The University of Reading

A singular vector perspective of 4D-Var: The spatial structure and evolution of baroclinic weather systems

C. Johnson

National Center for Atmospheric Research, Boulder CO, USA

B.J. Hoskins, N.K.Nichols

School of Mathematics, Meteorology and Physics, The University of Reading, Whiteknights, Reading UK

and

S.P. Ballard

Met Office, JCMM, Earley Gate, Reading UK

Department of Mathematics

Abstract

The extent to which Four-Dimensional Variational data assimilation (4D-Var) is able to use information about the time-evolution of the atmosphere to infer the vertical spatial structure of baroclinic weather systems is investigated. General results are derived using the singular value decomposition of the 4D-Var observability matrices in an idealized Eady model setting. These results are confirmed with 4D-Var analyses.

The results show that 4D-Var performs well at correcting the errors that would otherwise rapidly corrupt a forecast. However, in a few cases, 4D-Var may add a rapidly growing error instead of correcting a decaying error. The ability to extract the time-evolution information can be maximized by placing the observations as far apart as possible in time, and the specification of the case-dependent background error variances is crucial in being able to project the observational information onto analysis increments that lead to the appropriate growth rate.

1.   Introduction

The dynamical instability of the atmosphere means that small perturbations that are introduced into the flow may grow rapidly. For example, the flow at mid-latitudes is baroclinically unstable due to the vertical shear associated with the meridional temperature gradient. This wave-instability provides the dominant mechanism for disturbances to develop into mid-latitude weather systems. In such development, the vertical spatial structure of the disturbance plays a fundamental role in governing the development. For example, the normal-mode analysis of simple linear models (Charney, 1947; Eady, 1949) showed that the fastest growing structure exhibits a westward tilt with height in the pressure field. This vertical tilt leads to a process known as self-development where the upper and lower level waves act to intensify each other, leading to exponential modal growth. In contrast, the fastest decaying structure exhibits an eastward tilt with height so that the circulations associated with the upper and lower level waves act to weaken each other. More recent studies (Farrell, 1982, 1984) showed that it is possible for the growth rate of a disturbance to exceed the exponential growth of the fastest growing normal mode over a limited period of time. Such disturbances may be found by computing the singular vectors of the linear model (Farrell, 1989; Buizza and Palmer, 1995). The spatial structures of these disturbances are characterized by localized tilted interior potential vorticity (PV) structures which separate and untilt under the action of the shear, leading to amplification (Badger and Hoskins, 2001). Again, the vertical structure of the disturbance is fundamental for this rapid non-modal growth.

The rapid growth of small disturbances in a baroclinically unstable flow has two consequences for data assimilation. The first consequence is that data assimilation algorithms need to be able to generate analyses with the vertical structures that are necessary for both modal growth and decay and also for non-modal growth. The second consequence is that small errors in the initial conditions of a numerical weather forecast may rapidly develop into large forecast errors. It is therefore important that the data assimilation algorithm is able to correct such rapidly growing errors.

Algorithms such as Three-Dimensional Variational data assimilation (3D-Var) (Courtier et al., 1998), using specified error covariance matrices, generate an analysis by blending together observations near the analysis time with a background state. Algorithms such as Four-Dimensional Varia-

tional data assimilation (4D-Var) (Rabier et al., 2000) and Ensemble Filters (Evensen, 1994; Tippett et al., 2003; Lorenc, 2003a,b) extend this method by using a forecast model to link together observations that are distributed in time and also to evolve the background error covariance matrix (Lorenc, 1986). This means that observations are combined with dynamically evolved covariances so that 4D-Var is able to generate the vertical structures that are needed for baroclinic growth. For example, single observation experiments by Thépaut et al. (1996) showed that 3D-Var analysis increments do not exhibit any tilt with height, whereas 4D-Var analysis increments are anisotropic and also exhibit a westward tilt with height. Idealized experiments by Rabier and Courtier (1992) showed that 4D-Var is able to combine the information provided by the model dynamics with observations of only the eddy part of the flow to reconstruct a baroclinic wave.

Thépaut et al. (1996) demonstrated a strong link between the dominant singular vectors of the tangent linear model and 4D-Var analysis increments. Theoretical studies by Pires et al. (1996) and Rabier et al. (1996) also showed that the 4D-Var cost function is most sensitive to analysis increments given by the dominant singular vectors. Hence 4D-Var should provide accuracy of the unstable components of the flow by correcting the components in the initial errors that are rapidly growing.

Many of the previous 4D-Var studies have considered observations given at only the end of the window. However, one of the major advantages of 4D-Var is that it is able to use the model dynamics to link together a time-sequence of observations. This extra information can be used to build a better picture of both the spatial and temporal structure of the atmosphere by ensuring consistency between the observations and the predicted evolution of the atmospheric state. Here and in the companion paper Johnson et al. (2005a), hereafter JHN, we use the singular value decomposition (SVD) of the observability matrix to examine the spatial and temporal interpolation in 4D-Var that results from the interaction of a time-sequence of observations with the model dynamics. The technique was used in JHN to investigate the extent to which 4D-Var could use the time evolution information to reconstruct the state in an unobserved region, whilst filtering the observational noise. The technique is used in this paper to investigate the extent to which 4D-Var is able to use the information from a time-sequence of observations to generate the vertical spatial structures that are necessary for accurate

forecasts of baroclinic weather systems and also to correct errors in the initial conditions that result in rapid growth. For a complete assessment of whether 4D-Var is able to generate the appropriate vertical structures, we investigate whether 4D-Var is able to generate the appropriate structures for both baroclinic growth and decay.

The 2D Eady model is used throughout this paper. This linear model is one of the most simple models of baroclinic instability, and allows a clear understanding of the operation of 4D-Var. The 4D-Var algorithm, singular vector technique and Eady model are described in section 2. In section 3, we compare cases with growing and decaying modes and investigate the impact of the accuracy of the observations, the temporal position of the observations in the assimilation window and the position of the observations in the spatial domain. The experiments in section 3 only consider errors that result in modal growth or decay but the experiments in section 4 also consider cases with errors that result in non-modal growth. These final experiments are used to investigate the impact of the specified error variances on the growth rate of the following forecast. The main conclusions are given in section 5, and the paper then ends with a discussion. Further studies associated with this work may be found in Johnson (2003) and Johnson et al. (2005b).

2. Description of the Eady model experiments

a. 4D-Var algorithm

4D-Var finds the optimal state trajectory that is close to the observed values over a specified assimilation time window and is close to the background state at the beginning of the window. It is too expensive to compute the optimal state at every time level. Therefore, with the reduction of the control variable (Le Dimet and Talagrand, 1986), only the initial conditions are adjusted and the trajectory is required to satisfy the model equations exactly. In the following, we only consider linear models, but it is possible to apply 4D-Var to nonlinear models by using the nonlinear model in a non-incremental formulation, or using a linearized model in an incremental formulation (Courtier et al., 1994).

Mathematically, the analysis at time t_0, $\mathbf{x}^a$, is given by the initial state, $\mathbf{x}_0$, which minimizes the

cost function,

$$J(\mathbf{x}_0) = \mu^2(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \sum_{i=0}^{N}(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}_i)^T \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{H}_i\mathbf{x}_i), \tag{1a}$$

subject to the linear model constraint,

$$\mathbf{x}_{i+1} = \mathbf{M}(t_{i+1}, t_i)\mathbf{x}_i \ \text{ for } i = 0, \dots, N-1. \tag{1b}$$

Here, $\mathbf{x}_i$ is the state vector at time t_i, $\mathbf{x}^b$ is the background state, $\mathbf{y}_i$ is the vector of observations at time t_i, and $\mathbf{H}_i$ is the (linear) observation operator which converts from state space to observation space, and $\mathbf{M}(t_{i+1}, t_i)$ is the linear model operator from time t_i to t_{i+1}. In the majority of 4D-Var algorithms, $\mathbf{B}$ and $\mathbf{R}_i$ are the specified background state and observation error covariance matrices and μ^2 is taken as 1. Here we choose to instead consider an explicit dependence on the parameter μ^2; the parameter μ^2 is considered to be the error variance ratio between the observations and background state, and $\mathbf{B}$ and $\mathbf{R}_i$ are considered to be the specified background state and observation error correlation matrices. When $\mu^2 > 1$ it implies that the background state is more accurate than the observations and hence the algorithm draws close to the background state. When $\mu^2 < 1$, it implies that the background state is less accurate than the observations and hence the algorithm draws closer to the observations.

b.   *SVD framework*

The SVD of what is known as the observability matrix provides a useful tool to examine the interaction of observations and dynamics within 4D-Var, as shown by JHN. The observability matrix incorporates the observation and background error covariances, the observation locations and the linear forecast model. We denote the block diagonal matrix with the observation error correlations at each time on the diagonal as $\hat{\mathbf{R}}$, and the symmetric square roots of $\hat{\mathbf{R}}$ and $\mathbf{B}$ as $\hat{\mathbf{R}}^{1/2}$ and $\mathbf{B}^{1/2}$ respectively. The generalized observation operator is a block matrix linking the time-sequence of

6

observations, $\hat{\mathbf{y}} = \left[\mathbf{y}_0^T, \mathbf{y}_1^T, \ldots, \mathbf{y}_N^T \right]^T$, to the initial state vector, $\mathbf{x}_0$, and is denoted by

$$\hat{\mathbf{H}} = \left[\begin{array}{cccc} \mathbf{H}_0^T & (\mathbf{H}_1 \mathbf{M}(t_1, t_0))^T & \ldots & (\mathbf{H}_N \mathbf{M}(t_N, t_0))^T \end{array} \right]^T. \tag{2}$$

Then the observability matrix $\widetilde{\mathbf{H}}$ and its associated singular value decomposition is given by

$$\widetilde{\mathbf{H}} = \hat{\mathbf{R}}^{-1/2} \hat{\mathbf{H}} \mathbf{B}^{1/2} = \sum_{j=1}^{r} \lambda_j \mathbf{u}_j \mathbf{v}_j^T. \tag{3}$$

The singular values, left singular vectors (LSVs), right singular vectors (RSVs) and rank of the observability matrix are denoted by λ_j, $\mathbf{u}_j$, $\mathbf{v}_j$ and r.

The RSVs form an orthonormal basis in the uncorrelated state space, which means that the 4D-Var analysis increments can be written as:

$$\mathbf{x}^a - \mathbf{x}^b = \sum_{j=1}^{r} f_j c_j \mathbf{B}^{1/2} \mathbf{v}_j, \tag{4a}$$

where

$$f_j = \frac{\lambda_j^2}{\mu^2 + \lambda_j^2}, \tag{4b}$$

$$c_j = \frac{\mathbf{u}_j^T \hat{\mathbf{R}}^{-1/2} \hat{\mathbf{d}}}{\lambda_j}. \tag{4c}$$

and where $\hat{\mathbf{d}} = \hat{\mathbf{y}} - \hat{\mathbf{H}} \mathbf{x}^b$ is the generalized innovation vector. The RSVs are independent of the observed values and the background state. They give the possible spatial structures that can be analysed by 4D-Var for the specified linear model dynamics, linearization trajectory, error covariances and observation locations in both space and time.

The particular combination of RSVs that are included in the analysis increment is determined by the coefficients, c_j. If the vector $\hat{\mathbf{R}}^{-1/2} \hat{\mathbf{d}}$ has a large projection onto the LSV $\mathbf{u}_j$, then the corresponding RSV is given a large weight. The observational noise can have a large projection onto the LSVs with small singular values so that the corresponding RSVs would dominate the analysis increment. However, these are filtered by the filter factors, f_j, which damp the RSVs with small singular values,

$\lambda_j^2 << \mu^2$. Hence the algorithm selectively filters unrealistic structures and the analysis increment is dominated by the RSVs with large singular values.

c.   Eady model

The non-dimensional equations for the 2D Eady model (Eady, 1949) are now described. The basic state is given by a linear zonal wind shear with height in a domain between two rigid horizontal boundaries. The domain is infinite in the meridional direction and the only dependence on this direction is the uniform meridional temperature gradient which is in thermal wind balance with the zonal wind shear. The density, static stability and Coriolis parameters are all taken to be constants.

