singular vectors) and analysed separately from the rest of the flow (the rest being treated with standard VAR). Example methods are the reduced rank (or simplified) Kalman filter (Fisher 1998), and the errors of the day scheme (e.g. Barker and Lorenc 2005).

All of these variants rely on the control variable transform (CVT) method to formulate some or all of the **B**-matrix. As explained in Sec. 2, control variables are chosen whose background errors are assumed to be uncorrelated, which leads to a simplification of the background term in the cost function. In order to be uncorrleated, it is assumed that each control variable should represent the separate modes of atmospheric motion, ie a part to represent the slow Rossby modes, and other parts to represent the fast gravity modes, etc. The Rossby modes form the balanced component of the flow and are associated fundamentally by the potential vorticity (Gill 1982, Hoskins, McIntyre and Robertson 1985, Cullen 2003). The control variable that represents the Rossby modes is said to be 'balanced'.

In the usual formulation of the CVT method, the rotational wind field takes a prominent role as it is used directly in the mass-wind balance relationship to diagnose the mass field that is in balance with it (Sec. 2). The underlying assumption here is that vorticity is itself a fully balanced variable. In this work, this assumption is challenged, and in Sec. 3 arguments are presented (based on properties of potential vorticity) that vorticity has a component that is unbalanced, and this part should not be used with the mass-wind balance relationship. A new CVT formulation is presented in Sec. 4 that does not rely on the balanced vorticity assumption. The result is a formulation that gives rise to a **B**-matrix which has a degree of flow dependence in a way that is sensitive to the dynamical regimes present in the flow. Diagnostics for the new scheme are shown in Sec. 5 which highlight some of its properties. Conclusions are summarised in Sec. 6.

2 Control variable transforms in VAR

2.1 General principles

The following is a brief outline of the CVT method. A more detailed review of CVTs is given by Bannister (2007). The CVT is a change of variable that simplifies the treatment of the background term in the cost function. In the standard Ide et al. (1997) notation, the cost function, J, in the incremental formulation (Courtier et al. 1994) is the following

$$J(\delta\mathbf{x}) = \frac{1}{2}\delta\mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{x} + \frac{1}{2}\sum_i(\mathbf{y}_i^{\mathrm{o}} - H_i(\mathbf{x}^{\mathrm{b}} + \delta\mathbf{x}))^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{y}_i^{\mathrm{o}} - H_i(\mathbf{x}^{\mathrm{b}} + \delta\mathbf{x})). \tag{1}$$

Equation (1) is minimized with respect to $\delta\mathbf{x}$. In (1), $\mathbf{x}^{\mathrm{b}}$ is the background state, $\mathbf{y}_i^{\mathrm{o}}$ is the observation vector at time step t_i, H_i is the forward model operator (here including forecast model step to time t_i), and $\mathbf{R}_i$ is the observation error covariance matrix. The $t = 0$ state at any iteration is $\mathbf{x} = \mathbf{x}^{\mathrm{b}} + \delta\mathbf{x}$ and the analysis, $\mathbf{x}^{\mathrm{a}}$, follows from the $\delta\mathbf{x}$ that minimizes J. The CVT allows $\delta\mathbf{x}$ to be replaced with a control variable denoted χ (implicitly an incremental quantity so the δ-notation is dropped), which is a vector related to $\delta\mathbf{x}$ by the linear operator $\mathbf{B}_0^{1/2}$

$$\delta\mathbf{x} = \mathbf{B}_0^{1/2}\chi. \tag{2}$$

The square-root form of the CVT allows the background term in the cost function to simplify. Substituting (2) into (1), by noting that $\mathbf{B}$ can be expressed as $\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{T/2}$, and assuming that $\mathbf{B}_0^{1/2} \approx \mathbf{B}^{1/2}$ allows $\mathbf{B}$ to cancel in the background term. The cost function in terms of χ then takes the form

$$J[\chi] = \frac{1}{2}\chi^{\mathrm{T}}\chi + \frac{1}{2}\sum_i(\mathbf{y}_i^{\mathrm{o}} - H_i(\mathbf{x}^{\mathrm{b}} + \mathbf{B}_0^{1/2}\chi)^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{y}_i^{\mathrm{o}} - H_i(\mathbf{x}^{\mathrm{b}} + \mathbf{B}_0^{1/2}\chi). \tag{3}$$

In the χ representation the background error covariance matrix becomes the identity matrix ie, components of background error are uncorrelated and have unit variance. It is the transformed cost function (3) that is minimized with respect to χ in VAR. The unfeasible problem in the $\delta\mathbf{x}$-representation has been transformed into a feasible problem in the χ-representation as long as a square-root of $\mathbf{B}$ (the CVT) can be formulated. The operator $\mathbf{B}_0^{1/2}$ used in (2) is distinct from $\mathbf{B}^{1/2}$ to indicate that it is an approximate square-root. It is an important aim of data assimilation research to formulate a form of $\mathbf{B}_0^{1/2}$ which is close to $\mathbf{B}^{1/2}$. In practice, it is not possible to know the actual (ideally flow dependent) $\mathbf{B}$-matrix, but it is possible to have some idea of what some of its important structure functions look like. It is possible to evaluate a given formulation, $\mathbf{B}_0^{1/2}$ by examining structure functions of the implied error covariance matrix, $\mathbf{B}^{\mathrm{ic}}$, where

$$\mathbf{B}^{\mathrm{ic}} = \mathbf{B}_0^{1/2}\mathbf{B}_0^{T/2}. \tag{4}$$

2.2 The usual formulation of control variable transforms in operational VAR systems

In practice, $\mathbf{B}_0^{1/2}$ is formulated with physical arguments (e.g. that balanced and unbalanced control variables are uncorrelated) and with phenomonological arguments (e.g. that structure functions in the horizontal have a very different structure to those in the vertical, and that vertical structure functions should vary in a special way with latitude and scale). The problem of formulating $\mathbf{B}_0^{1/2}$ starts by considering separately the multivariate and spatial parts of the covariances in the following form, using the notation of Derber and Bouttier (1999)

$$\mathbf{B}_0^{1/2} = \mathbf{K}\mathbf{B}_{\mathrm{u}}^{1/2}, \tag{5}$$

where $\mathbf{K}$ is the parameter transform (or balance operator - not to be confused with the balance relation - see later), and takes care of the multivariate aspects of the problem, and $\mathbf{B}_{\mathrm{u}}^{1/2}$ is the spatial transform, which takes care of the spatial aspects of the problem. Each is described more fully in e.g. Bannister (2007). This paper is concerned with $\mathbf{K}$.

