

**A SINGULAR VECTOR**  
**PERSPECTIVE OF 4D-VAR:**  
**FILTERING AND INTERPOLATING**

C. Johnson <sup>1</sup>, N. K. Nichols <sup>2</sup>, B. J. Hoskins<sup>1</sup>

NUMERICAL ANALYSIS REPORT 1/04

Department of Mathematics

University of Reading

Whiteknights, PO Box 220

Reading, RG6 2AX

This work was funded by the Natural Environment Research Council (NERC) and supported by the Met Office through a CASE studentship.

---

<sup>1</sup>Department of Meteorology, University of Reading, UK

<sup>2</sup>Department of Mathematics, University of Reading, UK

## Abstract

Four-dimensional variational data assimilation (4D-Var) combines the information from a time-sequence of observations with the model dynamics and a background state to produce an analysis. In this paper, a new mathematical insight into the behaviour of 4D-Var is gained from an extension of concepts that are used to assess the qualitative information content of observations in satellite retrievals. It is shown that the 4D-Var analysis increments can be written as a linear combination of the singular vectors of a matrix which is a function of both the observational and the forecast model systems.

This formulation is used to consider the filtering and interpolating properties of 4D-Var using idealized case-studies with a simple model of baroclinic instability. The results of the 4D-Var case-studies exhibit the reconstruction of the state in unobserved regions, as a consequence of the interpolation of observations through time. The results also exhibit the filtering of components with small spatial scales that correspond to noise, and the filtering of structures in unobserved regions.

The singular vector perspective gives a very clear view of this filtering and interpolating by the 4D-Var algorithm and shows that the appropriate specification of the a priori statistics is vital to extract the maximal amount of useful information from the observations.

# Contents

<b>1</b>	<b>Introduction</b>	<b>4</b>
<b>2</b>	<b>Qualitative Information Content of Observations in 4D-Var</b>	<b>5</b>
<b>3</b>	<b>Eady Model Experiments: Description</b>	<b>8</b>
3.1	2D Eady Model . . . . .	8
3.2	Identical Twin Experiments . . . . .	10
3.2.1	True State. . . . .	10
3.2.2	Background Error Correlations. . . . .	10
3.2.3	Background State. . . . .	11
3.2.4	Observations. . . . .	11
3.3	4D-Var Algorithm. . . . .	12
3.4	SVD Computations. . . . .	12
<b>4</b>	<b>Eady model experiments: Results</b>	<b>12</b>
4.1	4D-Var Analyses . . . . .	12
4.2	SVD Interpretation . . . . .	15
<b>5</b>	<b>Conclusions</b>	<b>18</b>
<b>6</b>	<b>Discussion</b>	<b>19</b>
<b>7</b>	<b>Acknowledgements</b>	<b>20</b>
<b>8</b>	<b>References</b>	<b>21</b>

## List of Figures

- 1 Horizontal auto-correlation of the background state error for correlation length-scales  $l = 200km, 600km, \text{ and } 1000km$ . . . . . 11
- 2 4D-Var analyses ( $\mathbf{x}^a$ , solid), with (a)-(b) Exp. 1:  $l_{\text{specified}} = 1000km$  and (c)-(d) Exp. 2:  $l_{\text{specified}} = 200km$ . In both cases, the true state ( $\mathbf{x}_0^t$ , dotted) is given by the most unstable normal mode, and the background state ( $\mathbf{x}^b$ , dashed) has random correlated errors that are consistent with a covariance matrix with  $\sigma_b^2 = 1.5^2$  and  $l_{\text{actual}} = 1000km$ . Observations ( $\mathbf{y}$ , circles) of the lower level buoyancy are given at the beginning and the end of a 6 hour assimilation window. The observations have random uncorrelated errors with a Gaussian distribution and variance  $\sigma_o^2 = 1.5^2$ . The actual variance ratio is  $\mu_{\text{actual}}^2 = \sigma_o^2/\sigma_b^2 = 1$  and the specified variance ratio is also  $\mu_{\text{specified}}^2 = 1$ . All fields are shown at the beginning of the assimilation 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. . . . . 13
- 3 4D-Var analyses where the actual variance is  $\mu_{\text{actual}}^2 = 1$  whilst the specified variance ratio is  $\mu_{\text{specified}}^2 = 4 \times 10^{-3}$ . The observations in (c)-(d) are generated using a different random seed to the observations in (a)-(b). Other parameter values are:  $\sigma_o^2 = 1.5^2, \sigma_b^2 = 1.5^2, l_{\text{specified}} = 1000km, l_{\text{actual}} = 1000km$ . The details are as for Fig. 2. . . . . 14
- 4 4D-Var analyses where the specified variance ratio is (a)-(b)  $\mu_{\text{specified}}^2 = 4 \times 10^{-3}$  and (c)-(d)  $\mu_{\text{specified}}^2 = 1$ , whilst the actual variance ratio is  $\mu_{\text{actual}}^2 = 4 \times 10^{-3}$ . Parameter values are:  $\sigma_o^2 = 0.1^2, \sigma_b^2 = 1.5^2, l_{\text{specified}} = 1000km, l_{\text{actual}} = 1000km$ . The details are as for Fig. 2. . . . . 15
- 5 The singular values,  $\lambda_j$ , (upper panel) and the corresponding RSVs,  $\mathbf{v}_j$ , (lower panels) of the normalized observability matrix as a function of the singular vector index,  $j$ . The RSVs are shown by the amplitude of the upper and lower buoyancy fields as a function of horizontal distance  $x$ , with each RSV separated by one unit of (non-dimensional) buoyancy. Observations are given at the beginning and the end of a 6 hour window. The background error correlation has a length-scale of  $l_{\text{specified}} = 200km$ . . . . . 16
- 6 The singular values,  $\lambda_j$ , (upper panel) and the corresponding RSVs,  $\mathbf{v}_j$ , (lower panels) of the normalized observability matrix as a function of the singular vector index,  $j$ . The background error correlation has a length-scale of  $l_{\text{specified}} = 1000km$  and the details are as for Fig. 5. . . . . 17
- 7 Values of the Picard ratio  $\log(c_j) = \log(|\mathbf{u}_j^T \boldsymbol{\rho}_R^{-1/2} \hat{\mathbf{d}}|/\lambda_j)$  for observations with no noise (solid), and noise with standard deviation  $\sigma_o = 0.01$  (dashed), 0.1 (dot-dashed) and 1.5 (dotted). The background error correlations have a length-scale of  $l_{\text{specified}} = 1000km$ . Only the values between  $-1$  and  $20$  are shown for clarity, but the Picard ratios have a maximum value of  $40$  at  $j = 40$  and the values for perfect observations reach values of  $\log$  zero to machine accuracy. . . . . 18