The perturbation to the basic state is described by the non-dimensional buoyancy, b, on the upper and lower boundaries and by the non-dimensional quasi-geostrophic potential vorticity (QGPV), q, in the interior. Equivalently, the perturbation may also be described by the non-dimensional geostrophic streamfunction, ψ, which satisfies:

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial z^2} = q, \qquad \text{in } z \in \left[-\frac{1}{2}, \frac{1}{2}\right], \qquad x \in [0, X], \qquad (5\text{a})$$

$$\frac{\partial \psi}{\partial z} = b, \qquad \text{on } z = \pm\frac{1}{2}, \qquad x \in [0, X], \qquad (5\text{b})$$

where x is the non-dimensional distance in the zonal direction, and z is the non-dimensional height. The non-dimensional time will be denoted by t. The perturbation to the basic state is advected zonally by the basic shear flow as described by the non-dimensional QG thermodynamic equation and QGPV equation:

$$\left(\frac{\partial}{\partial t} + z\frac{\partial}{\partial x}\right) b = \frac{\partial \psi}{\partial x}, \qquad \text{on } z = \pm\frac{1}{2}, \qquad x \in [0, X], \qquad (6\text{a})$$

$$\left(\frac{\partial}{\partial t} + z\frac{\partial}{\partial x}\right) q = 0, \qquad \text{in } z \in \left[-\frac{1}{2}, \frac{1}{2}\right], \qquad x \in [0, X]. \qquad (6\text{b})$$

The perturbation is periodic in the horizontal so that the lateral boundary conditions are: $b(0, z, t) = b(X, z, t)$ and $q(0, z, t) = q(X, z, t)$. The model is discretized using 11 vertical levels for QGPV with 40 grid points in one periodic interval in x. The advection equations are discretized using a

leap-frog advection scheme. The computational details to compute the 4D-Var analysis and the SVD are identical to those in JHN.

Dimensional values for x, z and t will be used to discuss the results. These are based on a domain height of $10km$, buoyancy frequency $N = 10^{-2}s^{-1}$, Coriolis parameter $f = 10^{-4}s^{-1}$. The difference in the basic state zonal wind shear over a height of 10km is $40ms^{-1}$. The implied wavelength of the most rapidly growing and most rapidly decaying normal modes is about $4000km$.

3.   Results considering modal growth and decay

In this section, we use the Eady model to investigate the extent to which 4D-Var is able to use the temporal evolution information contained in the observations to correct the errors in the initial conditions that result in either modal growth or decay.

a.   *Experiment Description*

The true state is given by either the most rapidly growing or the most rapidly decaying normal mode, the equations for which are given in Appendix A. The growing mode exhibits a westward tilt with height of the streamfunction field, ψ, and an eastward tilt with height of the buoyancy field, $\partial\psi/\partial z$. This vertical tilt is associated with the growth in amplitude of the mode with time. In contrast, the decaying mode exhibits an eastward tilt in the streamfunction field and a westward tilt in the buoyancy field.

The background state is given by the true state but with either an amplitude or a phase error. For the amplitude error, $\mathbf{x}^b = 0.5\mathbf{x}^t$; for the phase error, the true state is shifted westward by 500km. Thus, in all the experiments, the background state error is also given by either a growing or decaying normal mode. These experiments can be used to compare cases in which the errors in the initial conditions are either growing or decaying.

The interior QGPV is zero for all the true and background states. The control variables used in the 4D-Var algorithm are given by only the upper and lower boundary buoyancy at the beginning of the window. This means that analysis increments may only be added to the buoyancy field and not to

the QGPV field. It is assumed that there are no observation error correlations, $\mathbf{R} = \mathbf{I}$. The specified background error correlations, $\mathbf{B}$, have a horizontal correlation length scale $l = 1000km$, and are defined in Appendix A.

Perfect synthetic observations of only the lower boundary buoyancy are given at two times during a 12h assimilation window. The assumption of perfect observations is not crucial here. In an equivalent experiment with noisy observations, the RSVs that correspond to noise would be damped by the filter factors and hence a similar analysis would be obtained. Thus, in these experiments, varying the size of the error variance, μ^2, has the same effect as varying the size of the observational noise. This assumes that in the assimilation of noisy observations the appropriate value for μ^2 is used. This is described in further detail in Appendix B.

b.   *Preliminary Examples*

Before we discuss the main results, we show some example 4D-Var analyses from these experiments. The analyses are shown at the middle of the 12h assimilation window. In Fig. 1a-b, the true state is given by a growing normal mode, the background state has an amplitude error and the specified error variance ratio is $\mu^2 = 1$. Observations of the lower level buoyancy are given at the beginning and the end of the 12h assimilation window. The 4D-Var algorithm draws close to the observations so that the analysis is close to the true state on the lower boundary. The algorithm also increases the amplitude of the upper boundary somewhat so that the growth rate of the analysis through the time window is more consistent with the observations. The equivalent analysis for the decaying mode is shown in Fig. 1c-d. The algorithm again draws close to the observations on the lower boundary, but the upper boundary wave is not adjusted as much as for the growing mode. In particular, the upper boundary wave is shifted eastward with only a slight increase in the amplitude.

Figure 2 shows similar experiments but with phase errors instead of amplitude errors. The analysis for the true state given by a growing mode is shown in Fig. 2a-b. The upper boundary wave is moved eastward so that it is closer to the true state, but there is also a reduction in the amplitude. The equivalent analysis for the true state given by the decaying mode is shown in Fig. 2c-d. Again the upper boundary wave has been shifted slightly but less information about the upper boundary

wave is inferred than for the growing mode. In both the amplitude and phase error cases, 4D-Var is much better able to correct the unobserved upper boundary wave for the case where the true state, and hence the background state error results in growth rather than decay.

These experiments illustrate that 4D-Var is able to use the time-evolution information to correct the upper level wave. However, it is not clear why 4D-Var is better at correcting the growing errors than the decaying errors, and how the analyses can be improved.

In the following experiments the SVD framework is first employed to consider further the extent to which 4D-Var is able to correct the upper level wave and produce an analysis with the correct growth rate. The concepts that are learned from the SVD framework are then confirmed by 4D-Var analyses. The analyses are verified by comparing the behaviour, including the magnitude and phase error, over the following forecast interval. The magnitude is evaluated using the non-dimensional kinetic energy (KE) norm:

$$KE = \iint v^2 dx dz \tag{7}$$

where $v = \partial \psi / \partial x$ is the perturbation meridional wind. The phase error is evaluated using the correlation between the streamfunction fields of the analysis and the true state. The error-correlation takes a value of one when the analysis is completely in phase with the true state and takes a value of minus one when they are completely out of phase.

We first consider the impact of the accuracy of the observations by varying the value of μ^2; we then examine the impact of the temporal position of the observations by varying the time of the first set of observations; and we finally examine the impact of the spatial position of the observations by observing horizontal lines of the buoyancy field at different heights.

c.   *Accuracy of the observations*

The ability of 4D-Var to reconstruct the upper level wave for different values of μ^2 is now examined. The observations are again of the lower level buoyancy at the beginning and the end of the window, but the value of μ^2 is varied. It is assumed that in the appropriate value of μ^2 is used in the assimilation of real data. A large value of μ^2 implies that relatively inaccurate observations are assimilated and so the analysis is close to the background state. A small value of μ^2 implies that relatively accurate

observations are assimilated and so the analysis is close to the true state. Hence varying the size of μ^2 allows us to investigate the impact of the accuracy of the observations.

1). SVD RESULTS

We first examine the RSVs of the observability matrix and discuss the general conclusions that are implied from these. Section 2 shows that the uncorrelated analysis increment can be written as a linear combination of the RSVs of the observability matrix. For our purposes, only the RSVs that contribute to the analysis increment are of interest; these are the RSVs that have non-zero values of c_j (4c). The values of c_j are large when the generalized innovation vector $\hat{\mathbf{d}}$ is in the same direction as the corresponding LSVs. In these carefully constructed experiments (with perfect observations, a true state given by a growing or decaying normal mode, and a background state that has either an amplitude or phase error) the innovations $\hat{\mathbf{d}}$ are also either given by a growing or decaying normal mode. This means that there are only four values of c_j that are non-zero, to machine precision. These are for $j = 2, 3, 6,$ and 7. Further, the RSVs occur in pairs so that RSVs 2 and 3 have the same singular value ($\lambda = 2.99$) and an identical structure except for a phase shift of $1000km$. Similarly RSVs 6 and 7 both have $\lambda = 0.74$. Thus, it is only necessary to examine the spatial structure of the second and sixth RSVs. These are shown in Figs. 3 and 4. Note that the RSVs, defined in uncorrelated state space, are premultiplied by the square root of the background error correlation matrix so that they are in correlated state space and to be consistent with (4). The structure of the same fields after they have been evolved by the linear model $\mathbf{M}$ are also shown. These give the associated structures at the end of the assimilation window. It should be noted that the evolved RSV at the final time is not the same as the LSV, as these RSVs are the singular vectors of the observability matrix $\widetilde{\mathbf{H}}$ and not of the model $\mathbf{M}$.

RSV 2, from the first pair of RSVs, is shown in Fig. 3. It has a large amplitude on the lower boundary, which is the observed region. The buoyancy field tilts eastward with height and the stream-function field tilts westward with height. This tilt is characteristic of a growing solution but it does not tilt as much as for a normal mode. When this RSV is evolved by the model to give the structure at the final time we find that the lower boundary wave increases slightly in amplitude and there is a large

increase in the amplitude of the wave on the upper boundary wave such that it is similar to that on the lower boundary. Also the upper boundary buoyancy wave moves westward and the lower boundary wave moves eastward so that the streamfunction field exhibits a greater tilt with height, close to that of the unstable normal mode. Thus, the structure of the RSV at the final time more closely resembles the structure of a normal mode.

RSV 6, from the second pair of RSVs, is shown in Fig. 4. At the initial time, it has a large amplitude on the unobserved upper boundary. The buoyancy field tilts westward with height and the streamfunction field tilts eastward with height, characteristic of a decaying mode, but again it does not tilt as much as for a normal mode. When the RSV is evolved to the end of the window, the circulation associated with the upper boundary wave acts to the weaken the lower boundary wave until it becomes zero and then begins to grow again with the opposite sign so that at the final time the buoyancy field tilts eastward with height and the streamfunction field tilts westward with height. Its structure at the final time is more similar to that of a growing wave than a decaying wave.

The coefficients c_j have a large value when the generalized innovation vector $\hat{\mathbf{d}}$ is in the same direction as the corresponding LSVs. Thus, the structure of the LSVs show how the observational information is projected onto the RSVs. The LSVs exist in observation space, which is the lower boundary buoyancy at the initial and final time. The LSVs corresponding to the first pair of RSVs have a similar structure to the lower boundary buoyancy fields of the first pair of RSVs and are therefore not shown. There is very little change in the shape and amplitude of the wave with time which implies that the first pair of RSVs give the structure corresponding to the general shape of the observed wave. The pair of LSVs corresponding to the second pair or RSVs (also not shown) have a similar structure to the lower boundary buoyancy fields of the second pair of RSVs. The wave at the final time has the opposite sign to that at the initial time. This implies that the second pair of RSVs correspond to extracting the growth rate from the observed wave.

From the Tikhonov filter factors, f_j (4b), the algorithm damps the RSVs with $\lambda_j^2 \ll \mu^2$ so that as the variance ratio, μ^2, is increased the algorithm damps more of the RSVs with small singular values. In this case, the first pair of RSVs with $\lambda = 2.99$ contain the information that is needed to reconstruct the state in observed regions and also to give a growing analysis increment. In contrast,

the second pair of RSVs, with $\lambda = 0.74$, contain the information needed to reconstruct the state in the unobserved regions and also to give a decaying analysis increment. In particular, these RSVs correspond to the information detecting the growth or decay of the observed wave. This growth is harder to detect than the general spatial structure of the lower boundary wave and hence these RSVs have a smaller singular value than the first pair of RSVs and are damped by the algorithm if the observational noise is relatively large i.e $\mu > 0.74$.