The PV-based CVT (Sec. 4) is tested in the VAR system of the Met Office (Lorenc et al. 2000, Ingleby 2001), and so it is necessary to first review the standard Met Office CVT. In this system the control variable χ has the structure $\chi = (\chi_{\delta\psi}, \chi_{\delta\chi}, \chi_{\delta p_{\mathrm{r}}}, \chi_{\delta\mu})$, which has parts representing fields of streamfunction, velocity potential, residual pressure and relative humidity increments. Each field is not represented in model space, but instead as weights of modes which are designed to be uncorrelated. $\mathbf{B}_{\mathrm{u}}^{1/2}$ is the first operator that acts on χ - see (5) and (2). It acts on each field separately and recovers these fields in the model's longitude, latitude and height representation. Let this intermediate field of variables be $\tilde{\chi} = \mathbf{B}_{\mathrm{u}}^{1/2}\chi = (\delta\tilde{\psi}, \delta\tilde{\chi}, \delta\tilde{p}_{\mathrm{r}}, \delta\tilde{\mu})$. This is the way that the scheme treats the spatial covariances. At the Met Office, the remaining operator $\mathbf{K}$ has the structure

$$\begin{pmatrix} \delta\psi \\ \delta\chi \\ \delta p \\ \delta T \\ \delta q \end{pmatrix} = \begin{pmatrix} \mathbf{I} & 0 & 0 & 0 \\ 0 & \mathbf{I} & 0 & 0 \\ \mathbf{H} & 0 & \mathbf{I} & 0 \\ \mathbf{TH} & 0 & \mathbf{T} & 0 \\ (\Gamma + \mathbf{YT})\mathbf{H} & 0 & \Gamma + \mathbf{YT} & \Gamma \end{pmatrix} \begin{pmatrix} \delta\tilde{\psi} \\ \delta\tilde{\chi} \\ \delta\tilde{p}_{\mathrm{r}} \\ \delta\tilde{\mu} \end{pmatrix}, \tag{6}$$

where the variables on the left hand side are the model variable increments streamfunction, velocity potential, pressure, temperature and specific humidity (the Met Office model actually uses wind components δu and δv instead of $\delta\psi$ and $\delta\chi$, but these are related trivially via the Helmholtz relations; there are model variable increments omitted in (6) - e.g. vertical velocity and density). The operators in (6) are as follows: $\mathbf{H}$ is the linear balance equation which operates on $\delta\tilde{\psi}$ to give the pressure field that is in balance with $\delta\tilde{\psi}$, $\mathbf{T}$ is the hydrostatic operator which calculates temperature from pressure increments, Γ, $\mathbf{Y}$ and Λ relate pressure, temperature and relative humidity increments respectively to specific humidity increments - see Bannister (2007). Line 3 of (6) for instance states that the full pressure increment has two contributions, a part that is related to $\delta\tilde{\psi}$, $\mathbf{H}\delta\tilde{\psi}$, and a residual part, $\delta\tilde{p}_{\mathrm{r}}$, and each contribution is treated as being uncorrelated with the other.

It is the process of recovering the model variables, $\delta\mathbf{x}$, from the uncorrelated fields, $\tilde{\chi}$, with $\mathbf{K}$ that prescribes the multivariate covariances between model variables in VAR. From a dynamical point of view there is a potential flaw with the use of the full rotational wind as a control variable in (6). The arguments, which are presented in the next section, require modification only to the upper left 3×3 matrix of $\mathbf{K}$ in (6), and so the rest of this paper will be concerned only with this part.

3 Shortcomings of the standard, 'vorticity-based' control variable transform

3.1 The balanced vorticity approximation

Scheme (6) is termed vorticity-based because of the assumption that the rotational part of the wind (actually represented not with vorticity but with streamfunction in the Met Office's scheme) is a balanced variable with no allowance of an unbalanced part. In general, mass and wind variables (including vorticity) have balanced and unbalanced components as follows

$$\delta\tilde{\psi} = \delta\tilde{\psi}_{\mathrm{b}} + \delta\tilde{\psi}_{\mathrm{u}}, \tag{7}$$
$$\delta\tilde{p} = \delta\tilde{p}_{\mathrm{b}} + \delta\tilde{p}_{\mathrm{u}}. \tag{8}$$

Substitution of (7) into the third line of (6) gives

$$\delta p = \mathbf{H}\delta\tilde{\psi} + \delta\tilde{p}_{\mathrm{r}} = \mathbf{H}\delta\tilde{\psi}_{\mathrm{b}} + \mathbf{H}\delta\tilde{\psi}_{\mathrm{u}} + \delta\tilde{p}_{\mathrm{r}}. \tag{9}$$

Owing to the second term of the right hand side, (9) is flawed as the linear balance equation is anomalously operating on an unbalanced state. The assumption that this term is negligible in the current scheme is referred to here as the balanced vorticity approximation (BVA). In the PV scheme, the BVA is relaxed, and a more realistic partition of wind and mass into balanced and unbalanced components as in (7)-(8) is performed.

