Journal of the Meteorological Society of Japan, Vol. 75, No. 1B, pp. 339–346, 1997

Development of an Operational Variational Assimilation Scheme

By Andrew C. Lorenc

Meteorological Office, London Road, Bracknell, RG12 2SZ, England.

(Manuscript received 30 May 1995, in revised form 28 November 1995)

Abstract

The United Kingdom Meteorological Office is developing a variational assimilation for its Unified Model forecast system, which contains a grid-point model, run operationally in global, regional, and mesoscale configurations. Key characteristics of the design are:

- development path from 3-dimensional to 4-dimensional scheme
- global and limited area configurations
- variational analysis of perturbations
- carefully designed, well conditioned 'background' term

This paper describes the variational scheme, with some example results from a simple 2-dimensional variational analysis which has been developed as a prototype.

1. Introduction

Over the past few years it has become apparent that variational assimilation schemes could be made practicable (*e.g.* see Courtier *et al.* 1993), and that possibly they might make a significant improvement in forecast quality:

1. in the extraction of useful information from satellite radiances, by three-dimensional retrieval (Andersson *et al.*, 1993).
2. in diagnosing dynamically consistent baroclinic structures, given observations that a system is developing (Thepaut *et al.* 1993).
3. in using observations affected by 'physical' atmospheric processes which are represented in the forecast model.

Most of the benefit from 1 might be realised from a static three-dimensional variational (3DVAR) system, while 2 and 3 probably need a four-dimensional (4DVAR) system containing a forecast model and its adjoint.

The bulk of the effort in developing a practical assimilation scheme goes in careful design and testing, and attention to detail in the observation processing. Currently the Met Office has a project to do this work, building a practical variational assimilation facility for the Met Office's Unified Model system, which contains a grid-point model, run operationally in global, regional, and mesoscale configurations. The project's targets are to match the current operational system, and to make possible the developments outlined above. We plan to test the basic 3DVAR system during 1996, followed by implementation in all operational configurations[1]. This will facilitate the developments necessary to get the benefits mentioned above; we might have a feasible 4DVAR scheme by 1997.

As a prototype for critical aspects of this development, a simple 2-dimensional variational analysis has been developed (2DVAR). Aspects studied include:

- The use of a filter and its adjoint for calculating the background penalty and its gradient (digital and spectral filters have been tried),
- Preconditioning using the filter,
- non-Gaussian observational errors.

This paper describes the full variational scheme, using illustrations from the 2DVAR when appropriate.

2. Variational analysis

The 'standard' formulation of variational analysis (Lorenc 1986) is — find the model state $\mathbf{x}$ which minimises a penalty J, made up from a background

1 We estimate that approximately 30 man-years effort will have been used by this stage.

term J^{b} and an observational term J^{o}:

$$J(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}^{\mathrm{b}} - \mathbf{x}\right)^{\mathrm{T}} \mathbf{B}^{-1}\left(\mathbf{x}^{\mathrm{b}} - \mathbf{x}\right) + \frac{1}{2}\left(\mathbf{y}^{\mathrm{o}} - \mathbf{y}\right)^{\mathrm{T}} (\mathbf{E}+\mathbf{F})^{-1}\left(\mathbf{y}^{\mathrm{o}} - \mathbf{y}\right) \quad (1)$$

where $\mathbf{x}^{\mathrm{b}}$ is a prior (background) estimate of $\mathbf{x}$, with error covariance $\mathbf{B}$, $\mathbf{y}^{\mathrm{o}}$ is a vector of observed values, with instrumental error covariance $\mathbf{E}$, and $\mathbf{y}$ is a prediction of the observed values, given by:

$$\mathbf{y} = H(\mathbf{x}) \quad (2)$$

$\mathbf{F}$ is the error covariance in the 'generalised interpolation' H, which in our 2DVAR examples is a simple interpolation, but which in 4DVAR includes an NWP forecast model. For the practical solution of this problem we make two transformations; to increments, and to a preconditioned control variable.

3. Analysis of increments

Following Courtier *et al.* (1994), we allow for a guess solution $\mathbf{x}^{\mathrm{g}}$, and solve for a model perturbation $\mathbf{w}'$, which may be at lower resolution than $\mathbf{x}$. That is, we find the perturbation model state $\mathbf{w}'$ which minimises:

$$J(\mathbf{w}') = \frac{1}{2}\left(\mathbf{w}'^{\mathrm{b}} - \mathbf{w}'\right)^{\mathrm{T}} \mathbf{B}^{-1}\left(\mathbf{w}'^{\mathrm{b}} - \mathbf{w}'\right) + \frac{1}{2}\left(\mathbf{y}^{\mathrm{o}} - \mathbf{y}\right)^{\mathrm{T}} (\mathbf{E}+\mathbf{F})^{-1}\left(\mathbf{y}^{\mathrm{o}} - \mathbf{y}\right) \quad (3)$$

where we use interpolation S to transform (simplify) the background $\mathbf{x}^{\mathrm{b}}$, and the guess $\mathbf{x}^{\mathrm{g}}$, to the lower resolution of $\mathbf{w}'$:

$$\mathbf{w}'^{\mathrm{b}} = \mathbf{w}^{\mathrm{b}} - \mathbf{w}^{\mathrm{g}} = S(\mathbf{x}^{\mathrm{b}}) - S(\mathbf{x}^{\mathrm{g}}) \quad (4)$$