From the SVD results, we expect that when the observations are extremely accurate, and hence $\mu = 0$, the algorithm is able to add both pairs of RSVs to the background state. When the observational noise is larger so that $\mu = 0.74$, the algorithm will filter out much of the second pair of RSVs. These are the vectors that are needed to reconstruct the upper level wave and to give a decaying analysis increment. When the observational noise is increased further, so that $\mu = 2.99$, both pairs of RSVs are strongly filtered, and so there is little difference between the background state and the analysis. This is now confirmed with the 4D-Var analyses and forecasts.

2). 4D-VAR ANALYSES

4D-Var analyses using observations of the lower level buoyancy at the beginning and the end of a 12h assimilation window, but with different specified variance ratios are now compared. We first consider the cases where the background state has an amplitude error. The KE values for the 12h window and following 36h forecast for the growing mode case are shown in Fig. 5a. The analysis is close to the true state when μ^2 is small ($\mu^2 = 0.1$) but as μ^2 increases, the RSVs are filtered more so that the analysis is close to the background state. The equivalent diagram for the decaying mode (Fig. 5b) is more interesting. When $\mu^2 = 0.1$, the analysis is close to the true state. When $\mu^2 = 1$, the first RSVs contribute to the analysis increment, but the second pair of RSVs are filtered. It is the second pair that are required to form a decaying increment. Hence, the KE decays through the assimilation window, but then begins to grow so that the 36h forecast is worse than that from the background state. When $\mu^2 = 100$, the observations are assumed to be so inaccurate that both pairs of RSVs are filtered, and the analysis remains close to the background state. It should be noted that the diagram for the decaying mode has a different scale, and so the impact on the 36h forecast is small compared

with the growing mode.

There are similar results for the cases where the background state has a phase error. For both the growing and decaying modes, the KE values for the background state are identical to those for the true state (Fig. 6). The error can only be seen in the error correlation values (Fig. 7). For the growing mode, when $\mu^2 = 0.1$, both pairs of RSVs contribute to the analysis increment, and both the phase and KE are close to the true state. When $\mu^2 = 1$, only the first pair of RSVs contribute to the analysis increment. This acts to correct the lower level wave, but not the upper level wave and hence the growth rate is reduced but the phase error is improved. When $\mu^2 = 100$, both pairs of RSVs are filtered so that the analysis is close to the background state. For the decaying mode, when $\mu^2 = 0.1$, the analysis has almost the correct rate of decay and no phase error for the majority of the forecast. When $\mu^2 = 1$, it is mostly only the first pair of RSVs that are added to the background state. This means that the growing increment eventually dominates the analysis so that the KE grows and the error correlation eventually has the wrong sign. This is because the streamfunction field of the analysis begins to tilt westward instead of eastward. When $\mu^2 = 100$, both RSVs are filtered and so the analysis is close to the background state. Again, for the decaying mode, the 36h forecast when observations are assimilated can in fact be worse than the forecast without assimilating observations.

d.   *Temporal position of the observations*

For the investigation of the impact of the temporal position of the observations in the assimilation window, observations of the lower boundary buoyancy are again given at two times during a 12h assimilation window. The first (initial) set of observations are given at either the beginning (T+0), middle (T+6) or end (T+12) of the window whilst the second (final) set of observations are always given at the end of the window. Note that the experiments with the initial set of observations at the end of the window are equivalent to experiments with a single set of observations at only the end of the window and with the variance ratio, μ^2, halved.

1). SVD RESULTS

Singular value decompositions of the corresponding observability matrices are first considered. For a fixed assimilation window length, the temporal position of the initial observations has very little impact on the spatial structure of the RSVs and so their structures are not shown. There is, however, a change in the singular values. Fig. 8 shows the singular values of the first and second pairs of RSVs that contribute to the analysis increment as selected by the values of c_j (not necessarily always numbers $2\&3$ and $6\&7$). When the initial observations are at T+0, the singular values are $\lambda = 2.99$ and 0.74. These correspond to the RSVs shown in Figs. 3 and 4. When the initial observations are moved towards the end of the window, the singular value of the first pair of RSVs increases whilst that of the second pair decreases. For initial observations at T+6, the singular values are 3.17 and 0.34. When the initial observations are at T+12, the singular value of the second pair of RSVs is zero; when the observations are only the end of the window, only the first pair of RSVs can contribute to the analysis increment as there is no observational information about the growth of the wave.

The second pair of RSVs correspond to the information about the growth rate of the wave. Intuitively, we know that more information about the growth rate can be extracted if the time between the initial and final time observations is longer. When the observations are far apart in time, there is a large difference in the amplitude of the observed wave at the initial and final times. When the observations are moved closer together in time, there is a small difference in the amplitude at the initial and final times and therefore it is much more difficult to infer the growth rate accurately when assimilating relatively noisy observations. It is for this reason that the singular value of the second pair of RSVs increases as the observations are moved further apart in time.

To be able to include the second pair of RSVs in the analysis increment, we need a value of μ^2 that is smaller than about 0.5. Further, the singular values show that the second pair of RSVs will have a larger contribution to the increment if the initial observations are closer to the beginning of the window. This is now confirmed by the 4D-Var analyses.

2). 4D-Var analyses

4D-Var analyses using observations of the lower level buoyancy but at different times are now compared. The first set of observations are given at the beginning, middle or end of the window and the second set of observations are always given at the end of the window. For all the cases, we select a variance ratio of $\mu^2 = 0.1$. This is close to the singular value of the second pair of RSVs and so allows the effect of the position of the observations to be illustrated clearly.

The results for the cases with an amplitude error are shown in Fig. 9. For the growing mode, there is little difference in the KE values for the different cases, and all are close to the truth. For the decaying mode, where the second pair of RSVs are more important, the analysis is indeed closer to the true state when the initial observations are at the beginning of the window. When the initial observations are at the end of the window, the analysis increment leads to growth rather than decay because the second pair of RSVs are filtered completely.

The results for the cases with a phase error (Figs. 10 and 11) are somewhat similar. For the growing mode, the phase errors are significantly improved for all three cases as the lower level wave has been shifted to the correct position and the KE values are slightly worse as the upper level wave is not shifted completely. Further, both the phase error and KE values for the growing mode are worse when the initial observations are at the end of the window rather than at the beginning. For the decaying mode, the differences are again more pronounced. With observations at the end of the window, only the first pair of RSVs are included and the KE values begin to grow whilst the streamfunction field eventually tilts westward instead of eastward, giving a negative correlation. With observations at the beginning of the window, the RSVs needed to give decay are filtered less, improving both the KE values and the phase error.

From the equivalence between 4D-Var and the Kalman Filter, 4D-Var implicitly propagates the background error covariance matrix through time. This means that the error covariance at the end of the interval is more sensitive to the flow regime (more flow-dependent) than at the beginning (Thépaut et al., 1996), which suggests that you would expect observations to be more useful at the end of the window. However, these experiments show that the 4D-Var analyses are improved when the initial and final observations are further apart in time. Thus, in these cases, the observations at

17

the beginning of the window are also important.

e. Spatial position of the observations

The impact of the spatial position of the observations is considered by repeating the SVD analysis and 4D-Var experiments with the horizontal line of observations given at different heights. This is achieved by observing the interior non-dimensional buoyancy, which is the vertical derivative of the non-dimensional streamfunction. The observations in each case are given at the beginning and the end of a 12h assimilation window.

1). SVD RESULTS

The singular values of the first and second pairs of RSVs, as a function of the height of the buoyancy observations, are shown in Fig. 12. Due to the symmetry of the of the Eady model, the experiments using observations on the lower boundary are equivalent to those using observations on the upper boundary. Hence, the curves are symmetrical about 5km. As the height of the observations is increased from 0 to 4.5km, the singular values of both the first and second RSVs decrease. The former decreases from 2.99 to 1.64 and the latter from 0.74 to 0.57.

 The SVD shows that we expect the 4D-Var analyses to be closest to the truth when the observations are nearest to the upper or lower boundaries, as these give the largest singular values. The normal modes are driven from the boundaries and their buoyancy fields have their largest amplitudes at the boundaries. Hence observing the boundary buoyancy could be expected to be optimum and the difference between the observations and the background state is largest at the boundaries. Therefore, more accurate information can be extracted in the presence of relatively noisy observations when the observations are closest to the relevant boundary, here the lower one. This is now confirmed with the actual 4D-Var analyses.

2). 4D-VAR ANALYSES

When the true state is given by the growing mode, the variance ratio is specified as $\mu^2 = 10$. With this value, the position of the observations has the most impact on the filtering of the first pair of

RSVs. When the height of the observations is increased, the first pair of RSVs are filtered more. The KE values (Fig. 13a) show that the analysis is indeed closest to the background state when the observations are near to the centre of the domain and close to the true state when the observations are on the lower boundary.

When the true state is given by the decaying mode, the variance ratio is specified as $\mu^2 = 0.1$. With this value, the position of the observations has the most impact on the filtering of the second pair of RSVs. When the height of the observations is increased, the second pair of RSVs are filtered more so that the first pair dominate the solution, leading to growth instead of decay. The KE values (Fig. 13b) show that the best analysis is indeed obtained when the observations are on the lower boundary.

f. Summary

Cases where the true state is given by either a growing or decaying normal mode, and where the background state has either an amplitude or phase error have been considered. The right singular vectors (RSVs) of the 4D-Var observability matrix provide a useful tool to examine how the observational information is projected into the analysis increment. In these cases, there are only two pairs of RSVs that are needed to form the analysis increment. The first pair of RSVs both have large singular values, and lead to a growing analysis increment. The second pair of RSVs have small singular values and are needed to reconstruct the upper level wave and to give a decaying analysis increment.

The 4D-Var algorithm filters the RSVs with relatively small singular values, compared with the size of the observational noise. This implies that 4D-Var preferentially adds an analysis increment that results in growth. Such behaviour means that 4D-Var is extremely efficient in correcting the errors in the initial conditions that would otherwise lead to large forecast errors. However, if the errors in the background state lead to decay, adding an analysis increment that leads to growth is detrimental to the forecast.

4D-Var is able to extract the information about the growth rate of the true state from the time-sequence of observations. This information is contained in the second pair of RSVs and is filtered less when either μ^2 decreases or when their singular value increases. We have examined three ways

in which this can occur. First, as the accuracy of the observations increases, the value of μ^2 decreases. Second, as the time between the initial and final sets of observations increases, the singular value of the second pair of RSVs increases. Third, as the line of buoyancy observations is moved from the centre of the domain to either the upper or lower boundary, the singular values of both the first and second pairs of RSVs increases.

4.   Results considering modal and non-modal growth

The experiments in section 3 are based on background state errors that are given by either a growing or decaying normal mode. Such experiments are useful for understanding the differences in the behaviour of 4D-Var in the presence of growing and decaying errors. However it is possible for an error in the initial conditions, typically associated with interior QGPV anomalies, to exceed the exponential growth of the most rapidly growing normal mode. In this final section of results, the extent to which 4D-Var is capable of correcting such an error is investigated by comparing the behaviour of 4D-Var in cases where the true state is given by either the most rapidly growing normal mode or a PV-dipole perturbation that exhibits non-modal rapid growth.

a.   Experiment Description

The experiments in section 3 only allowed the analysis increments to be added to the boundary buoyancy. The experiments in this section allow the analysis increments to be added to both the boundary buoyancy and the interior QGPV. The background error variance for buoyancy and QGPV can be expected to be different, and this is allowed for. The ratio between the observed and background buoyancy error variance is denoted by μ_b^2 and μ_q^2 is the ratio between the buoyancy observation error variance and the background QGPV error variance. Horizontal background error correlations are specified by the matrices $\mathbf{B}_{bb}$ and $\mathbf{B}_{qq}$, both with length scales $l = 500km$. $\mathbf{B}_{bb}$ is block diagonal, with two horizontal correlation matrices on the diagonal, which correspond to the buoyancy on the upper and lower levels. $\mathbf{B}_{qq}$ is also block diagonal, but with 11 horizontal correlation matrices on the diagonal that correspond to the 11 vertical levels for QGPV. The details of the horizontal correlation

matrices are given in Appendix A. To simplify the experiments, we consider the case where all the background state values are zero, $\mathbf{x}^b = 0$. The cost function is then given by:

$$J(\mathbf{x}_0) = \mathbf{x}_0^T \begin{pmatrix} \mu_b^2 \mathbf{B}_{bb} & 0 \\ 0 & \mu_q^2 \mathbf{B}_{qq} \end{pmatrix} \mathbf{x}_0 + \left(\hat{\mathbf{y}} - \hat{\mathbf{H}} \mathbf{x}_0 \right)^T \left(\hat{\mathbf{y}} - \hat{\mathbf{H}} \mathbf{x}_0 \right) \tag{8}$$

The true state is given by either the most rapidly growing normal mode or a PV-dipole that exhibits non-modal growth, the equations for which are given in Appendix A. The horizontal domain is increased to 8000km, with 80 grid points in the horizontal. Horizontal lines of perfect observations of the non-dimensional buoyancy are given at the beginning and the end of a 6h window at a height of 4.5km. Providing observations in the middle of the spatial domain allows an interior buoyancy anomaly to be well observed.