# 1 Introduction

In weather forecasting, data assimilation is used to generate an analysis that provides the initial conditions for numerical weather prediction. The data assimilation algorithm combines together various sources of information about the present state of the atmosphere such as observations, a model forecast valid at the same time (known as a background state), statistics of the expected observation and background state errors, and a forecast model.

The algorithm known as three-dimensional variational data assimilation (3D-Var) gives an analysis which is a weighted average of the observations and the background state. For linear observation operators, this is the Best Linear Unbiased Estimate (BLUE), (Lorenz 1986). The weights are defined by the error covariances so that when the observations are relatively accurate, the analysis is drawn to the observations, but when the observations are relatively inaccurate or there are no observations, the analysis is drawn to the background state. Since the observational coverage of the atmosphere is not complete, the specified background error correlations are important for spreading information from the observations to the regions surrounding the observation locations. Thus, the a priori statistics play an important role in determining the filtering and interpolating properties of the algorithm (Hollingsworth and Lönnberg 1986).

3D-Var uses observations given at only a single point in time. However, it is possible to make use of a full time-sequence of observations by linking the observations with the model equations. This was first considered by Thompson (1961), who used dynamical equations to propagate information from data rich regions into data holes. Sasaki (1970) subsequently developed a variational method to combine the information from a time-sequence of observations with a numerical model. Finding the optimal state at every time is expensive, but the cost may be reduced by using only the initial conditions as the control variables (Le Dimet and Talagrand 1986), giving what is now known as four-dimensional variational data assimilation (4D-Var).

4D-Var is the most advanced data assimilation algorithm used operationally (Rabier *et al.* 2000). The use of the model equations means it is much more computationally expensive than 3D-Var, but it does yield many advantages. Courtier and Talagrand (1987), Rabier and Courtier (1992), Thépaut *et al.* (1993), and Tanguay *et al.* (1995) have demonstrated that 4D-Var is able to extract information from a time-sequence of observations to infer the state in unobserved regions. Also, from the equivalence with the Kalman Filter, it is known that 4D-Var is able to generate flow-dependent structure functions and hence give westward tilting analysis increments that are vital for baroclinic growth. This has been demonstrated with single-observation experiments by Thépaut *et al.* (1993). Operational experiments (Rabier *et al.* 1998; Desroziers *et al.* 1999) have also shown that 4D-Var gives better analyses than 3D-Var, particularly in baroclinic regions.

Although the advantages of 4D-Var have been demonstrated, they are not well understood theoretically. Single-observation experiments have usefully illustrated the flow-dependent structure functions that are implicitly generated by 4D-Var, but they have not provided insight into how the information from an observation interacts with both the model dynamics and the other observations in a 4D-Var algorithm. The purpose of this paper is to provide a new insight into the use of a time-sequence of observations in 4D-Var. This will be achieved mathematically through an extension of an information content technique that is widely used in satellite retrieval studies. This technique is used later in this paper to provide a new

interpretation of 4D-Var analyses from idealized case-studies using the 2D Eady model.

In section 2, the singular value decomposition (SVD) technique for examining the qualitative information content of observations in 4D-Var is introduced. The Eady model and experiments are introduced in section 3. In section 4, the 4D-Var analyses are first discussed and then interpreted using the SVD technique. The main conclusions of the study are given in section 5 and are followed by a brief discussion. The notation used in the paper follows Ide *et al.* (1997).