$\mathbf{w}^{\mathrm{g}}$ is also used as linearisation state for forecasting, and other manipulations to, the perturbations $\mathbf{w}'$. $\mathbf{y}$, the prediction of the observed values, is a function of the guess (which may itself be iterated in an outer-loop), the linearisation state $\mathbf{w}^{\mathrm{g}}$, and the perturbation $\mathbf{w}'$ calculated each iteration in the inner-loop, as explained in more detail in Section 5:

$$\mathbf{y} = \tilde{H}(\mathbf{x}^{\mathrm{g}}, \mathbf{w}^{\mathrm{g}}, \mathbf{w}') \quad (5)$$

This transformation to a variational problem in $\mathbf{w}'$ is based on the belief that $\mathbf{x}^{\mathrm{g}} + \mathbf{S}^{-\mathrm{I}}\mathbf{w}'$ will be a good approximation to the $\mathbf{x}$ which minimises (1). $\mathbf{S}^{-\mathrm{I}}$ is the generalised inverse of the linearisation of S; it transforms from the low resolution of $\mathbf{w}'$ to that of $\mathbf{x}$. It is possible to iterate this correction process for $\mathbf{x}$, outside of the minimisation iteration which finds $\mathbf{w}'$. The updated $\mathbf{x}^{\mathrm{g}} + \mathbf{S}^{-\mathrm{I}}\mathbf{w}'$ is used as guess in a new incremental variational analysis.

Other groups working on 4DVAR have started from a full-fields approach, which needs the adjoint of the linearisation of the full model about its four-dimensional trajectory — usually called the tangent-linear model. A tangent-linear model is derived by differentiating the equations used in the full model. For a model with full physical parametrisations this is a complex task, requiring a large coding effort, or automatic differentiation software. Our approach is different. The first-guess estimate of the atmosphere's four-dimensional trajectory is going to differ from the truth by a finite amount, with a spread governed by the background error variance. So we design a perturbation forecast model which gives an approximation to the evolution of finite perturbations. For instance if, in the trajectory of the full model, it is not raining, but nearly saturated, then some finite perturbations will be such as to make it rain. Thus the perturbation forecast model, designed for the best average of all perturbations, should allow for some latent heating. Threshold processes, which can lead to difficulties in differentiating in the tangent-linear approach, should be smoothed in the perturbation forecast approach. The perturbation forecast model is designed from physical principles, and can have different resolution and algorithms from the full model; it is *not* its tangent-linear model. (For computational efficiency, in order to have a linearisation state which does not change each iteration, we do choose to make the perturbation forecast model linear). We use it, and its adjoint, to find the increment which most reduces the misfit to observations (show as $\mathbf{y}$ in Fig. 1) and the background. This is done using an iterative descent algorithm; the process is shown using solid arrows in Fig. 1. Adding it, we make a new full resolution four-dimensional trajectory (show using dotted arrows in Fig. 1), and can then repeat the inner incremental variational step.

4. Pre-conditioned control variable

Secondly, we transform to a variable $\mathbf{v}$ designed to improve the conditioning of the Hessian matrix in the minimisation process. The Hessian is a matrix of second-order partial derivatives with respect to the control variables. *e.g.* for (1) the Hessian is defined as:

$$\left(\frac{\partial^2 J}{\partial \mathbf{x}^2}\right) = \begin{pmatrix} \frac{\partial^2 J}{\partial x_1 \partial x_1} & \frac{\partial^2 J}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 J}{\partial x_1 \partial x_{n_x}} \\ \frac{\partial^2 J}{\partial x_2 \partial x_1} & \frac{\partial^2 J}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 J}{\partial x_2 \partial x_{n_x}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 J}{\partial x_{n_x} \partial x_1} & \frac{\partial^2 J}{\partial x_{n_x} \partial x_2} & \cdots & \frac{\partial^2 J}{\partial x_{n_x} \partial x_{n_x}} \end{pmatrix} \quad (6)$$

For (3), ignoring derivatives of $\tilde{\mathbf{H}}$ (defined in Section 5), the Hessian is given by

$$\left(\frac{\partial^2 J}{\partial \mathbf{w}'^{\,2}}\right) = \mathbf{B}_{(w)}^{-1} + \tilde{\mathbf{H}}^{\mathrm{T}} (\mathbf{E}+\mathbf{F})^{-1} \tilde{\mathbf{H}} \quad (7)$$