3.2 The unbalanced vorticity

The role of unbalanced vorticity is most easily understood by examining its role in the shallow water equations (Katz et al. 2006). The shallow water equation system is often used as a simple tool to understand the 3-D atmosphere, which may be regarded as a system of shallow water systems with one system per vertical mode. The general principles from this analysis are transferable to 3-D. The perturbation form of the shallow water potential vorticity (PV), δQ is as follows

$$\delta Q = gh\nabla^2\delta\psi - fg\delta h, \tag{10}$$

where g is the acceleration due to gravity, h is the height of the shallow water layer (this is akin to pressure in the 3-D equations) and f is the Coriolis parameter (assumed in this section to be constant). In (10), products of incremental quantities are ignored. Both streamfunction and height increments have balanced and unbalanced parts akin to (7)-(8). The linear balance equation for this system relates balanced mass and balanced wind

$$0 = g\delta h_{\mathrm{b}} - f\delta\psi_{\mathrm{b}}. \tag{11}$$

By (i) noting that unbalanced parts of the fields, $\delta\psi_{\mathrm{u}}$ and δh_{u}, do not contribute to the PV (see below) and (ii) scaling horizontal distance, x, by the characteristic length scale, L_0, $x = L_0\hat{x}$ (giving $\nabla^2 = L_0^{-2}\hat{\nabla}^2$), allows (10) and (11) to be developed as follows

$$\delta Q = gh\nabla^2\delta\psi_{\mathrm{b}} - fg\delta h_{\mathrm{b}} = (gh\nabla^2 - f^2)\delta\psi_{\mathrm{b}} = f^2(\mathrm{Bu}^2\hat{\nabla}^2 - 1)\delta\psi_{\mathrm{b}}. \tag{12}$$

where Bu is the dimensionless Burger number

$$\mathrm{Bu} = \frac{\sqrt{gh}}{fL_0} = \frac{L_R}{L_0}. \tag{13}$$

The quantity $\sqrt{gh}/f$ is the Rossby radius of deformation, L_R, which allows Bu to be interpreted as the ratio of L_R to L_0 in (13). In the 3-D quasi-geostrophic system, $\mathrm{Bu} = NH/fL_0$, where N is the static stability and H is the characteristic vertical lengthscale.

Since $\hat{\nabla}^2$ is $O(1)$ when acting on $\delta\psi_{\mathrm{b}}$ in (12), the relative importance of the wind part (first term) and the mass part (second term) of the PV in (12) is controlled by Bu. Two limiting regimes are shown in Fig. 1 as follows (Wlasak et al. 2006).

- For $\mathrm{Bu} \gg 1$ (where $L_0 \ll L_R$, achieved e.g. when h is large and f is small), the PV does not depend on the mass term in (12) but is controlled entirely by the balanced rotational wind. In the absence of the mass term, the balanced rotational wind can take the value of the total rotational wind, resulting in $\delta\psi_{\mathrm{u}} = 0$ (top left corner of Fig. 1). In this regime the BVA is a good approximation.
- For $\mathrm{Bu} \ll 1$ (where $L_0 \gg L_R$, achieved e.g. when h is small), the PV does not depend on the wind term in (12) and is controlled entirely by the balanced mass field. In the absence of the wind term, the balanced mass can take the value of the total mass field, resulting in $\delta h_{\mathrm{u}} = 0$ (bottom right corner of Fig. 1). In this regime there is no reason to expect that the unbalanced wind is negligible and so the BVA (and hence the standard VAR scheme) is not expected to be a good approximation.

In the rest of this paper, a new PV-based formulation of the CVT is introduced and tested which should not suffer the shortcomings highlighted in (9), as it recognises the presence of an unbalanced vorticity component, and therefore applies the balance relationship, **H**, more appropriately. The PV-scheme is introduced in Sec. 4. In any scheme, the control variables (e.g. $\delta\tilde{\psi}$ and $\delta\tilde{p}_{\mathrm{r}}$ in the standard scheme) are assumed to be mutually

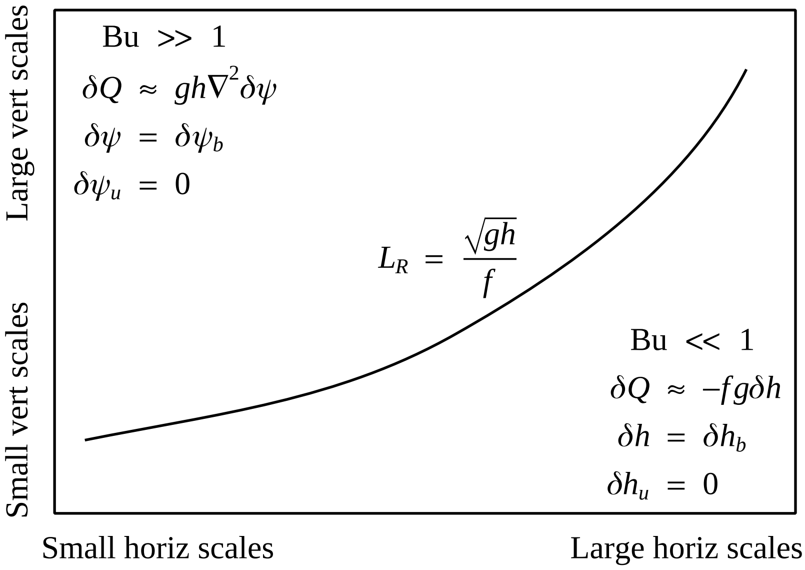

Figure 1: Regimes of the shallow water equations. The Rossby radius, L_R (curved line), separates the regimes into Bu > 1 *(left of the curve) and* Bu < 1 *(right of the curve). In the extreme cases shown, the PV is described only by the rotational wind for* Bu $\gg 1$ *and mass for* Bu $\ll 1$.

uncorrelated. This property is rarely strictly true in reality, but a successful scheme should result in new control variables that have only weak correlations. Even though the PV-based scheme is expected to be better on dynamical grounds than the standard scheme, its new control variables must be tested for their degree of correlation. These, and other diagnostics are shown in Sec. 5.