The ratio μ_q^2/μ_b^2 gives the variance of the non-dimensional boundary buoyancy errors divided by the variance of the non-dimensional interior QGPV errors in the background state. In practice, the size of this ratio will depend on the forecast model used to generate the background state, previous analysis errors and the particular flow situation. The following experiments compare the results of the SVD of the observability matrices and 4D-Var analyses with different specifications for the a priori error variance ratios.

Extreme values are chosen to illustrate clearly the behaviour of 4D-Var. These are summarized in Table 1. Specification 1 uses a large value of μ_q^2/μ_b^2 (10^5). This assumes that we think that the background state errors in the boundary buoyancy field are much larger than those in the interior QGPV field. Specification 3 uses a small value of μ_q^2/μ_b^2 (10^{-5}). This assumes that we think that the background state errors in the buoyancy field are much smaller than those in the QGPV field. Specification 2 uses values that are a compromise of the two extremes ($\mu_q^2/\mu_b^2 = 5 \times 10^{-2}$).

b. SVD results

The observability matrix corresponding to the cost function given by (8) can be written as:

$$\widetilde{\mathbf{H}} = \widehat{\mathbf{H}} \begin{pmatrix} \mu_b \mathbf{B}_{bb}^{1/2} & 0 \\ 0 & \mu_q \mathbf{B}_{qq}^{1/2} \end{pmatrix} \tag{9}$$

We now examine the RSVs of the observability matrices for different specifications of the error variance ratios. We only examine the RSVs that have large values of c_j when the background state errors are given by a growing mode and the observations are perfect. It should be noted that for the non-modal case, there are many values of c_j that are non-zero and hence there are many RSVs with a variety of spatial scales that contribute to the analysis increments. However, for our purposes, it suffices to examine the structures of only a few of these RSVs.

The RSVs for the case with specification 1 ($\mu_q^2/\mu_b^2 = 10^5$) are shown in Fig. 14a-b. The interior QGPV has a relatively small magnitude whilst the buoyancy field has a relatively large magnitude. The buoyancy field for RSV4 exhibits an eastward tilt with height and that for RSV10 exhibits a westward tilt, similar to the RSVs in section 3. Such structures lead to modal growth and are therefore likely to be appropriate for correcting such errors. The RSVs for the case with specification 3 ($\mu_q^2/\mu_b^2 = 10^{-5}$) are shown in Fig. 14c-d. The interior QGPV now has a relatively large magnitude in comparison to the buoyancy field. Such structures lead to non-modal growth through PV-unshielding and are therefore likely to be appropriate for correcting such errors.

c. 4D-var analyses

We now consider the impact of different error variance specifications on analyses where the background state errors are either given by modal growth or non-modal growth. The analyses are verified by comparing the spatial structure at the beginning of the window, and also the exponential growth rate, which is defined as:

$$\sigma_{KE}(t) = \frac{1}{2\Delta t} ln \left(\frac{KE_{t+\Delta t}}{KE_{t-\Delta t}} \right) \tag{10}$$

where Δt is the timestep.

The spatial structures of the analyses are shown at the beginning of the window. We first consider cases where the true state is given by the growing normal mode. The true state, shown in Fig. 15a, is given by buoyancy waves and zero interior QGPV. This gives an exponential KE growth rate, shown in Fig. 16a, that is constant throughout the forecast period. The analyses using the three different variance ratio specifications are shown in Fig. 15. For specification 1, a large analysis increment (amplitude 3.5) is added to the buoyancy fields and a small analysis increment (amplitude 10^{-4}) is added to the interior QGPV. This gives an analysis that looks similar to the true state at the initial time, and the KE growth rate values (shown in Fig. 16a) are also similar. For specification 2, a smaller analysis increment (amplitude 2.4) is added to the buoyancy field and a larger analysis increment (amplitude 4.1) is added to the QGPV. As the analysis has a large amplitude in the interior, the increment grows by the PV-unshielding mechanism, leading to a larger growth rate than that of the true state. For specification 3, the structure of the RSVs are such that 4D-Var adds an even larger analysis increment (amplitude 8.5) to the interior QGPV, leading to an even larger growth rate.

Finally, we consider cases where the true state is given by a PV-dipole perturbation that exhibits rapid finite-time non-modal growth. The true state, shown in Fig. 17a, corresponds to a positive temperature anomaly which may have resulted from, for example, diabatic heating. As discussed by Badger and Hoskins (2001), the anomaly evolves by the PV-unshielding mechanism to give rapid finite time growth which peaks at 6h, as shown in Fig. 16b. The analyses using the three different variance ratio specifications are shown in Fig. 17. For specification 1, a large analysis increment (amplitude 3.7) is added to the buoyancy fields and a small analysis increment (amplitude 7×10^{-4}) is added to the QGPV fields. This results in both an extremely unrealistic analysis and growth rate. Although the time-evolution information is contained in the observations, it is projected onto the inappropriate RSV structures. For specification 2, a larger analysis increment (amplitude 2.9) is added to the interior QGPV field, giving an analysis that has a similar structure to the true state. The growth rate also has a similar behaviour to that of the true state, but never exceeds that of the most unstable normal mode. For specification 3, the observational information can be projected onto the appropriate RSV structures. A larger analysis increment (amplitude 5.9) is added to the interior QGPV and a smaller analysis increment is added to the boundaries so that in the analysis,

the buoyancy anomaly has a small vertical scale, similar to the true anomaly. The growth rate now achieves a value that exceeds that of the fastest growing normal mode. However, it still does not reach the maximum value of the true state and it is likely that many more accurate observations are required to produce a growth rate that is close to that of the true state.

d.   Summary

We have compared cases where the background state errors exhibit either modal or non-modal growth and where there are different specifications for the background error variances for the interior QGPV and the boundary buoyancy.

The RSVs of the 4D-Var observability matrix illustrated that the possible analysis increment structures are pre-determined by the model, observation locations, and the covariances, without the knowledge of the observed values. When the specified boundary buoyancy variance is relatively large, the RSVs have maxima on the boundaries. Such structures result in modal growth. When the specified interior QGPV variance is relatively large, the RSVs have maxima in the interior. Such structures result in non-modal growth.

With an inappropriate specification of the error variance ratios, the observational information is projected onto the inappropriate structures. Although the observations may contain information that can be used to infer that the true state is for example a PV-dipole, this may be mistakenly filtered by the 4D-Var algorithm if the error covariances are specified inappropriately. The growth rates show clearly that this can lead to poor forecasts, especially for the case where the true state exhibits non-modal growth. Hence the specification of the background error covariance is crucial to enable the benefits of 4D-Var to be maximized.

5.   Conclusions

In this paper we have used the singular value decomposition (SVD) of the 4D-Var observability matrix to investigate the extent to which 4D-Var is able to use the information about the temporal development of idealized baroclinic weather systems to infer the vertical spatial structure. Simple

cases using the 2D Eady model with a single horizontal line of buoyancy observations at two times during the assimilation period have been investigated. The SVD provides a general framework, without the need for repeating numerous 4D-Var identical twin experiments. However, to confirm the anticipated results from the SVD, we have also examined many 4D-Var analyses and compared the evolution of the KE values and the correlation of the streamfunction field throughout the assimilation window and following forecasts.

Three main conclusions from this work can be drawn. First, we have shown that 4D-Var preferentially generates an analysis increment that leads to growth, but a decaying analysis increment can be generated provided that the information about the evolution can be extracted from the observations. Second, the ability of 4D-Var to extract the valuable time-evolution information can be maximized by adjusting the locations of the observations in space and time. Third, the specification of the appropriate background error covariances is crucial in being able to generate analysis increments that lead to the appropriate growth rate.

6. Discussion

The fact that 4D-Var is efficient in correcting rapidly growing errors is not a new result. Previous studies by Thépaut et al. (1996); Pires et al. (1996) and Rabier et al. (1996) demonstrated that a 4D-Var algorithm with no background term is able to use observations given at the end of the assimilation window to correct such errors. However, this paper has extended this conclusion to show that 4D-Var is more efficient in correcting such errors when information about the time-evolution of the observed system can be inferred. Further, we have shown that in the case where the actual errors are decaying with time, it is possible for the observations to be projected onto analysis increment structures that lead to growth instead of decay; such increments are detrimental to the forecast. An increment that leads to decay can be generated provided there is accurate information about the growth rate of the system, but otherwise 4D-Var generates an increment that leads to growth. So not only is 4D-Var efficient in correcting growing errors, it is also efficient in adding growing errors. The cases where 4D-Var corrects such errors most likely outweigh the cases where 4D-Var adds these errors. However, there are likely to be a few occasions where 4D-Var adds rapidly growing errors to the background

state, resulting in a worse forecast than if observations had not been assimilated.

The benefits of 4D-Var can be maximized by placing the observations in the optimal locations in space and time. Such adaptive observation strategies have previously been considered (Snyder, 1996), and include placing the observations in regions that are sensitive to error growth (Buizza and Montani, 1999) or where the observations have a large impact on a measure of the forecast error (Bishop and Toth, 1999). This paper suggests that an alternative strategy could be based on optimizing the information that can be extracted by the 4D-Var algorithm. Further, the 4D-Var observability matrix enables us to consider the optimal location for observations in both time and space (Daescu and Navon, 2004), and takes into account both the model dynamics and error covariances.

This work has also highlighted the importance of using case-dependent background error variances, so that the maximum amount of information can be extracted from the observations and so that the algorithm can generate the modal and non-modal growth structures. Inappropriate error variances can lead to useful information being mistakenly filtered from the solution, or information can be projected onto the inappropriate structures. Case-dependent covariances may be achieved using methods such as error-breeding (Toth and Kalnay, 1997), or alternatively by computing the appropriate values from the data (Dee, 1995; Gong et al., 1998; Desroziers and Ivanov, 2001).

Definitions of the true state initial conditions and background error correlation matrices

The normal modes are defined (e.g. Hoskins and Bretherton, 1972) as:

$$\psi(x,z) = \cosh(kz)\cos(kx) - \alpha\sinh(kz)\sin(kx), \qquad \text{Growing Mode} \qquad \text{(A.1a)}$$

$$\psi(x,z) = \cosh(kz)\cos(kx) + \alpha\sinh(kz)\sin(kx), \qquad \text{Decaying Mode} \qquad \text{(A.1b)}$$

where $k = 1.6$, $\alpha = 2.59$ and the interior QGPV is zero. The interior QGPV-dipole perturbation consists of two spatially confined vortices by defining $q(x,z) = f(x)g(z)$, where

$$f(x) = \begin{cases} -\frac{L}{4} - \frac{x}{2} + \frac{\sin(2kx)}{4k} & -\frac{L}{2} < x < -\frac{L}{4} \\ -\frac{L}{8} - \frac{\cos(kx)}{k} & -\frac{L}{4} < x < \frac{L}{4} \\ -\frac{L}{4} + \frac{x}{2} - \frac{\sin(2kx)}{4k} & \frac{L}{4} < x < \frac{L}{2} \end{cases} \qquad \text{(A.2a)}$$

$$g(z) = \cos^2(mz)\sin(mz) \quad 2 < z < 8km \qquad \text{(A.2b)}$$

with $L = 800km$, $k = 2\pi/L$, $m = \pi/0.8$ and both $f(x)$ and $g(z)$ are zero outside the specified ranges.