The generalised interpolation $\tilde{\mathbf{H}}$ in the second term in (7) depends on the positions of the observations being used. It is hard to analyse its conditioning in a general way, so we concentrate on the first

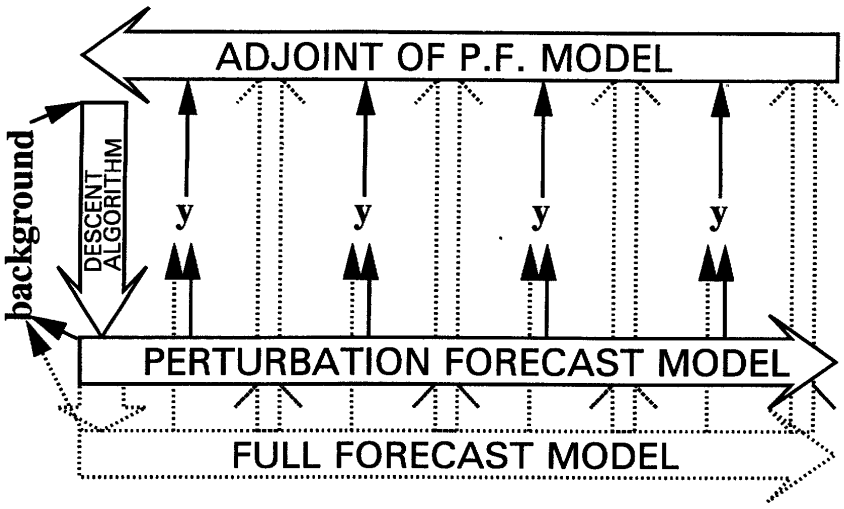

Fig. 1. Incremental four-dimensional variational assimilation.

term, which depends on the background error covariance. It has been observed that the errors in $\mathbf{x}^{\mathrm{b}}$ are usually balanced, and smooth. We assume that $\mathbf{x}^{\mathrm{g}}$ is similarly balanced and smooth. This means that balanced and smooth modes will correspond to small eigenvalues of $\mathbf{B}_{(w)}^{-1}$, while imbalanced, or rough modes will correspond to large eigenvalues. This large range of eigenvalues means that $\mathbf{B}_{(w)}^{-1}$ is ill-conditioned. To alleviate this ill-conditioning, we use a filter $\mathbf{U}$ designed to reduce the power in unbalanced or rough modes, and its generalised[2] inverse $\mathbf{U}^{-\mathrm{I}}$. We design these such that, approximately:

$$\begin{aligned}\mathbf{B}_{(w)}^{-1} &\approx \left(\mathbf{U}^{-\mathrm{I}}\right)^{\mathrm{T}} \mathbf{U}^{-\mathrm{I}} \\ \mathbf{B}_{(w)} &\approx \mathbf{U}\mathbf{U}^{\mathrm{T}}\end{aligned} \tag{8}$$

Then, defining a new control variable $\mathbf{v}$ such that

$$\mathbf{w}' = \mathbf{U}\mathbf{v} \tag{9}$$

our transformed variational problem is to find the $\mathbf{v}$ which minimises

$$\begin{aligned}J(\mathbf{v}) = &\frac{1}{2}\left(\mathbf{v}^{\mathrm{b}} - \mathbf{v}\right)^{\mathrm{T}} \mathbf{B}_{(v)}^{-1} \left(\mathbf{v}^{\mathrm{b}} - \mathbf{v}\right) \\ &+\frac{1}{2}\left(\mathbf{y}^{\mathrm{o}} - \mathbf{y}\right)^{\mathrm{T}} (\mathbf{E}+\mathbf{F})^{-1} \left(\mathbf{y}^{\mathrm{o}} - \mathbf{y}\right)\end{aligned} \tag{10}$$

where

$$\mathbf{v}^{\mathrm{b}} = \mathbf{U}^{-\mathrm{I}}\, \mathbf{w}'^{\mathrm{b}} \tag{11}$$

and the estimates of the observations are now given by:

$$\mathbf{y} = \tilde{H}(\mathbf{x}^{\mathrm{g}}, \mathbf{w}^{\mathrm{g}}, \mathbf{U}\mathbf{v}) \tag{12}$$

The Hessian of (10) is given by:

$$\left(\frac{\partial^2 J}{\partial \mathbf{v}^2}\right) = \mathbf{B}_{(v)}^{-1} + \mathbf{U}^{\mathrm{T}} \tilde{\mathbf{H}}^{\mathrm{T}} (\mathbf{E}+\mathbf{F})^{-1} \tilde{\mathbf{H}}\mathbf{U} \tag{13}$$

Because of (8)

$$\mathbf{B}_{(v)}^{-1} \approx \mathbf{I} \tag{14}$$

so the first term in (13) is much better conditioned than in (7).