4 New control variables and transforms based on potential vorticity

The equations are developed for the PV-based scheme for the 3-D atmosphere. The new scheme must

- recognise the presence of an unbalanced vorticity,
- apply mass-wind balance relationships only to balanced variables,
- allow the system to adjust to the dynamical regime so that it resembles the existing scheme at high Bu, but not at low Bu, and
- use control variables that are only weakly correlated, and with less correlation than the existing variables.

4.1 Basic equations

The scheme centres on two equations. The first is an approximate perturbation form of Ertel PV. By ignoring products of perturbations and horizontal components of vorticity, and assuming small Rossby number, PV has the form

$$\alpha_0 \nabla_h^2 \delta\psi + \beta_0 \delta p + \gamma_0 \frac{\partial \delta p}{\partial z} + \varepsilon_0 \frac{\partial^2 \delta p}{\partial z^2} = \delta Q, \tag{14}$$

where α_0, β_0, γ_0 and ε_0 are reference state values (actually in this work they are the zonal mean of the linearisation state), ∇_h is the horizontal Laplacian and z is height. These elements introduce flow dependence into the system. A derivation of this PV is given in Bannister and Cullen (2006). This PV is assumed to have the same qualitative properties as the shallow water PV (10) in terms of the prominence of mass and wind variables in different regimes. The second equation is the linear balance equation (LBE) which relates the balanced parts of mass and wind

$$0 = \nabla_h \cdot (f\rho_0 \nabla_h \delta\psi_{\mathrm{b}}) - \nabla_h^2 \delta p_{\mathrm{b}}, \tag{15}$$

where ρ_0 is the reference state density and f now is variable.

4.2 Equations for the balanced and unbalanced components

Equations (14)-(15) are used to formulate a set of equations to relate new control variables to model variables, and the reverse. Equation (14) can be used to derive a diagnostic equation between the unbalanced mass and wind variables. The unbalanced components have no PV

$$\alpha_0 \nabla_h^2 \delta\psi_{\mathrm{u}} + \beta_0 \delta p_{\mathrm{u}} + \gamma_0 \frac{\partial \delta p_{\mathrm{u}}}{\partial z} + \varepsilon_0 \frac{\partial^2 \delta p_{\mathrm{u}}}{\partial z^2} = 0, \tag{16}$$

which may be regarded as the anti-balance analogue to (15). Substituting (7)-(8) into (14) (ignoring for now the tildes), and using (16) gives

$$\alpha_0 \nabla_h^2 \delta\psi_{\mathrm{b}} + \beta_0 \delta p_{\mathrm{b}} + \gamma_0 \frac{\partial \delta p_{\mathrm{b}}}{\partial z} + \varepsilon_0 \frac{\partial^2 \delta p_{\mathrm{b}}}{\partial z^2} = \delta Q, \tag{17}$$

which states that only balanced variables can influence PV. The unbalanced components cannot obey the LBE, but substituting $\delta\psi_{\mathrm{b}}$ with $\delta\psi_{\mathrm{u}}$ and δp_{b} with δp_{u} in (15) gives a residual, which is called the anti-PV or linear imbalance, $\delta\bar{Q}$

$$\nabla_h \cdot (f\rho_0 \nabla_h \delta\psi_{\mathrm{u}}) - \nabla_h^2 \delta p_{\mathrm{u}} = \delta\bar{Q}. \tag{18}$$

Equations (15)-(18) are two pairs of equations, each associated with either the balanced or the unbalanced components of the flow. The variables $\delta\tilde{\psi}_{\mathrm{b}}$ and $\delta\tilde{p}_{\mathrm{u}}$ (now with tildes to indicate that they are members of the new vector $\tilde{\chi}$) replace $\delta\tilde{\psi}$ and $\delta\tilde{p}_{\mathrm{r}}$ of the standard scheme as balanced and unbalanced control variables in VAR. The velocity potential variable, $\delta\tilde{\chi}$ remains the same. These particular variables have been chosen as they represent the simplest change to the current system, they lead to a relatively efficient **K**-operator and are analogous to those used by Cullen (2003) as implemented in the ECMWF's VAR system (the way of generating the equations is also similar).

Although Cullen's trials were successful, his scheme has some drawbacks. Firstly his equations are posed in spectral representation. Although this allows many of the equations to be solved relatively easily, it has the consequence that the reference state quantities cannot be prescribed as a function of latitude. The scheme presented here for the Met Office's VAR system allows latitude dependence of reference state quantities. Secondly, Cullen's scheme is tied to the ECMWF grid, which has Lorenz grid staggering in the vertical. The equations to be solved are ill-conditioned in their own right, but the Lorenz grid exaggerates this effect. The Met Office uses the Charney-Phillips grid staggering in the vertical which is not expected to suffer the same problems.

4.3 Control-to-model variables

Based on (15)-(16), the PV-based $\mathbf{K}$ in (5) is as follows (for the three dynamical variables only)

$$\begin{pmatrix} \delta\psi \\ \delta\chi \\ \delta p \end{pmatrix} = \begin{pmatrix} \mathbf{I} & 0 & \bar{\mathbf{H}} \\ 0 & \mathbf{I} & 0 \\ \mathbf{H} & 0 & \mathbf{I} \end{pmatrix} \begin{pmatrix} \delta\tilde{\psi}_{\mathrm{b}} \\ \delta\tilde{\chi} \\ \delta\tilde{p}_{\mathrm{u}} \end{pmatrix}, \tag{19}$$

This should be compared to the part of (6) concerned with dynamical variables. In (19), $\mathbf{H}$ is the matrix representation of the LBE operator from (15) which has input $\delta\tilde{\psi}_{\mathrm{b}}$ and output $\delta\tilde{p}_{\mathrm{b}}$, and $\bar{\mathbf{H}}$ is the matrix representation of the anti-balance operator from (16) which has input $\delta\tilde{p}_{\mathrm{u}}$ and output $\delta\tilde{\psi}_{\mathrm{u}}$. The key differences with (6) are (i) the presence of the $\bar{\mathbf{H}}$-operator and (ii) the changes of the first and third control variables. This scheme recognises the existence of an unbalanced vorticity and applies the LBE only to the balanced variable $\delta\tilde{\psi}_{\mathrm{b}}$, instead of $\delta\tilde{\psi}$. The operator $\bar{\mathbf{H}}$ is flow dependent due to the presence of state-dependent matrices present in (16). The way that these operators are expected to work is along the lines of the arguments presented in Sec. 3 for the shallow water system (that $\delta\tilde{\psi}_{\mathrm{b}} = \delta\tilde{\psi}$ for Bu $\gg 1$, and $\delta\tilde{p}_{\mathrm{u}} = 0$ for Bu $\ll 1$).