The background error correlation matrices are defined by assuming that the background state errors are correlated in the horizontal, but that there are no correlations between vertical levels. The matrix $\mathbf{B}$ is block-diagonal with matrices $\boldsymbol{\rho}$ on the diagonal, which define the horizontal correlation between variables on one vertical level. The correlations are specified by defining the inverse correlation matrix, as in JHN:

$$\boldsymbol{\rho}^{-1} = \gamma\left(\mathbf{I} + \frac{l^4}{2}(\mathbf{L}_{xx})^2\right). \qquad \text{(A.3)}$$

$\mathbf{L}_{xx}$ is a finite difference second derivative matrix in the x direction which incorporates periodic boundary conditions, l is the horizontal correlation length-scale, and γ is a scalar parameter that is specified so that diagonal elements of $\boldsymbol{\rho}$ have a value of one.

APPENDIX B

On the use of perfect observations

The experiments in this paper use perfect observations rather than observations with errors, but still include the effect of filtering provided by the background term in the cost function. We now describe, first theoretically and then with experiments, why these results are relevant to assimilation with real data. The effects of observational noise and filtering by 4D-Var are described in further detail in JHN.

Following from (4), when imperfect observations $\hat{\mathbf{y}}^\varepsilon = \hat{\mathbf{H}}\mathbf{x}^t + \boldsymbol{\varepsilon}$ are assimilated, the analysis increments can be written in terms of components from the true signal $\hat{\mathbf{H}}\mathbf{x}^t$ and the observational noise $\boldsymbol{\varepsilon}$.

$$\mathbf{B}^{-1/2}(\mathbf{x}^a - \mathbf{x}^b) = \sum_j f_j c_j^t \mathbf{v}_j + \sum_j f_j c_j^\varepsilon \mathbf{v}_j \tag{B.1a}$$

where

$$c_j^t = \mathbf{u}_j^T \mathbf{R}^{-1}(\hat{\mathbf{H}}\mathbf{x}^t - \hat{\mathbf{H}}\mathbf{x}^b)/\lambda_j \tag{B.1b}$$

$$c_j^\varepsilon = \mathbf{u}_j^T \mathbf{R}^{-1}\boldsymbol{\varepsilon}/\lambda_j \tag{B.1c}$$

Typically, the observational noise has a large projection, c_j^ε, onto the RSVs with small spatial scales and associated with small singular values, whilst the true signal, c_j^t, has a large projection onto the RSVs with large spatial scales and associated with large singular values. This is illustrated in Fig.B.1. The role of the filter factor, f_j, is to filter the contribution from the noise whilst retaining the contribution from the true signal. When the appropriate value for the variance ratio is specified, the contribution from the noise, $\sum_j f_j c_j^\varepsilon \mathbf{v}_j$ (the second term in equation B.1a), should be close to zero so that the equations for perfect and imperfect observations are almost identical. This means that the analysis with perfect observations should be close to the analysis with imperfect observations. Thus the relevance of perfect observations relies on the assumption that the observational noise is projected onto the RSVs with small singular values and therefore that the noise is filtered from the

solution. It is only with this same assumption that variational assimilation methods can be of use.

The analysis with perfect observations does account for the fact that the observational noise is filtered, because the appropriate amount of filtering is also applied to the true signal. This is very different to a similar analysis with perfect observations but with no background term and hence no filtering.

There may be some small differences between the analyses for perfect and imperfect observations as it is possible for the observational noise to project onto the RSVs with large singular values. These small differences will depend on the actual observational errors and are likely to become important for a series of assimilation windows in which the forecasts become the background states for the next analysis. However, for a single assimilation window it is unlikely that these differences have a significant impact on the results. Further, as these differences depend on the structure of the observational noise, an ensemble of experiments, with different values for the random noise, would be needed to deduce concrete conclusions. The use of perfect observations eliminates this need.

To finally confirm that the results with perfect observations are almost equivalent to those with imperfect observations, we compare an analysis that uses perfect observations with an analysis that uses noisy observations. The true state is given by a growing mode and the background state has a phase error. The case with noisy observations has random noise added to the perfect observations that has a Gaussian distribution with standard deviation 1.5. The appropriate value of μ^2, and hence the appropriate amount of filtering, is found by repeating the analyses with different values of μ^2 and selecting the case where the value of the cost function at the minimum is equal to the number of observations (Talagrand, 1998). This gives an appropriate value of $\mu^2 = 3.5$. The analysed wave, shown at the beginning of the window in Fig. B.2a-b, is close to the true state on the lower boundary and mid-way between the background state and true state on the upper boundary. The case for perfect observations again uses $\mu^2 = 3.5$. The analysis (Fig. B.2c-d) is almost identical to that for the perfect observations, with the analysed wave close to the true state on the lower boundary and mid-way on the upper boundary. There are some slight differences, which depend on the values of the observational noise.

References

Badger, J. and B. J. Hoskins, 2001: Simple initial value problems and mechanisms for baroclinic growth. *J. Atmos. Sci.*, **58**, 38–49.

Bishop, C. H. and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.*, **56**, 1748–1765.

Buizza, R. and A. Montani, 1999: Targeting obsevations using singular vectors. *J. Atmos. Sci.*, **56**, 2965–2985.

Buizza, R. and T. N. Palmer, 1995: The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52**, 1434–1456.

Charney, J. G., 1947: The dynamics of long waves in a baroclinic westerly current. *Journal of Meteorology*, **4**, 135–162.

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

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.

Daescu, D. N. and I. M. Navon, 2004: Adaptive observations in the context of 4d-var data assimilation. *Meteorol. Atmos. Phys.*, **85**, 205–226.

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

Desroziers, G. and S. Ivanov, 2001: Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Q. J. R. Meteorol. Soc.*, **127**, 1433–1452.

Eady, E. T., 1949: Long waves and cyclone waves. *Tellus*, **1**, 33–52.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *J. Geophys. Res.-Oceans*, **99**, 10143–10162.

Farrell, B. F., 1982: The initial growth of disturbances in a baroclinic flow. *J. Atmos. Sci.*, **39**, 1663–1686.

— 1984: Modal and non-modal baroclinic waves. *J. Atmos. Sci.*, **41**, 668–673.

— 1989: Optimal excitation of baroclinic waves. *J. Atmos. Sci.*, **46**, 1193–1206.

Gong, J. J., G. Wahba, D. R. Johnson, and J. Tribbia, 1998: Adaptive tuning of numerical weather prediction models: simultaneous estimation of weighting, smoothing and physical parameters. *Mon. Weather Rev.*, **126**, 210–231.

Hoskins, B. J. and F. P. Bretherton, 1972: Atmospheric frontogenesis models: Mathematical formulation and solution. *J. Atmos. Sci.*, **29**, 11–37.

Johnson, C., 2003: *Information Content of Observations in Variational Data Assimilation*. Ph.D. thesis, University of Reading.

Johnson, C., B. J. Hoskins, and N. K. Nichols, 2005a: A singular vector perspective of 4d-var: Filtering and interpolation. *Q. J. R. Meteorol. Soc.*, **131**, 1–20.

Johnson, C., N. K. Nichols, and B. J. Hoskins, 2005b: Very large inverse problems in atmosphere and ocean modelling. *Int. J. Numer. Meth. Fluids*, **47**, 759–771.

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

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177–1194.

— 2003a: Modelling of error covariances by 4d-var data assimilation. *Q. J. R. Meteorol. Soc.*, **129**, 3167–3182.

— 2003b: The potential of the ensemble Kalman filter for NWP-a comparison with 4d-var. *Q. J. R. Meteorol. Soc.*, **129**, 3183–3203.

Pires, C., R. Vautard, and O. Talagrand, 1996: On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus*, **48A**, 96–121.

Rabier, F. and P. Courtier, 1992: Four-dimensional assimilation in the presence of baroclinic instability. *Q. J. R. Meteorol. Soc.*, **118**, 649–672.

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

Rabier, F., E. Klinker, P. Courtier, and A. Hollingsworth, 1996: Sensitivity of forecast errors to initial conditions. *Q. J. R. Meteorol. Soc.*, **122**, 121–150.

Snyder, C., 1996: Summary of an informal workshop on adaptive observations and FASTEX. *Bull. Amer. Meteor. Soc.*, **77**, 953–961.

Talagrand, O., 1998: A posteriori evaluation and verification of analysis and assimilation algorithms. *Diagnosis of data assimilation systems*, ECMWF, Reading, UK, Workshop Proceedings, 17–28.

Thépaut, J.-N., P. Courtier, G. Belaud, and G. Lemaître, 1996: Dynamical structure functions in 4D variational assimilation: A case study. *Q. J. R. Meteorol. Soc.*, **122**, 535–561.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Weather Rev.*, **131**, 1485–1490.

Toth, Z. and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Weather Rev.*, **125**, 3297–3319.

List of Tables

Specification	μ_q^2	μ_b^2	μ_q^2/μ_b^2
1	1	10^{-5}	10^5
2	5×10^{-4}	10^{-2}	5×10^{-2}
3	10^{-5}	1	10^{-5}

Table 1: Three different specifications of the variance ratios for the interior QGPV, μ_q^2, and for the buoyancy on the boundaries, μ_b^2, for use in 4D-Var experiments.