5. Observation operators

The steps in calculating (5) are:

1. Calculate from $\mathbf{x}^{\mathrm{g}}$ by horizontal interpolation, and save for later use, columns $c_{\mathbf{x}}$.
2. For each new estimate of $\mathbf{w}'$, calculate some extra perturbation fields (*e.g.* relative humidity). This calculation will use linearisation state values from $\mathbf{w}^{\mathrm{g}}$.
3. Calculate columns $c'_{\mathbf{w}}$ by horizontal interpolation of $\mathbf{w}'$ and these extra fields.
4. Use a column version of the generalised inverse of the linearisation of S to calculate an incremented column:

$$c_{\mathbf{x}}^{+} = c_{\mathbf{x}} + \mathbf{S}_{col}^{-\mathrm{I}} c'_{\mathbf{w}} \tag{15}$$

5. Calculate $\mathbf{y}$ from $c_{\mathbf{x}}^{+}$. To simplify the adjoint, some terms may be calculated instead from $c_{\mathbf{x}}$:

$$\mathbf{y} = H_{col}(c_{\mathbf{x}}^{+}, c_{\mathbf{x}}) \tag{16}$$

This procedure means that, despite having a linear perturbation model, wa can use full, nonlinear, calculations in step 5. For instance radiance observations which are nonlinearly related to the model's temperatures and humidities can be used directly. On the other hand, nonlinear terms that are not

2 Modes which should be very strongly damped by $\mathbf{U}$ are, for computational reasons, omitted altogether, so $\mathbf{U}$ is not square. The generalised inverse sets these modes to zero.

important (perhaps the dependence of the drag coefficient on stability when predicting surface wind) can be calculated from $c_{\mathbf{x}}$ in (16), rather than from $c^+_{\mathbf{x}}$. This means they need not be considered in the 'adjoint': $\tilde{\mathbf{H}}^{\mathrm{T}} = (\partial \mathbf{y}/\partial \mathbf{w}')^{\mathrm{T}}$, which is needed in the descent algorithm to calculate

$$\left(\frac{\partial J^{\mathrm{o}}}{\partial \mathbf{w}'}\right)^{\mathrm{T}} = \left(\frac{\partial \mathbf{y}}{\partial \mathbf{w}'}\right)^{\mathrm{T}} \left(\frac{\partial J^{\mathrm{o}}}{\partial \mathbf{y}}\right)^{\mathrm{T}}$$
$$= \left(\frac{\partial c'_{\mathbf{w}}}{\partial \mathbf{w}'}\right)^{\mathrm{T}} \left(\frac{\partial c^+_{\mathbf{x}}}{\partial c_{\mathbf{w}'}}\right)^{\mathrm{T}} \left(\frac{\partial \mathbf{y}}{\partial c^+_{\mathbf{x}}}\right)^{\mathrm{T}} \left(\frac{\partial J^{\mathrm{o}}}{\partial \mathbf{y}}\right)^{\mathrm{T}} \tag{17}$$

6. Variable transforms for the full model

As well as the above conditioning consideration, we need to be able to evaluate J^{b} and its gradient, for which multiplication by $\mathbf{B}^{-1}_{(v)}$ is needed, so again a simple diagonal form is desired. Note however that there is no requirement that (14) should be exactly obeyed. Any discrepancy will lead to a discrepancy in (14), and the background penalty term in (10) allows for this.

Although it is the transform $\mathbf{U}$ and its transpose which are required in the minimisation, it is easier first to understand the generalised inverse transform $\mathbf{U}^{-\mathrm{I}}$. This is constructed in stages:

- using simple physical ideas to transform parameters ($\mathbf{U}_p^{-\mathrm{I}}$),
- using zonal and seasonal-average statistics to transform into empirical modes (EOFs) in the vertical ($\mathbf{U}_v^{-\mathrm{I}}$),
- filtering, to allow for different scales in the horizontal ($\mathbf{U}_h^{-\mathrm{I}}$).

Statistics on the residual covariances $\mathbf{B}_{(v)}$ can be collected, and modelled in a simple way for use in the background penalty term.

We know physical relationships between variables, such as the closeness to balance, and non-divergence, which imply that elements of $\mathbf{w}'$ which are different physical parameters, *e.g.* temperature and wind, are correlated[3]. Following the ideas of Parrish and Derber (1992), we use these relationships to design parameter transform so as to separate $\mathbf{w}'$ into three-dimensional fields of variables which are uncorrelated with each other. In the first version these are: velocity potential, stream function, the unbalanced part of the hydrostatic pressure[4], and relative humidity.

3 strictly, the errors are correlated

4 We use the linear balance equation to calculate a pressure increment from the vorticity increment, scale it by an empirical factor near one to get an estimate of the balanced pressure increment, and subtract it from the full pressure increment, to get the unbalanced pressure increment.