4.4 Model-to-control variables

The most frequently used CVT is the control-to-model part shown in Sec. 4.3, which is needed to evaluate the cost function (3) and its gradient (not shown). The inverse transform (model-to-control) is needed however before the assimilation to derive background error statistics for each new variable in order to determine parameters in the spatial transform, $\mathbf{B}_{\mathrm{u}}^{1/2}$ in (5). The inverse transform derives a population of control variable $\delta\tilde{\psi}_{\mathrm{b}}$, $\delta\tilde{\chi}$ and $\delta\tilde{p}_{\mathrm{u}}$ from a population of model fields $\delta\psi$, $\delta\chi$ and δp. The equations that need to be solved to achieve this are discussed briefly here (the process of modelling spatial error covariances is beyond the scope of this paper).

The equation for $\delta\tilde{\psi}_{\mathrm{b}}$ is found by eliminating $\delta\tilde{p}_{\mathrm{b}}$ between (15) and (17)

$$\begin{aligned} \alpha_0\nabla_h^2\delta\tilde{\psi}_{\mathrm{b}} + \beta_0\nabla_h^{-2}[\nabla_h\cdot(f\rho_0\nabla_h\delta\tilde{\psi}_{\mathrm{b}})] + \quad & \gamma_0\tfrac{\partial}{\partial z}\nabla_h^{-2}[\nabla_h\cdot(f\rho_0\nabla_h\delta\tilde{\psi}_{\mathrm{b}})] + \\ & \varepsilon_0\tfrac{\partial^2}{\partial z^2}\nabla_h^{-2}[\nabla_h\cdot(f\rho_0\nabla_h\delta\tilde{\psi}_{\mathrm{b}}] = \delta Q, \end{aligned} \tag{20}$$

where tildes have been added to the control fields to distinguish them from model fields. The PV, δQ, is calculated from the model fields via (14). The equation for $\delta\tilde{\chi}$ from $\delta\chi$ is trivial (in practice the model variables are δu and δv instead of $\delta\chi$, which involves solution of a Poisson equation). The equation for $\delta\tilde{p}_{\mathrm{u}}$ is found by eliminating $\delta\tilde{\psi}_{\mathrm{u}}$ between (16) and (18)

$$\nabla_h\cdot\left(f\rho_0\nabla_h\nabla_h^{-2}\left[-\alpha_0^{-1}\beta_0\delta\tilde{p}_{\mathrm{u}} - \alpha_0^{-1}\gamma_0\frac{\partial\delta\tilde{p}_{\mathrm{u}}}{\partial z} - \alpha_0^{-1}\varepsilon_0\frac{\partial^2\delta\tilde{p}_{\mathrm{u}}}{\partial z^2}\right]\right) - \nabla_h^2\delta\tilde{p}_{\mathrm{u}} = \delta\bar{Q}. \tag{21}$$

The anti-PV, $\delta\bar{Q}$, is calculated via (18) by substituting-in the model fields $\delta\psi_{\mathrm{u}} \to \delta\psi$ and $\delta p_{\mathrm{u}} \to \delta p$ (the equation with these substitutions is equivalent to (18) because the LBE (15) eliminates the balanced contributions to $\delta\bar{Q}$). Equations (20)-(21) give a set of control variable increments from a set of model variable increments. The latter set are derived from differences between forecasts of different lengths, which are meant to have the same properties of forecast errors under the approximation of the NMC method (Parrish and Derber 1992, Bouttier 1996, Berre et al. 2006).

Equations (20)-(21) are treated as 3-D elliptic problems which are solved approximately using the generalised conjugate residual (GCR) method with appropriate boundary conditions and preconditioning. Both (20) and (21) have the form $\delta\mathbf{y} = \mathbf{C}\delta\mathbf{x}$. The GCR works by solving a variational problem for $\delta\mathbf{x}$ which minimizes the

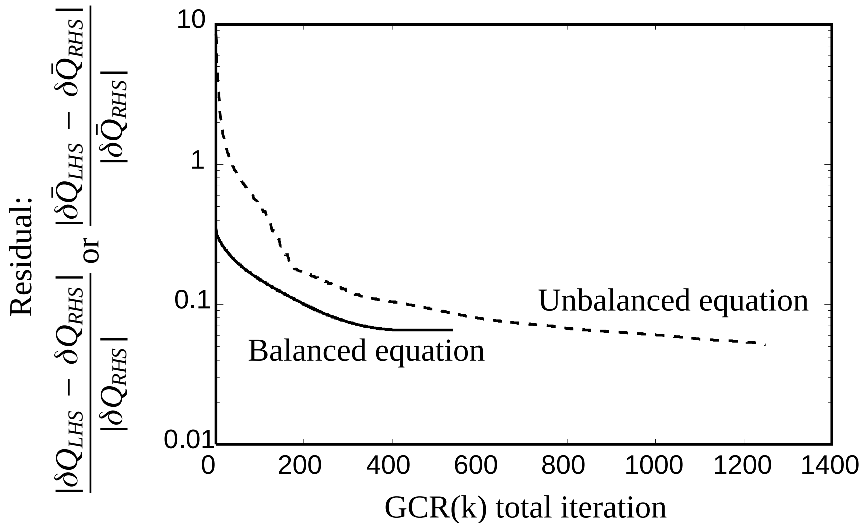

Figure 2: Residuals (relative to $|\delta Q|_2$ and $|\delta\bar{Q}|_2$ for the balanced and unbalanced equations respectively.

residual $|\delta \mathbf{y} - \mathbf{C}\delta\mathbf{x}|_2^2$ where $||_2$ represents an L_2 norm. The residuals of (20)-(21) after each iteration of the GCR minimization for an example case are shown in Fig. 2. Equations (20)-(21) are believed to have a large condition number and the residuals remain high even after a large number of iterations. Even though these equations have not been solved to a very high accuracy, they are still used in this study.