List of Figures

Figure 1: 4D-Var analyses for the cases where the background state has an amplitude error and the true state is given by the most rapidly (a)-(b) growing and (c)-(d) decaying mode. The analysis ($\mathbf{x}^a$, solid), background state ($\mathbf{x}^b$, dashed) and true state ($\mathbf{x}^t$, dotted) fields are all shown at the middle of the time window. The upper panels (a) and (c) show the buoyancy on the upper boundary and the lower panels (b) and (d) show the buoyancy on the lower boundary.
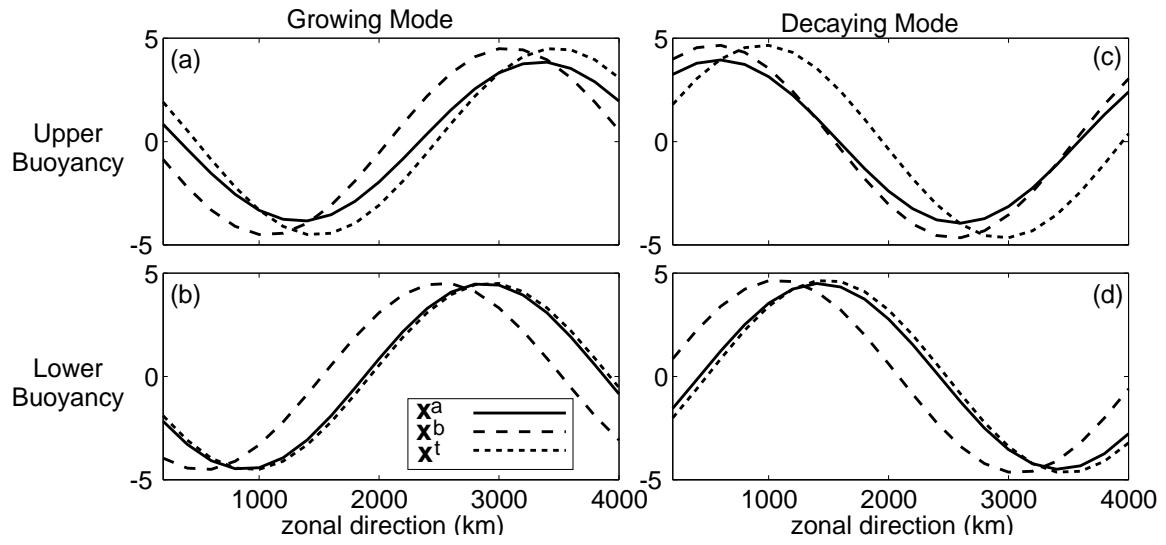
**Figure 2:** 4D-Var analyses for the cases where the background state has a phase error and the true state is given by the most rapidly (a)-(b) growing and (c)-(d) decaying mode. The details are as for Fig. 1.
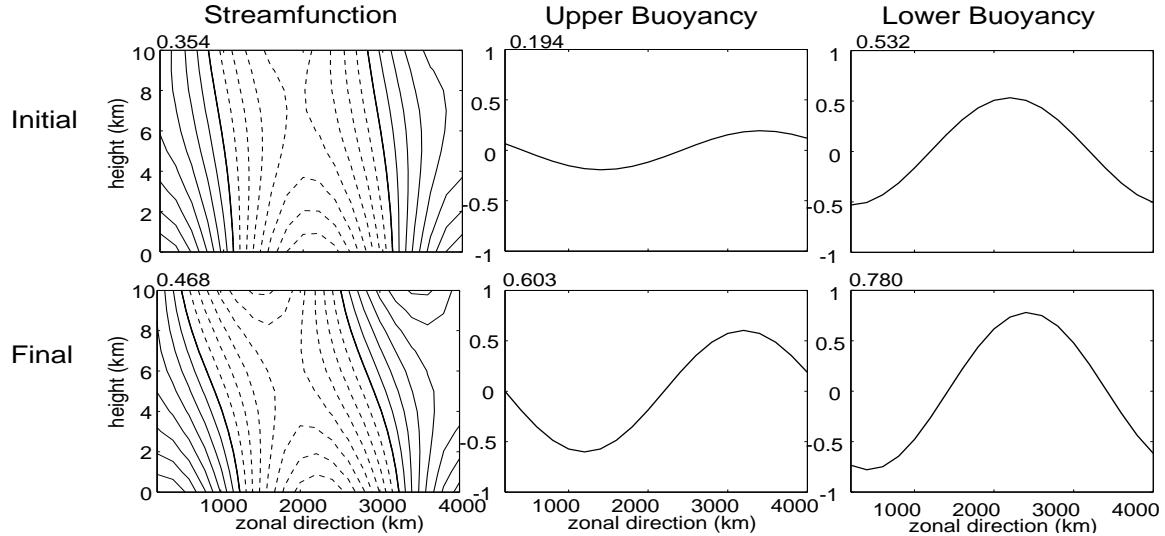
Figure 3: The top panels show the second RSV, $\mathbf{v}_2$, of the observability matrix, premultiplied by the square root of the background error correlation matrix, $\mathbf{B}^{1/2}$. The lower panels show the result of integrating these fields by the Eady model over a 12h interval. The numbers at the top left of each plot indicates the maximum magnitude of each plotted field.
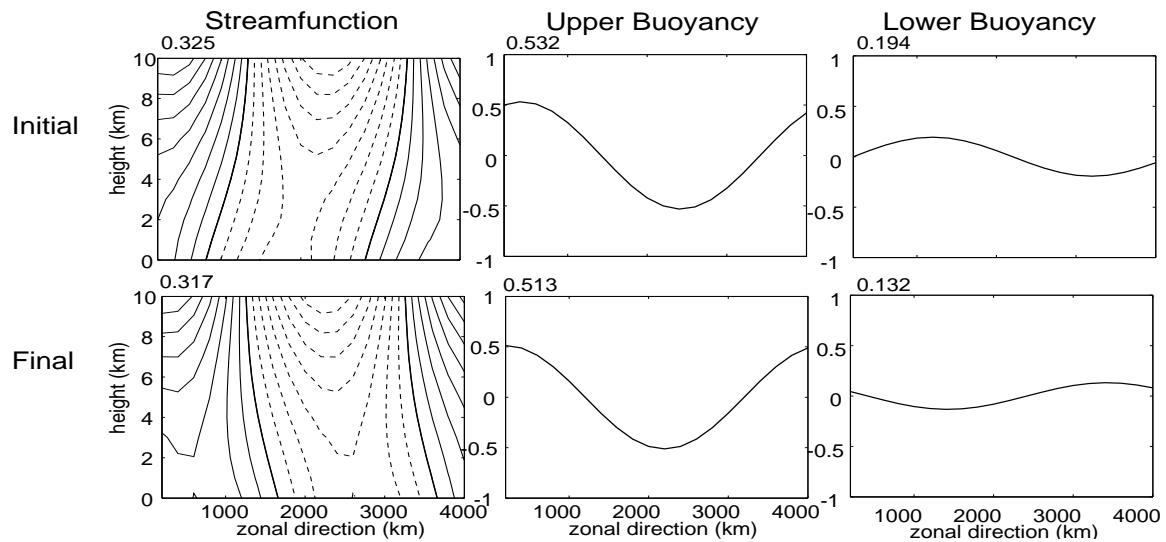
Figure 4: *The structure of the sixth RSV, ($\mathbf{B}^{1/2}\mathbf{v}_6$). The details are as for Fig. 3.*
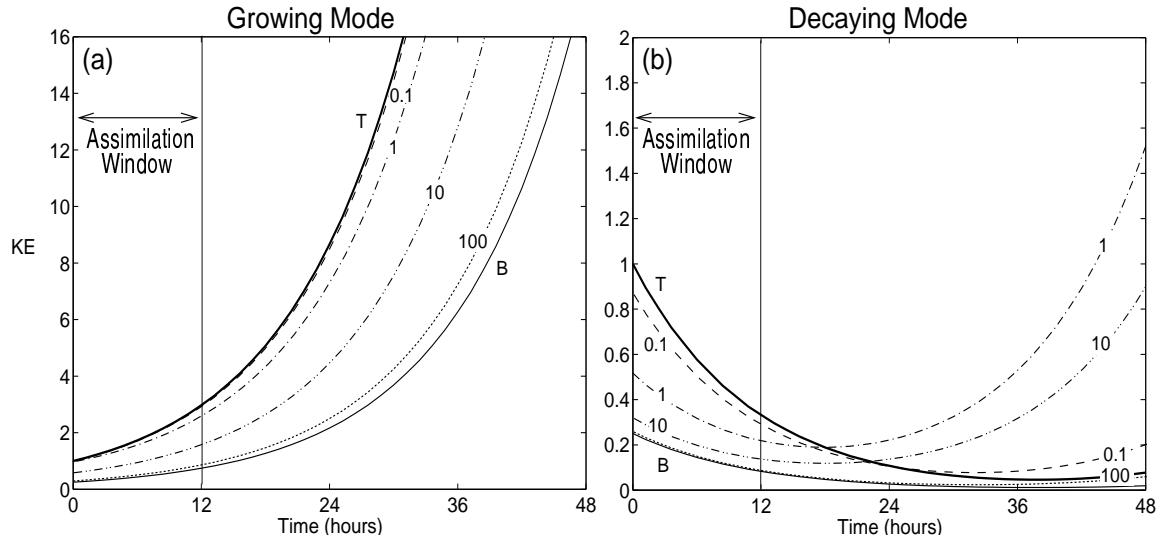
Figure 5: The evolution of the kinetic energy (KE) for the cases where the background state has an amplitude error and the true state is given by the most rapidly (a) growing and (b) decaying mode. The true state KE is shown by the thick solid line (T) and the background state KE is shown by the thin solid line (B). The analyses have specified error variance ratios μ^2 of 0.1 (dash), 1 (dot-dash), 10 (dot-dot-dash) and 100 (dot).
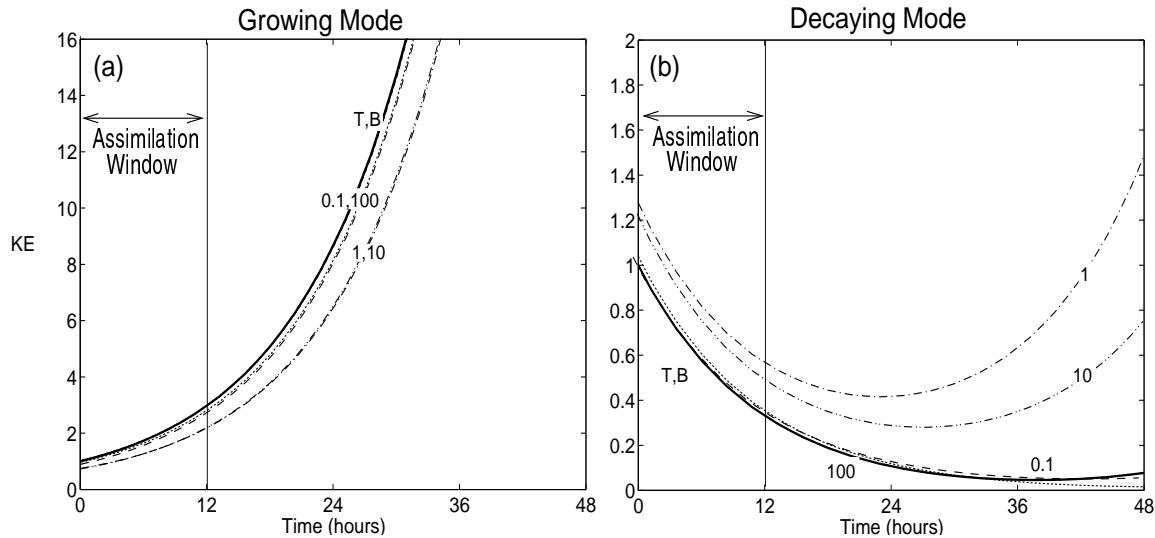
Figure 6: The evolution of the KE for the cases where the background state has a phase error and the true state is given by the most rapidly (a) growing and (b) decaying mode. The details are as for Fig. 5.
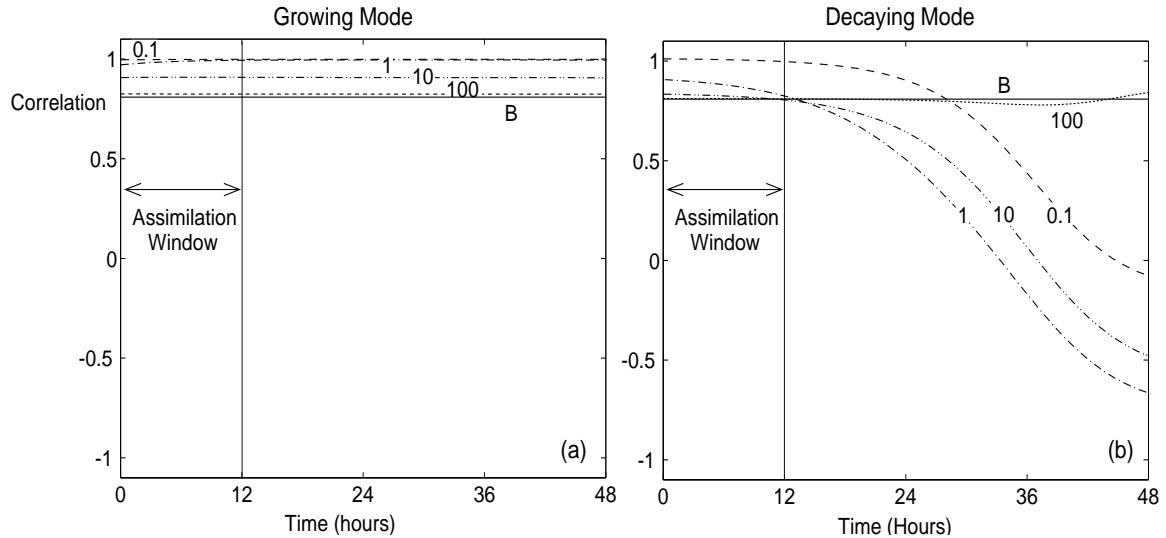
Figure 7: The evolution of the error correlation for the cases where the background state has a phase error and the true state is given by the most rapidly (a) growing and (b) decaying mode. The background state correlation is shown by the solid line (B). The analyses have specified error variance ratios μ^2 of 0.1 (dash), 1 (dot-dash), 10 (dot-dot-dash) and 100 (dot).