Within each three-dimensional field there are still correlations between points close in space. We can accumulate average vertical covariances within each three-dimensional field, for instance by comparing forecasts valid at the same time. Making some assumptions we can design $\mathbf{U}_v^{-\mathrm{I}}$ so as to separate each three-dimensional field into two-dimensional fields of EOF coefficient.

Finally we design $\mathbf{U}_h^{-\mathrm{I}}$ to act on each two-dimensional field, allowing for horizontal correlations. This is described further below. Having designed the transform from the (physically meaningful) $\mathbf{w}'$ into the control variable $\mathbf{v}$, it is relatively straightforward to derive its inverse $\mathbf{U}$, and then the transpose (or adjoint) $\mathbf{U}^{\mathrm{T}}$.

7. Implementation in 2DVAR

2DVAR is designed to test the above for a single two-dimensional field on a sphere, either globally, or for a rectangular limited area. For a single field there is no concept of 'balance'; the only prior knowledge about background errors is that they are likely to be smooth. Thus the transformation $\mathbf{U}$ is implemented using a horizontal filter. (In the equations below we omit the subscript h). It is assumed that (8) and (14) are exactly true. The $\mathbf{v}$ which minimises (10) is found using a descent algorithm.

7.1 Horizontal transform U

Horizontal correlations between grid-points are normally defined using a continuous correlation function of the grid-point positions. Usually they are taken to be (locally) homogeneous and isotropic, so that the correlation is a function of the distance between the points only. In this case it is a standard result that the fourier transform of the covariance function is the power spectrum. Our equations use matrix notation: $\mathbf{B}^{-1}$ is a symmetric matrix and $\mathbf{w}'$ is a column vector. Multiplication by the matrix $\mathbf{B}^{-1}$ does not represent a simple physical operation on the spatial field represented by $\mathbf{w}'$; the physical interpretation of the resulting vector depends on the grid. In contrast a filter does represent a physical process; for resolved scales the result should be independent of the grid. So $\mathbf{B}$ cannot be represented solely by a filter, with $\mathbf{B}^{-1}$ its inverse. Lorenc (1992) showed how to relate $\mathbf{B}$ to a filter $\mathbf{S_B}$, using a (symmetric) inner-product matrix $\mathbf{P}$, which is a diagonal matrix of grid box areas. Using the 'square root' of $\mathbf{P}$ to transform variables, gives:

$$\mathbf{B} = \mathbf{S_B}\,\mathbf{P}^{-1}$$
$$\mathbf{B}^{-1} = \mathbf{P}\,\mathbf{S_B}^{-1} \tag{18}$$

Not all filters are convenient; we shall see below that the square-root, adjoint, and inverse are needed. We have two options for implementing $\mathbf{S_B}$; either as a grid-point or spectral filter:

7.1.1 Grid-point filter

Some meteorological variables have error correlations which can be approximated using a Second Order Auto-Regressive (SOAR) function. Lorenc (1992) showed that this was approximately equivalent to two passes of a recursive filter. We assume that $\mathbf{S_B}$ can be expressed as two applications of a filter $\mathbf{R}$, a one-pass recursive filter, followed by a scaling proportional to the filter scale, designed to make the filter equivalent to a correlation[5]. We construct the two-pass filter from a one-pass filter and its adjoint in such a way that the result is exactly self-adjoint:

$$\mathbf{S_B} \equiv \mathbf{R}\mathbf{R}^* \tag{19}$$

Using the recursive filter method allows us to vary the horizontal correlation scale smoothly in the horizontal, for example we can have different values in northern hemisphere, southern hemisphere, and tropics. (19) ensures that the resulting filter is self-adjoint, even if the component recursive filter is not.

7.1.2 Spectral filter

An alternative way to perform a scale selective filter is via a spectral transform.

$$\mathbf{S_B} \equiv \mathbf{T}\mathbf{D}^2\mathbf{T}^{-1} \tag{20}$$

Here $\mathbf{T}$ is the spectral transform from wave-space to grid-point-space, and $\mathbf{D}^2$ is a diagonal matrix defining the damping required for each wave. For a spectral transform, the inverse is also the adjoint:

$$\mathbf{T}^{-1} \equiv \mathbf{T}^* \tag{21}$$

So, remembering that $\mathbf{D}$ is diagonal, we can define:

$$\mathbf{R} \equiv \mathbf{T}\mathbf{D}$$
$$\mathbf{S_B} \equiv \mathbf{R}\mathbf{R}^* \tag{22}$$

Then for either the grid-point or spectral filter we get:

$$\mathbf{U} = \mathbf{R}\mathbf{P}^{-\frac{1}{2}} \tag{23}$$

$$\mathbf{U}^{-\mathrm{I}} = \mathbf{P}^{\frac{1}{2}}\mathbf{R} \tag{24}$$

7.1.3 Comparison

These two approaches were tested in the 2DVAR program for a limited area grid. The spectral approach, using a double-sine FFT, proved to be more flexible; it could be tuned to match the grid-point filter nearly exactly, or some other power spectrum could be modelled.