5 Diagnostics of the standard and PV-based schemes

This section contains some preliminary diagnostics from trials of the PV-based and the standard BVA-based VAR. Only minimal discussion is given here as a more detailed paper will follow.

5.1 Correlations of control variables

The point-by-point correlations $\mathrm{cor}(\delta\tilde{\psi}, \delta\tilde{p}_{\mathrm{r}})$ under the BVA and $\mathrm{cor}(\delta\tilde{\psi}_{\mathrm{b}}, \delta\tilde{p}_{\mathrm{u}})$ under the PV scheme are plotted in Fig. 3 as a function of latitude and level (correlations have been averaged zonally), found from a population of six cases. VAR assumes zero correlations, but both plots show significant values (maxima of ± 0.9), although the PV-based set has smaller correlations overall (rms of BVA is 0.349, rms of PV is 0.255), indicating that the latter is a more suitable choice for uncorrelated control variables.

5.2 Lengthscales of control variables

The predicted properties of the control variables as summarised in Fig. 1 can be examined for real model data by looking at the correlation lengthscales. Figure 4 shows vertical correlations for the variables $\delta\tilde{\psi}$ and $\delta\tilde{p}_{\mathrm{r}}$ of the standard BVA scheme (left panels), and $\delta\tilde{\psi}_{\mathrm{b}}$ and $\delta\tilde{p}_{\mathrm{u}}$ of the PV scheme (right panels) as a function of wavenumber. The vertical structures are broader for the PV scheme than for the BVA scheme, particularly at large horizontal scales (small wavenumbers). This is consistent with the assessment of Fig. 1 that at large horizontal scales: (i) $\delta\tilde{\psi}_{\mathrm{b}} < \delta\tilde{\psi}$ in magnitude and (ii) $\delta\tilde{p}_{\mathrm{u}}$ is small unless in each case the vertical scale is large.

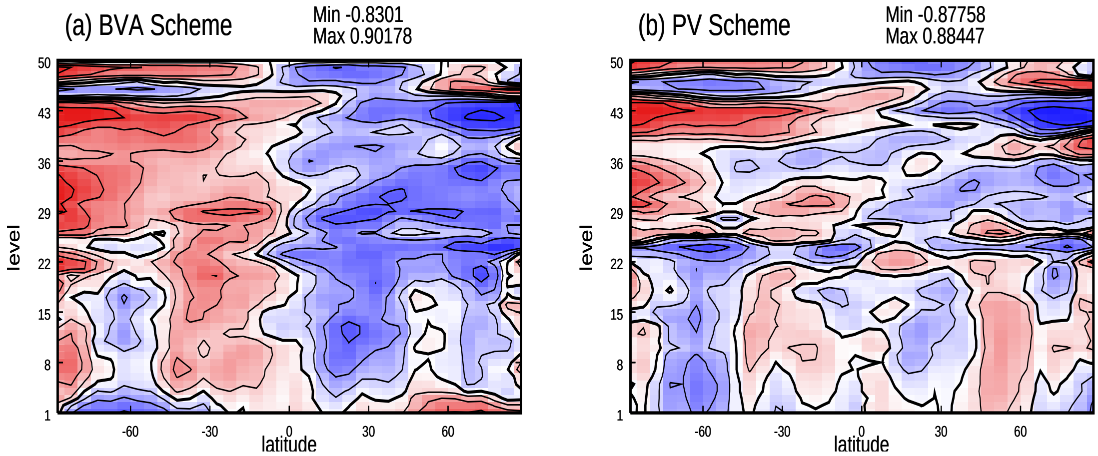

Figure 3: Correlations for the BVA scheme, $\mathrm{cor}(\delta\tilde{\psi},\delta\tilde{p}_{\mathrm{r}})$ *(panel a) and for the PV scheme,* $\mathrm{cor}(\delta\tilde{\psi}_{\mathrm{b}},\delta\tilde{p}_{\mathrm{u}})$ *(panel b).*