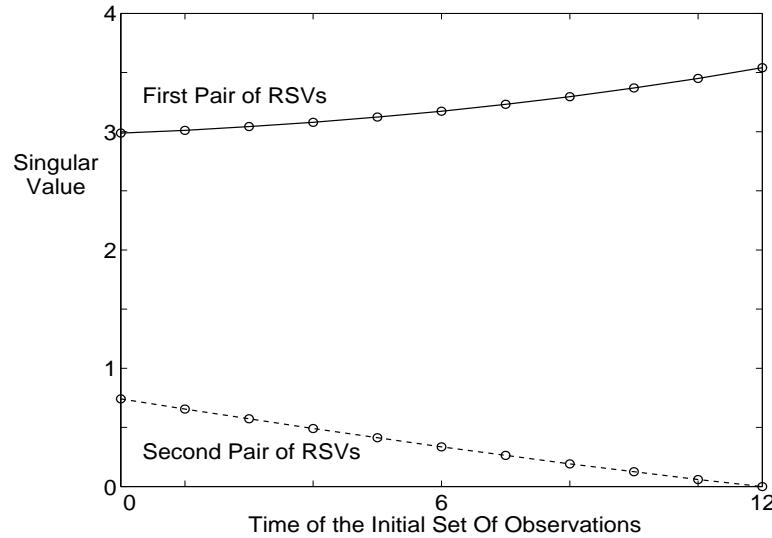
Figure 8: The singular values of the observability matrix that correspond to the first (solid) and second (dashed) pairs of RSVs that contribute to the analysis increment, plotted against the time of the initial observations. In all cases, the final set of observations are given at T+12.
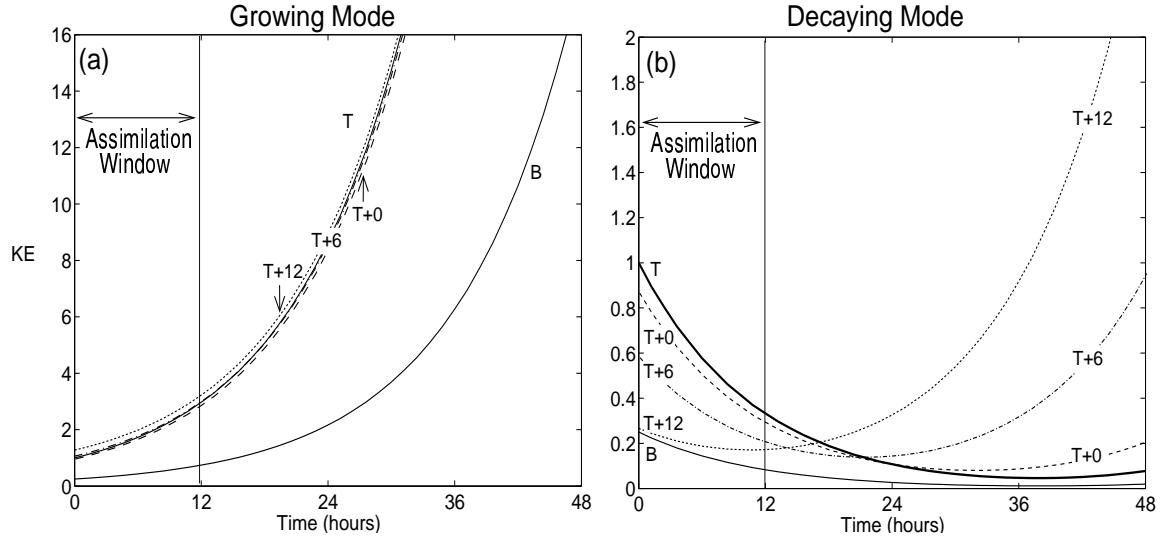
Figure 9: The evolution of the KE for the cases where the background state has an amplitude error and the true state is given by the most rapidly (a) growing and (b) decaying mode. The first set of observations are either at the beginning (T+0, dash), the middle (T+6, dot-dash) or the end (T+12, dot) and the second set are always at the end of the window. The true state KE is shown by the thick solid line (T) and the background state KE is shown by the thin solid line (B).
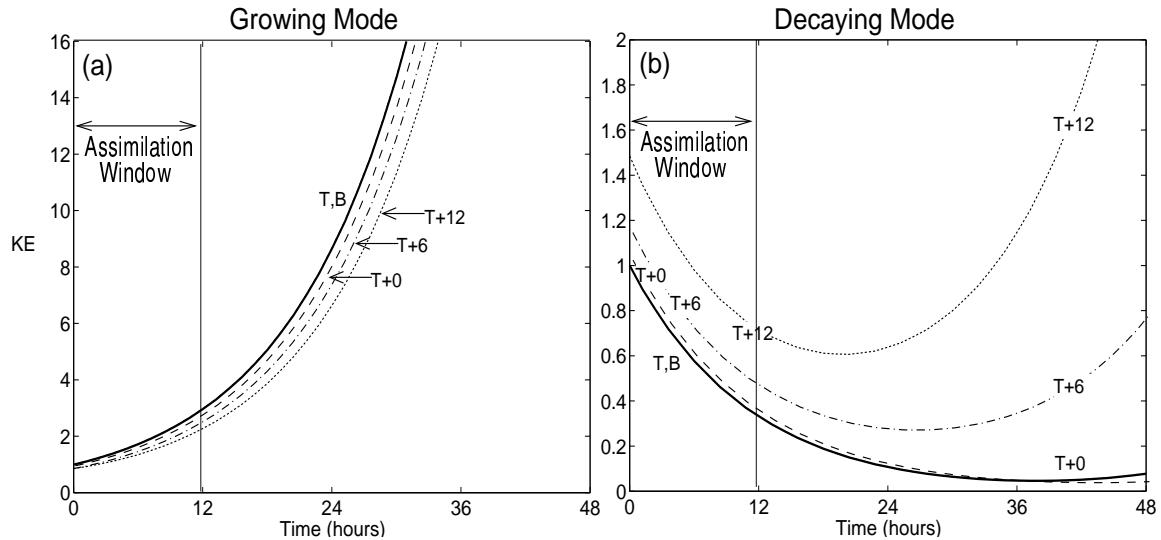
Figure 10: The evolution of the KE for the cases where the background state has a phase error and the true state is given by the most rapidly (a) growing and (b) decaying mode. The details are as for Fig. 9.
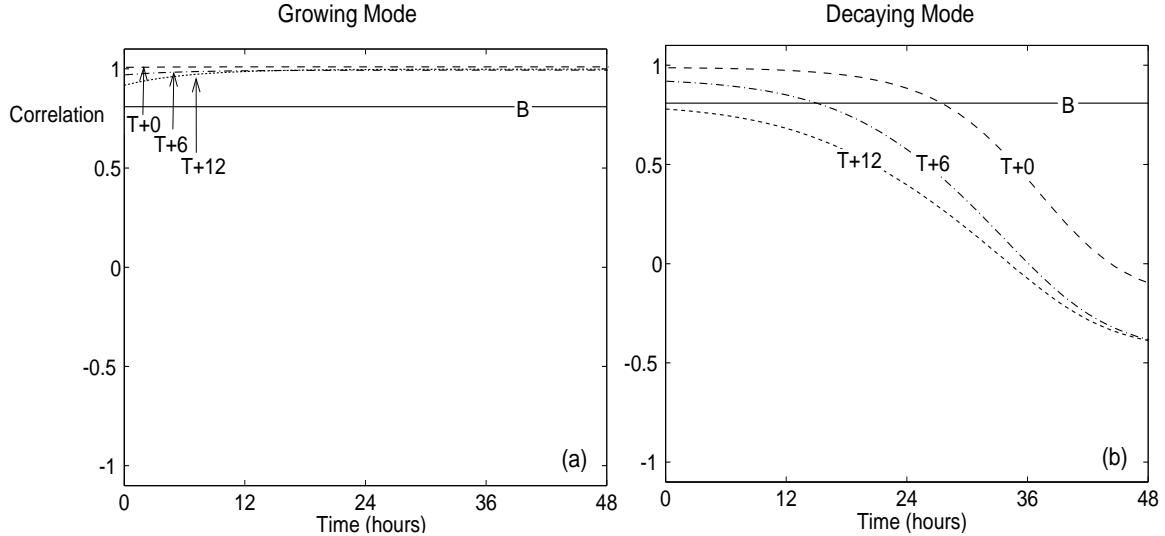
Figure 11: The evolution of the error correlation for the cases where the background state has a phase error and the true state is given by the most rapidly (a) growing and (b) decaying mode. The first set of observations are either at the beginning (T+0, dash), the middle (T+6, dot-dash) or the end (T+12, dot) and the second set are always at the end of the window. The background state error correlation is shown by the solid line (B).
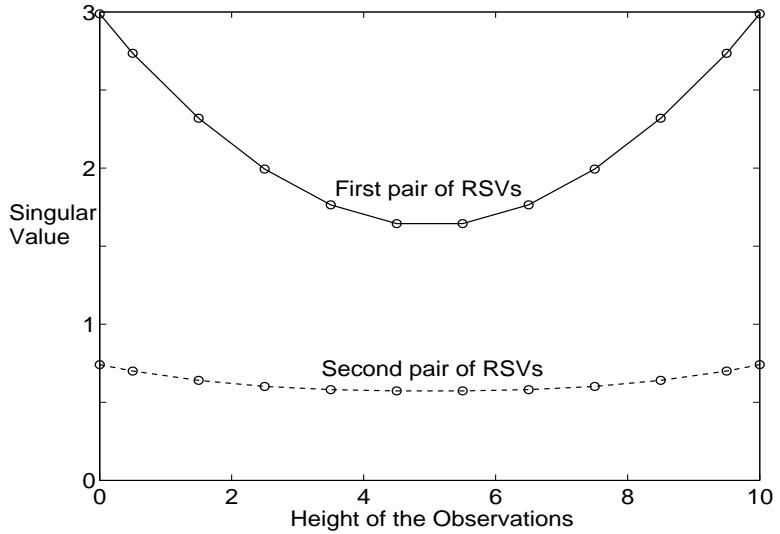
Figure 12: The singular values of the observability matrix that correspond to the first (solid) and second (dashed) pairs of RSVs that contribute to the analysis increment, plotted against the height of the horizontal line of observations.
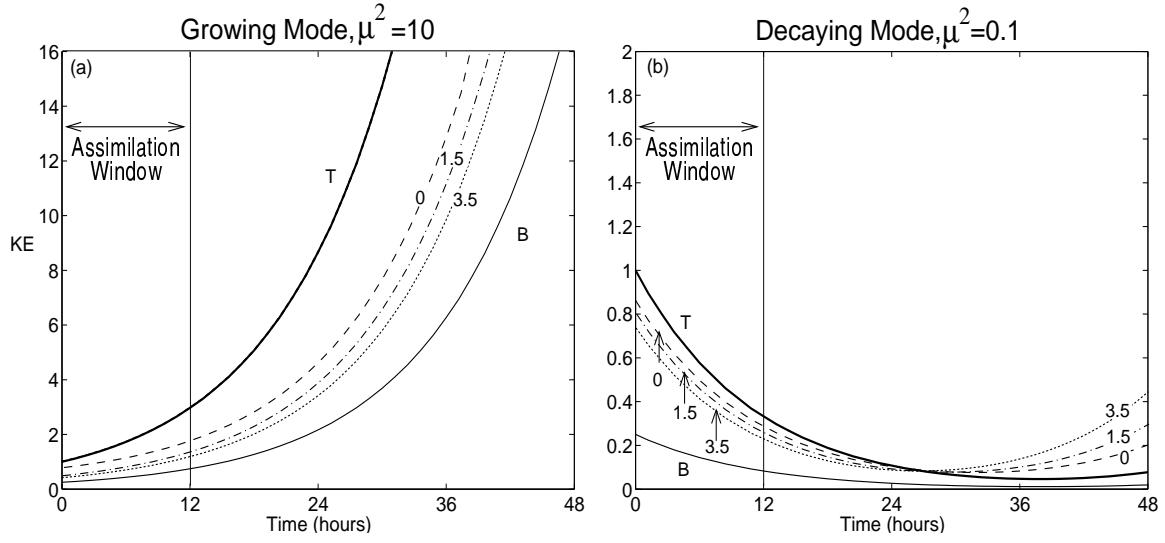
Figure 13: The evolution of the KE for the cases where the background state has an amplitude error and (a) the true state is given by the growing mode and the specified error variance μ^2 is 10, and (b) the true state is given by the decaying mode and the specified error variance μ^2 is 0.1. The horizontal line of observations is given at heights of 0km (dash), 1.5km (dot-dash) and 3.5km (dot).