This is illustrated in Fig. 2, which shows the analysis increments due to a single observation, using the recursive filter designed to match a SOAR (top left), and a spectral filter designed to match the recursive filter (top right). The recursive filter, as it acts along the grid rows and columns, does not have an isotropic response; it is easy to make the spectral filter isotropic (bottom left). For this example all methods converged in one iteration. In a similar test using 1000 randomly placed observations, 18 penalty evaluations were needed, to get convergence to almost machine precision. Computational costs for the two types of filter were similar.

7.2 The effect of preconditioning

The system is designed to find the minimum of (10). We also did experiments to instead find the minimum of (3). This is referred to below as the method without preconditioning. In the recursive filter which matches a SOAR (Lorenc 1992), a one-dimensional wave of length $2\pi/k$ is damped each pass by a factor given by:

$$S(k) = \frac{1}{1 + \frac{\alpha}{(1-\alpha)^2}\left[2\sin\left(\frac{k\ \delta x}{2}\right)\right]^2} \tag{25}$$

where the filter coefficient α is given by:

$$\alpha = 1 + E - \sqrt{E(E+2)}$$
$$E = 2\ N_{pass}\ \delta x^2/4s^2 \tag{26}$$

where s is the horizontal correlation scale, δx is the grid- length, and $N_{pass} = 1$.

Using (25) we can calculate the eigenvalues of $\mathbf{B}_{(w)}^{-1}$. Figure 3 shows a plot of the relative[6] eigenvalues for the configuration used in most of the tests described below, for a 31 by 31 grid with a gridlength of 100 km, and a horizontal correlation scale of 400 km. From the ratio of the largest and smallest eigenvalues we can see the contribution of $\mathbf{B}_{(w)}^{-1}$ to the condition number of the Hessian. For our example this is larger than 10^7, and it increases rapidly as the grid-length is reduced below 100 km. So if J^{b} dominates, the problem is badly ill-conditioned, and we can expect convergence of minimisation routines without preconditioning to be slow.

This is demonstrated in a simple experiment with a single observation, with value $y^{\mathrm{o}} = 1.0$, at the centre of the square grid area. The background field, and first-guess, is zero everywhere. For this case the preconditioning is perfect, so the descent algorithm converges in one iteration (which requires two penalty evaluations). The method without preconditioning however needs several thousand penalty evaluations before nearing convergence, (judged by the approach of the norm of the gradient of J to zero), as shown in Fig. 4.

Despite the many iterations, the method without preconditioning has still not converged to the exact solution, as can be seen in Fig. 5. We verified

5 filtering a unit delta function should give a correlation field, with maximum value one at the position of the delta function

6 The full inverse covariance also contains a scaling dependent on the background error variance, and the horizontal scale coefficient s.

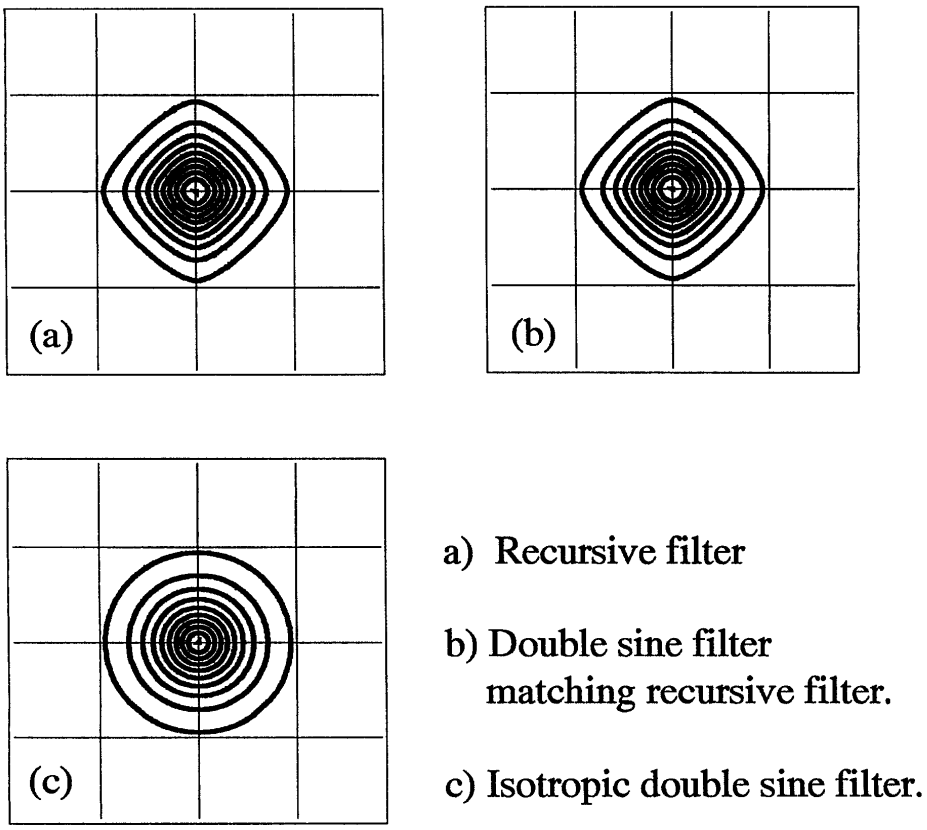

Fig. 2. Increments due to a single observation, for various filters.