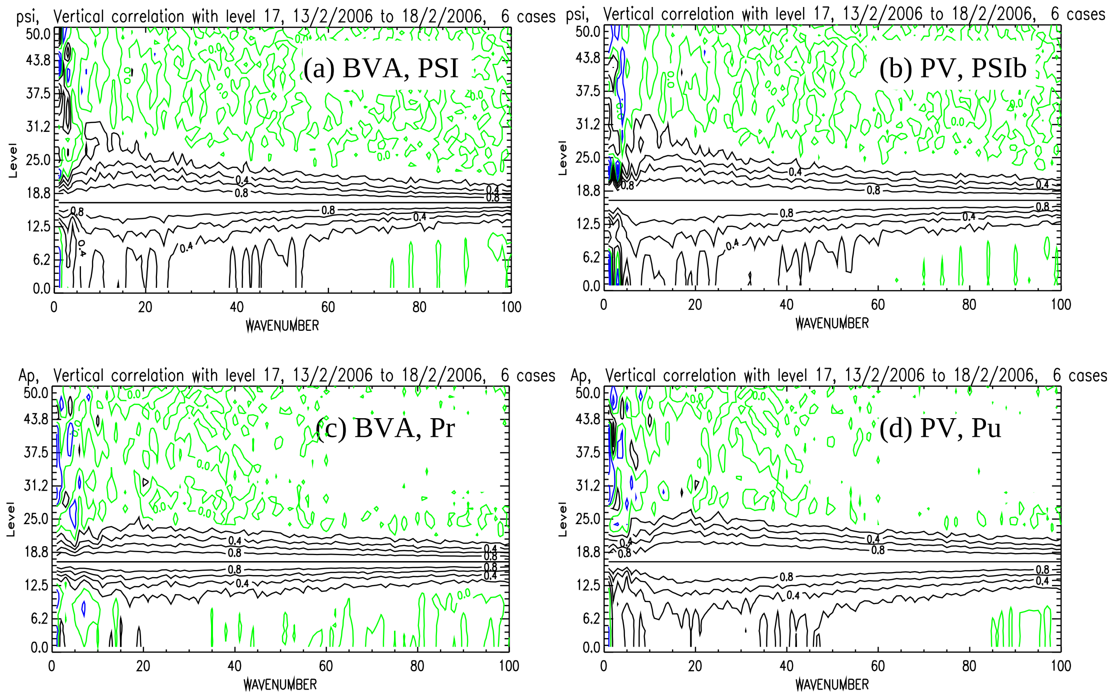

Figure 4: Vertical correlations of control variables with level 17 (≈ 500 hPa) as a function of wavenumber. The top panels are for the rotational wind control variables and the bottom panels are for the mass control variable (left: BVA, right: PV).

5.3 Pseudo observation tests

The PV-scheme has been run with a small number of synthetic observations. The analysis increments that are produced are indicative of the structure functions that are implied by the error covariance formulation. Figure 5 comprises cross sections of the analysis increments of pressure (left panels), zonal wind (middle panels) and meridional wind (right panels) from an experiment that assimilates a number of point pressure observations (the effect of one observation is visible in the cross sections shown). The top panels are for the BVA scheme which show the region of influence of the pressure observation and the geostrophic dipole patterns of the wind response. The bottom panels are for the PV-based scheme which show a pressure response similar to the BVA (but with a narrower vertical scale). The wind response is unsatisfactory however and it is dominated by large horizontal and small vertical scales. These winds are primarily unbalanced (the balanced parts - not shown - are similar to the BVA response) and are due to amplification at these scales. From (16), the unbalanced streamfunction is, from the $\bar{\mathbf{H}}$ operator

$$\delta\tilde{\psi}_{\mathrm{u}} = \nabla_h^{-2}\left[-\alpha_0^{-1}\beta_0\delta\tilde{p}_{\mathrm{u}} - \alpha_0^{-1}\gamma_0\frac{\partial\delta\tilde{p}_{\mathrm{u}}}{\partial z} - \alpha_0^{-1}\varepsilon_0\frac{\partial^2\delta\tilde{p}_{\mathrm{u}}}{\partial z^2}\right] = \bar{\mathbf{H}}\delta\tilde{p}_{\mathrm{u}}. \tag{22}$$

This operator amplifies large horizontal and small vertical scales. The results in Fig. 5 are clearly anomalous and following possible explanations are being considered.

- The variances of $\delta\tilde{p}_{\mathrm{u}}$ in the training set may be too large at large horizontal and small vertical scales due to problems of removing the residuals from the GCR solver (Fig. 2). This is a symptom that the $\mathbf{K}^{-1}$ operator defined by the GCR solution of (20)-(21) is not the exact inverse of $\mathbf{K}$ operator defined by (19). A fix might be to damp-out variance in $\delta\tilde{p}_{\mathrm{u}}$ at these scales. Note that a similar correction is used in the BVA scheme at the Met Office as a result of the operation $\mathbf{H}\delta\tilde{\psi}$. Since $\delta\tilde{\psi}$ contains unbalanced parts at small vertical scales, these scales are removed from $\mathbf{H}\delta\tilde{\psi}$ with vertical regression.

- A second order auto-regressive function (SOAR) is used to model the horizontal variance spectra for each independent vertical mode (performed as part of $\mathbf{B}_{\mathrm{u}}^{1/2}$). The horizontal scale of this function may be too large. A solution might be to replace the SOAR with the actual spectrum found from the training set.

5.4 Summary

A new formulation to model the multivariate aspects of forecast error covariances for data assimilation, based on the properties of potential vorticity (PV) is formulated and tested in the Met Office's VAR scheme. The standard scheme makes the balanced vorticity approximation (BVA), which assumes that the rotational wind is universally a balanced variable. The PV scheme relaxes this assumption and allows a contribution to the rotational wind that is unbalanced in a way that is consistent with the flow regime.

In outline, the scheme looks similar to the standard scheme, as the standard control variables of streamfunction and residual pressure are replaced with the balanced component of streamfunction and unbalanced pressure, but is more difficult to apply owing to the three-dimensional elliptic equations that need to be solved to determine the error statistics of the new variables, and the over sensitivity of the unbalanced wind response in the regime of large horizontal and small vertical scales.