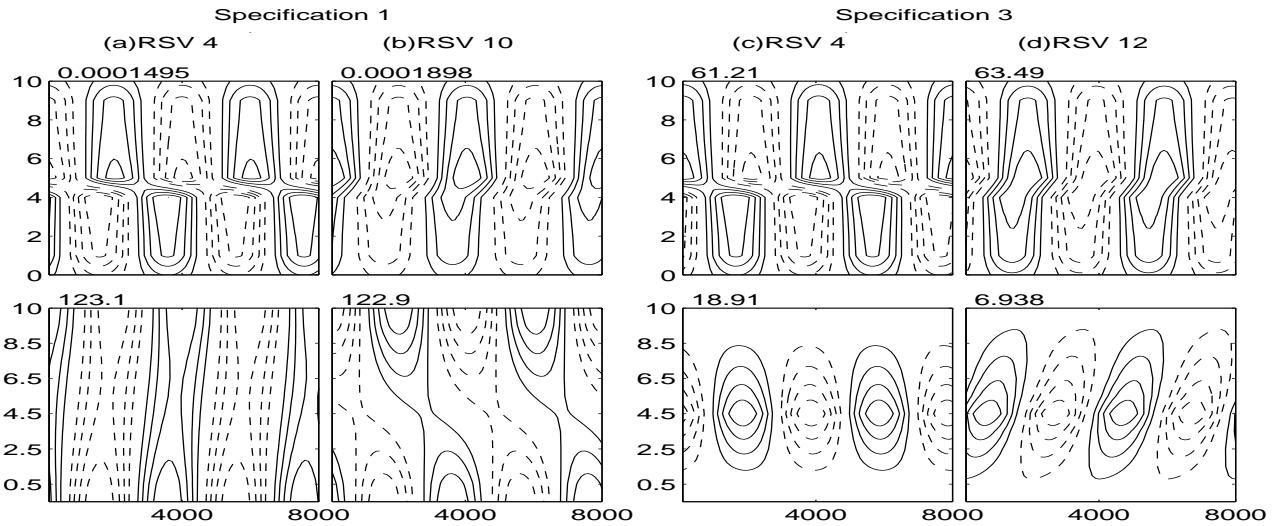
Figure 14: The structure of the non-dimensional QGPV (upper panels) and buoyancy (lower panels) for selected RSVs $\mathbf{B}^{1/2}\mathbf{v}$ of the observability matrices where the variance ratios are (a)-(b) Specification 1, and (c)-(d) Specification 3. These have corresponding singular values, (a) $\lambda_4 = 802$, (b) $\lambda_{10} = 213$, (c)$\lambda_4 = 194$ and (d)$\lambda_{12} = 56$. The values at the top left of each panel denote the maximum magnitude of the field.
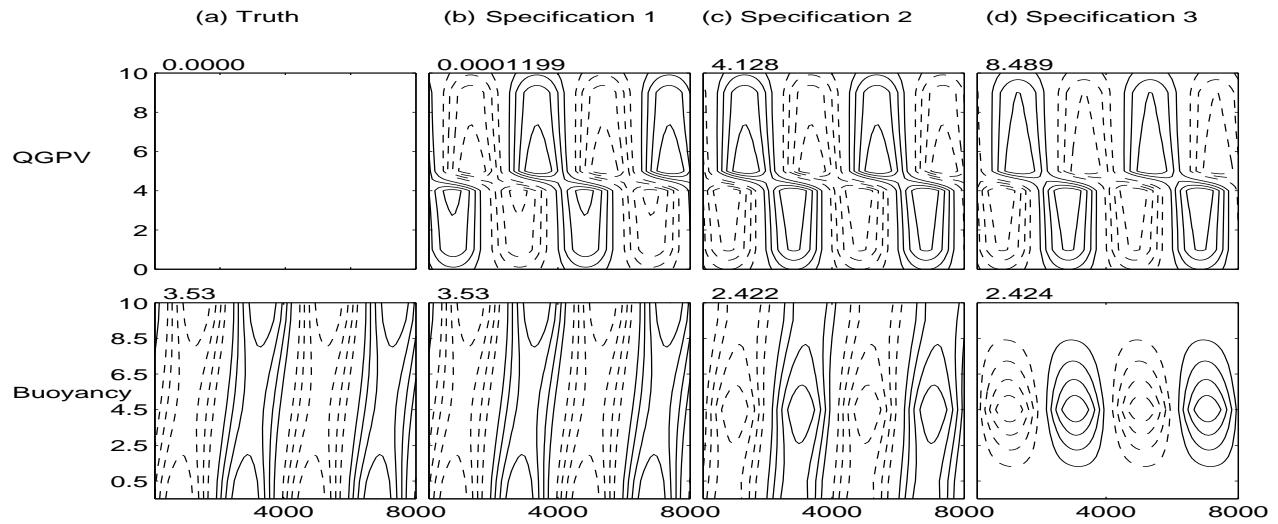
Figure 15: 4D-Var analyses non-dimensional QGPV (upper) and buoyancy (lower) for the cases where the true state is given by the most rapidly growing mode (shown in (a)). The specified error variance ratios are: (b) Specification 1, (c) Specification 2 (d) Specification 3.
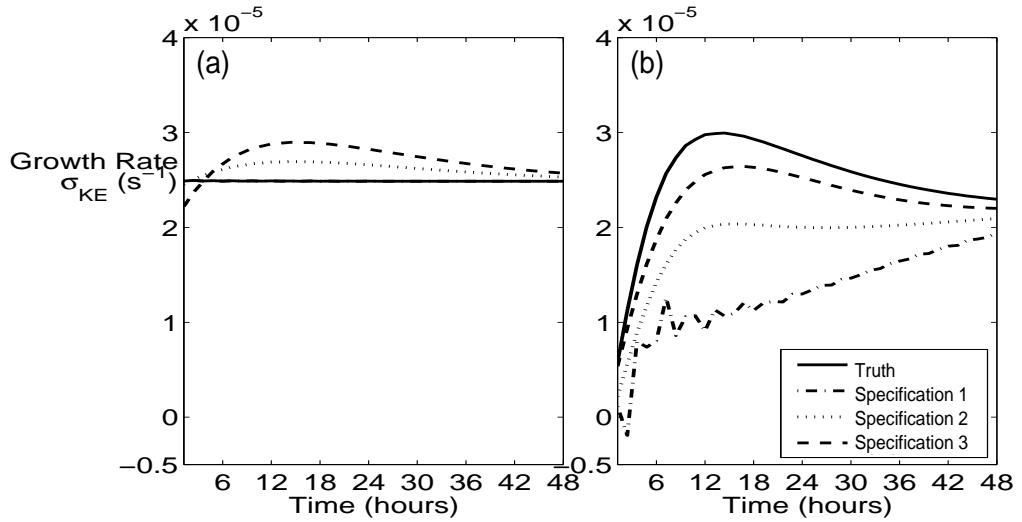
Figure 16: The evolution of the KE growth rates for the cases where the background state values are all zero and the true state is given by (a) the most rapidly growing mode and (b) a PV-dipole perturbation. The true state is shown by the solid line and the other lines show the analyses using the different specifications for the variance ratios.
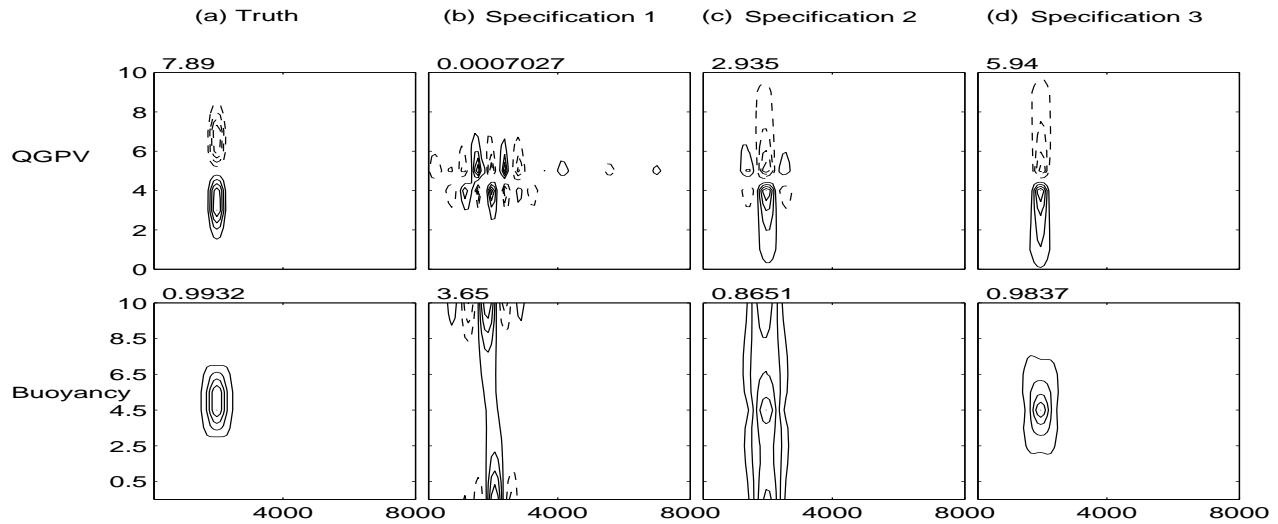
Figure 17: 4D-Var analyses non-dimensional QGPV (upper) and buoyancy (lower) for the cases where the true state is given by an interior QGPV dipole perturbation (shown in (a)) that exhibits rapid finite-time non-modal growth. The details are as for Fig. 15.
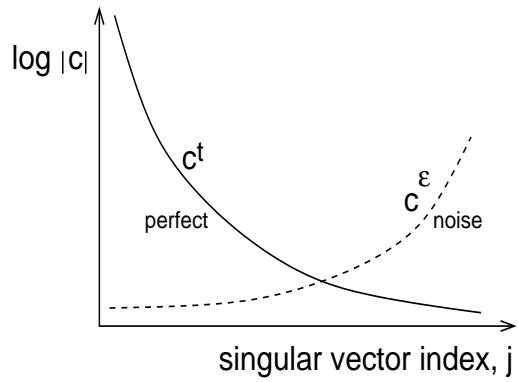
Figure B.1: Illustration of typical Picard ratios, $log|c_j|$, for the true signal and the noise. The Picard ratio for the true signal, $log|c_j^t|$, decreases with increasing singular vector index, j. The Picard ratio for the error, $log|c_j^\varepsilon|$, increases with j.
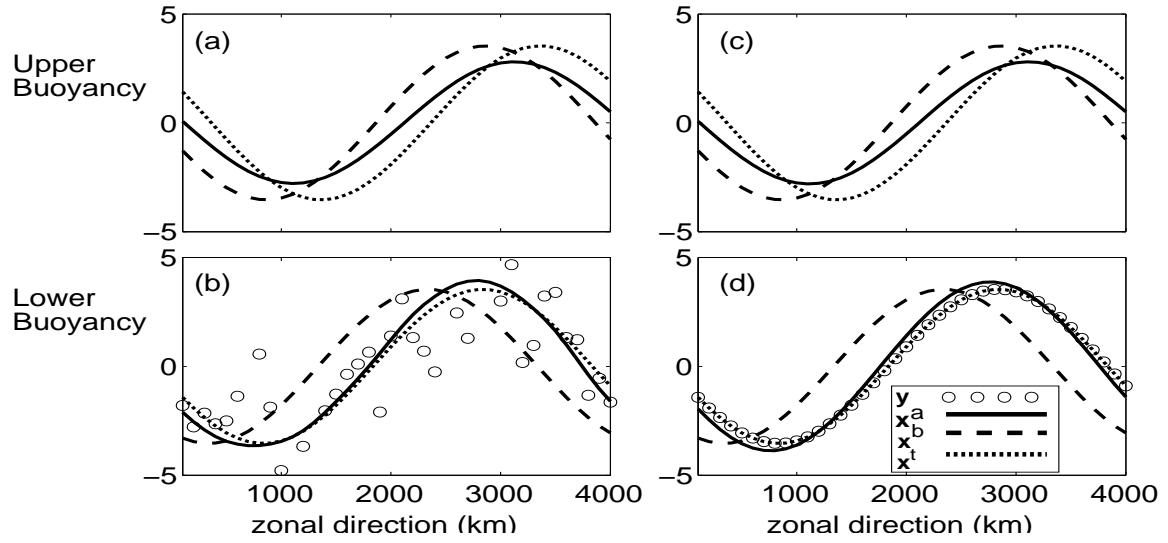
Figure B.2: 4D-Var analyses created using observations with (a-b) noise with standard deviation 1.5 and (c)-(d) no noise. Both experiments use $\mu^2 = 3.5$ and the analyses are shown at the beginning of a 12 hour assimilation window.