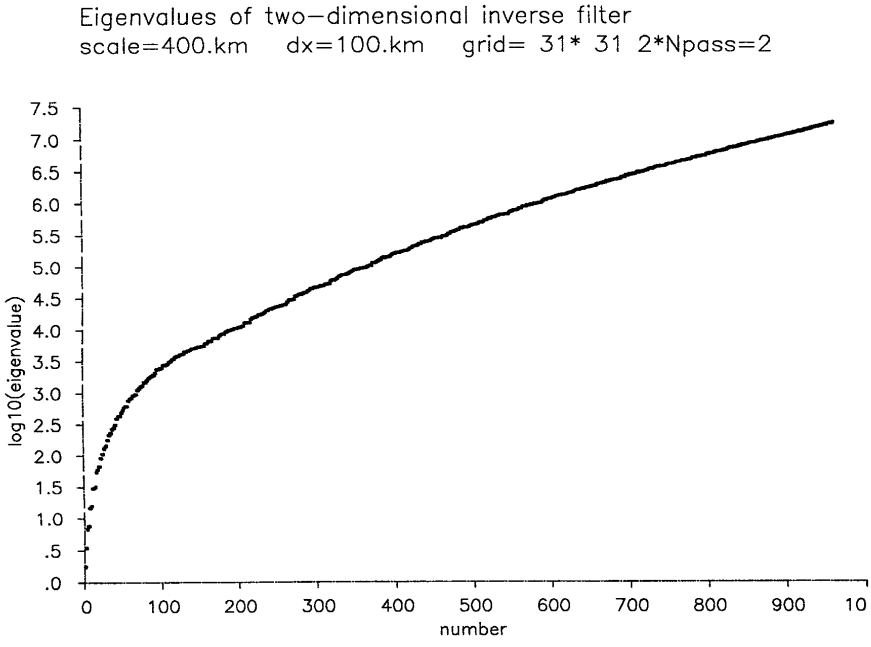

Fig. 3. Eigenvalues of an inverse filter proportional to the inverse error covariance.