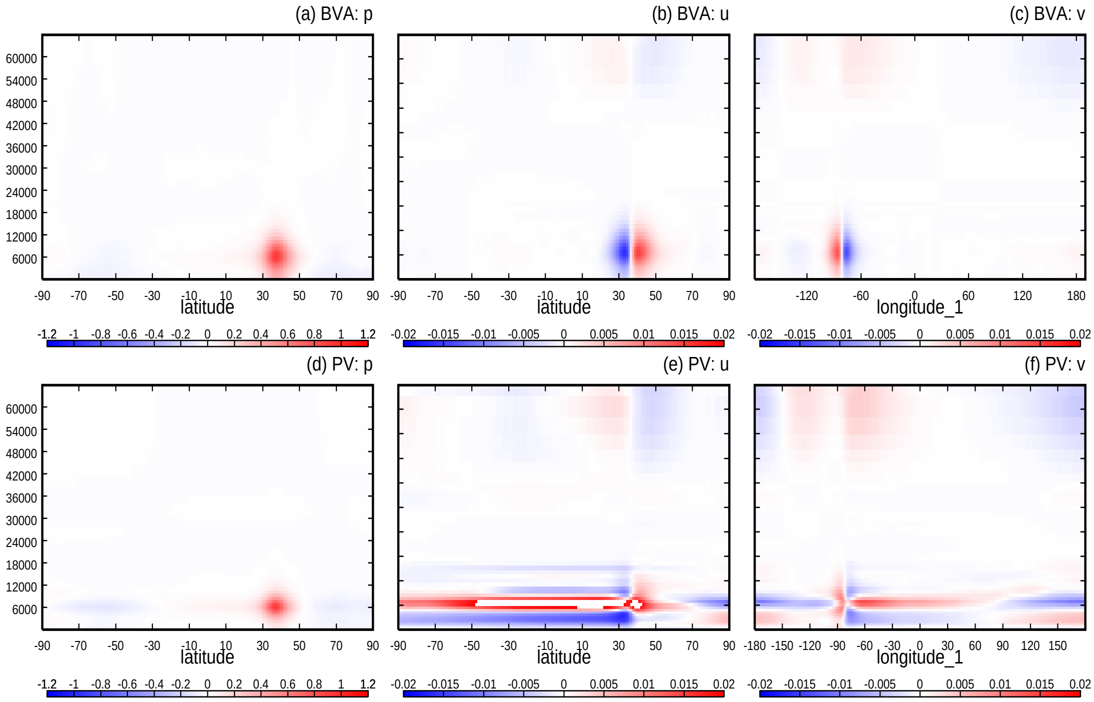

Figure 5: Analysis increments of pressure and zonal and meridional winds from assimilation tests with pseudo observations of pressure. The top row is for the BVA scheme and the bottom row is for the PV scheme.

Acknowledgements

Thanks to Paul Berrisford, Mark Dixon, Dingmin Li, David Pearson, Ian Roulstone and Marek Wlasak for scientific or technical discussions. This work is funded by NERC and the Met Office.

References

Bannister R.N., M.J.P. Cullen (2006). The implementation of a PV-based control variable in variational data assimilation. Part I: Transforms, DARC Report 6. Available from *www.met.rdg.ac.uk/ ross/DARC/DataAssim.html*.

Bannister R.N. (2007). A review of forecast error covariance statistics in atmospheric variational data assimilation. II: modelling the error covariance statistics for numerical weather prediction. To be submitted for publication, available soon from *www.met.rdg.ac.uk/ ross/DARC/DataAssim.html*.

Barker D.M. and A.C. Lorenc (2005). The use of synoptically-dependent background error structures in 3dVar., Var. Scientific Documentation Paper 26, Met Office, 11pp. Available from the Met Office.

Berre L., S.E. Stefanescu and M. Belo (2006). The representation of the analysis effect in three error simulation techniques, Tellus, 58A, 196-209.

Bouttier F. (1996). Application of Kalman filtering to numerical weather prediction, Proceedings of the 1996 ECMWF seminar on data assimilation, pp. 61-90. Available from *www.ecmwf.int*.

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.

Cullen M.J.P. (2003). Four-dimensional variational data assimilation: A new formulation of the background-error covariance matrix based on a potential-vorticity representation, Q.J.R. Meteorol. Soc. 129, 2777-2796.

Desroziers G. (1997). A coordinate change for data assimilation in spherical geometry of frontal structures, Mon. Wea. Rev., 125, 3030-3038.

Derber J. and F. Bouttier (1999). A reformulation of the background error covariance in the ECMWF global data assimilation system, Tellus, 51A, 195-221.

Fisher M. and P. Courtier (1995). Estimating the covariance matrix of analysis and forecast error in variational data assimilation, ECMWF Tech. Memo. 220, 26 pp. Available from *www.ecmwf.int*.

Fisher M. (1998). Development of a simplified Kalman filter, ECMWF Tech Memo 260, 16 pp. Available from *www.ecmwf.int*.

Fisher M. (2003). Background error covariance modelling, Proc. ECMWF seminar on recent developments in data assimilation for atmosphere and ocean, 8-12 Sept., 45-64. Available from *www.ecmwf.int*.

Gill A.E. (1982). Atmosphere-Ocean Dynamics, Vol. 30. Academic Press, San Diego.

Hoskins B.J., M.E. McIntyre and A.W. Robertson (1985). On the use and significance of isentropic potential vorticity maps, Q.J.R. Meteorol. Soc., 111, 877-946.

Ide K., P. Courtier, M. Ghil and A.C. Lorenc (1997). Unified notation for data assimilation: operational, sequential and variational, J. Meteor. Soc. Japan, 75, No. 1B, 181-189.

Ingleby N.B. (2001). The statistical structure of forecast errors and its representation in the Met Office global 3-dimensional variational data assimilation system, Q.J.R. Meteorol. Soc., 127, 209-231.

Katz D., A.S. Lawless, N.K. Nichols, M.J.P. Cullen and R.N. Bannister (2006). A comparison of potential vorticity-based and vorticity-based control variables, Dept. of Mathematics, Univ. of Reading, Numerical Analysis Report, 34 pp. Available from *www.maths.rdg.ac.uk/Research/Publications/Na_reports.asp*.

Lorenc A.C., S.P. Ballard, R.S. Bell, N.B. Ingleby, P.L.F. Andrews, D.M. Barker, J.R. Bray, A.M. Clayton, T. Dalby, D. Li, T.J. Payne and F.W. Saunders (2000). The Met Office global 3-dimensional variational data assimilation scheme, Q.J.R. Meteorol. Soc., 126, 2991-3012.

Parrish D.F. and J.C. Derber (1992). The National Meteorological Center's spectral statistical interpolation analysis system, Mon. Wea. Rev., 120, 1747-1763.

Wlasak M.A., N.K. Nichols NK and I. Roulstone (2006). Use of potential vorticity for incremental data assimilation, Dept. of Mathematics, Univ. of Reading, Numerical Analysis Report, 20 pp. Available from *www.maths.rdg.ac.uk/Research/Publications/Na_reports.asp*.