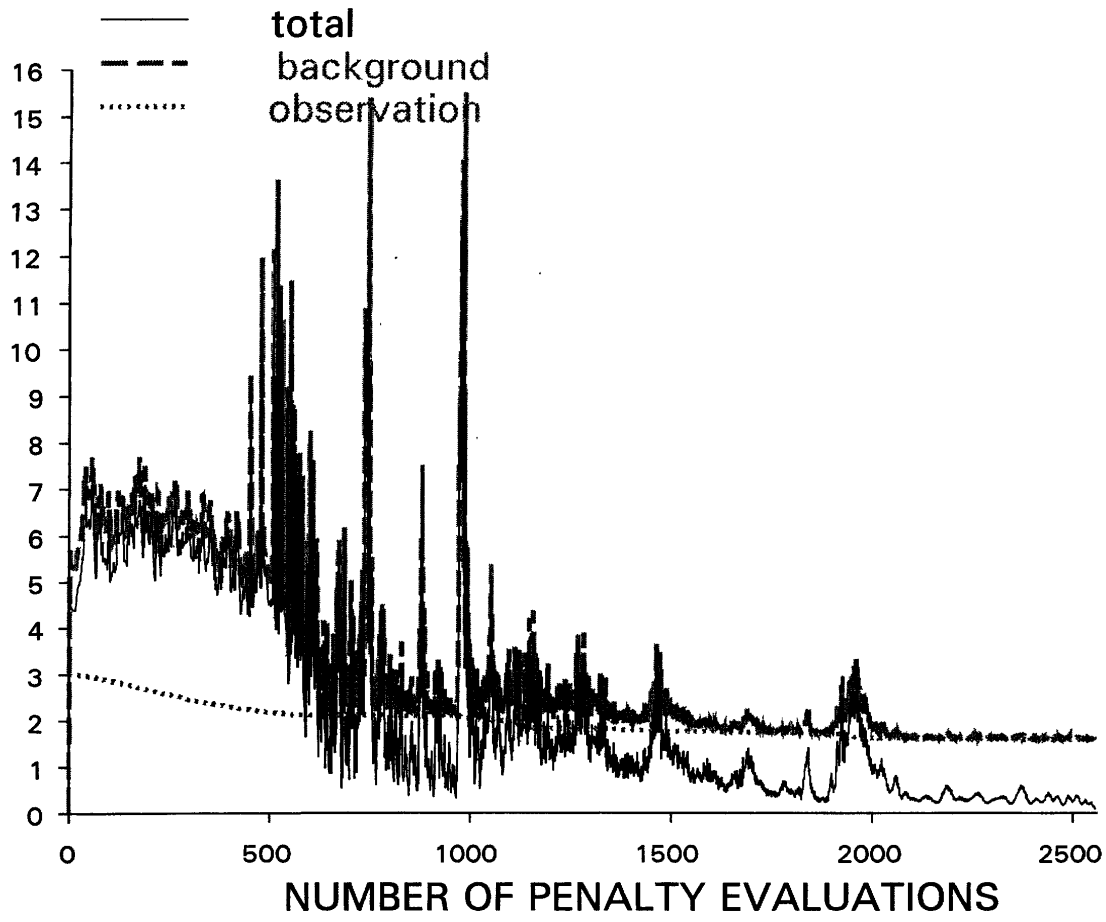

Fig. 4. Solid line — norm of gradient of J at each iteration during the minimisation (at convergence it should be zero). Dashed and dotted lines — norm of contribution to gradient from J^{b} and J^{o} terms.

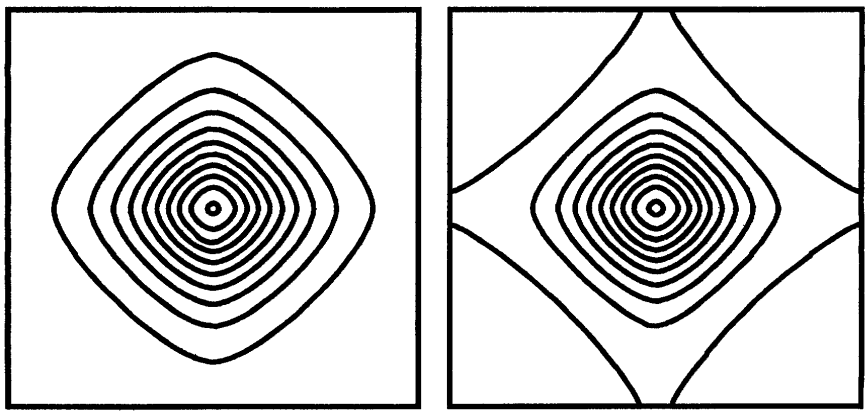

Fig. 5. Left — solution field obtained with preconditioned method for one observation at the pole. Right — solution field obtained with the method without preconditioning.

that the method with preconditioning was indeed giving the exact solution by using it as first-guess for another descent iteration. Neither method could improve on it.

8. Conclusions

The design of a variational analysis system has been set out which should meet our requirements for: global and limited-area versions, a development path from three-dimensional to four-dimensional, and a computational cost that should be affordable. Horizontal smoothness may be constrained using a grid-point filter, or via a transformation to spectral space. The recursive filter tested has the advantage of allowing horizontal variations of scale, but has limited flexibility in the shape of its response curve. Because it is computed along the grid directions, it cannot in general be made isotropic. We

expect to use the spectral option for most variables. Careful preconditioning is essential for convergence, particularly if there are isolated observations and prior constraints such as smoothness are important. Conditioning can be achieved while calculating the smoothness penalty, by a change of variables.

Acknowledgments

These ideas have benefited from discussions with those implementing the system, especially Phil Andrews, Dale Barker, Stuart Bell, Bruce Ingleby, Jon Shurlock, and Mike Thurlow.

References

Andersson, E., J. Pailleux, J.-N. Thepaut, J.R. Eyre, A.P. McNally, G.A. Kelly and P. Courtier, 1993: Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. *Quart. J. Roy. Met. Soc.*, **120**, 627–653

Courtier, P., J.-N. Thepaut and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Met. Soc.*, **519**, 1389–1408

Courtier, P., J. Derber, R. Errico, J.-F. Louis and T. Vukicevic, 1993: Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology *Tellus*, **45A**, 342–357

Lorenc, A.C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Met. Soc.*, **112**, 1177–1194.

Lorenc, A.C., 1992: Iterative analysis using covariance functions and filters *Quart. J. Roy. Met. Soc.*, **118**, 569–591

Parrish, D.F. and J.C. Derber, 1992: The National Meteorological Center's Spectral Statistical Interpolation analysis system. *Mon. Wea. Rev.* **120**, 1747–1763

Thepaut, J.-N., D. Vasiljevic, P. Courtier and J. Pailleux, 1993 Variational assimilation of conventional meteorological observations with a multi-level primitive-equation model. *Quart. J. Roy. Met. Soc.*, **199**, 153–186