## 2 Qualitative Information Content of Observations in 4D-Var

4D-Var considers observations,  $\mathbf{y}_i$ , at time,  $t_i$ , distributed over an assimilation time window,  $[t_0, t_N]$ . It is assumed that the observational errors are unbiased, temporally uncorrelated and have Gaussian distributions with covariances,  $\mathbf{R}_i$ . The observations,  $\mathbf{y}_i$ , are related to the state variables,  $\mathbf{x}_i$ , by the observation operator,  $\mathbf{H}$ . Throughout this paper, it is assumed that the observation operator is linear. The background state,  $\mathbf{x}^b$ , is specified at the initial time,  $t_0$ , and it is assumed that the background state errors are also unbiased and have a Gaussian distribution with covariance,  $\mathbf{B}$ . It is assumed that there are no cross-correlations between the observations and the background state.

The 4D-Var algorithm then finds the initial state,  $\mathbf{x}_0$ , that is close to the background state at the initial time and such that the trajectory,  $\mathbf{x}_i$ , satisfying the model equations, is close to the observations through the assimilation window. Again, it is assumed that the model,  $\mathbf{M}$ , is perfect and linear. The linearity assumption simplifies the problem considerably, and allows the use of the SVD. In operational assimilation, both the observation operator and the model are weakly nonlinear. The following interpretation, however, could be applied to the linearized models in an incremental 4D-Var formulation.

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

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

subject to the linear model constraint,

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

where  $\mathbf{x}_i$  is the state vector at time  $t_i$ , and  $\mathbf{M}(t_{i+1}, t_i)$  is the linear model operator from time  $t_i$  to  $t_{i+1}$ . The first and second terms in the cost function are known as the  $J^b$  and the  $J^o$  terms respectively.

The 4D-Var algorithm is a constrained minimization but we may rewrite it as an unconstrained minimization by substituting the model (1b) into the  $J^o$  term in (1a):

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}(\hat{\mathbf{y}} - \hat{\mathbf{H}}\mathbf{x}_0)^T \hat{\mathbf{R}}^{-1}(\hat{\mathbf{y}} - \hat{\mathbf{H}}\mathbf{x}_0), \quad (2a)$$

where the block matrix,

$$\hat{\mathbf{H}} = \begin{bmatrix} \mathbf{H}^T & (\mathbf{H}\mathbf{M}(t_1, t_0))^T & \dots & (\mathbf{H}\mathbf{M}(t_N, t_0))^T \end{bmatrix}^T, \quad (2b)$$

is known in control theory as the observability matrix (Zou *et al.* 1992). The observability matrix links the time-sequence of observations to the state vector. It can be thought of as a generalized observation operator as it acts in a similar way to the observation operator in 3D-Var. The vector

$$\hat{\mathbf{y}} = [\mathbf{y}_0^T \quad \mathbf{y}_1^T \quad \dots \quad \mathbf{y}_N^T]^T \quad (2c)$$

is the generalized observation vector. The generalized observation error covariance,  $\hat{\mathbf{R}}$ , is a block diagonal matrix with the observation error covariances,  $\mathbf{R}_i$ , on the diagonal. Note that this formulation can be extended to include temporal observation error correlations, in which case  $\hat{\mathbf{R}}$  would no longer be block diagonal.

We now assume that the observation error variance is  $\sigma_o^2$  for all observations, and the background state error variance is  $\sigma_b^2$  for all control variables. This is not true in general, but simplifies the equations considerably. The covariance matrices can then be written as:

$$\hat{\mathbf{R}} = \sigma_o^2 \boldsymbol{\rho}_R, \quad \mathbf{B} = \sigma_b^2 \boldsymbol{\rho}_B, \quad (3)$$

where  $\boldsymbol{\rho}_R$  and  $\boldsymbol{\rho}_B$  are the observation and background state correlation matrices respectively. If these are substituted into (2a), then the analysis is also the minimum of

$$2\sigma_o^2 J(\mathbf{x}) = \mu^2 (\mathbf{x}_0 - \mathbf{x}^b)^T \boldsymbol{\rho}_B^{-1} (\mathbf{x}_0 - \mathbf{x}^b) + (\hat{\mathbf{y}} - \hat{\mathbf{H}}\mathbf{x}_0)^T \boldsymbol{\rho}_R^{-1} (\hat{\mathbf{y}} - \hat{\mathbf{H}}\mathbf{x}_0), \quad (4)$$

where  $\mu^2 = \sigma_o^2/\sigma_b^2$  is the variance ratio, which can be considered as a weighting parameter or signal-to-noise ratio that determines the relative weight given to the observations in comparison to the background state. Such a parameter is also considered by Gong *et al.* (1998).

Writing the algorithm in form (4) clearly illustrates that 4D-Var solves the least-squares equations. Least-squares equations have been used for many years to solve other inverse problems such as deducing unknown constants in dynamic oceanography (Wunsch 1977) and determining the vertical distribution of ozone in remote sensing (Mateer 1965). Current satellite retrieval algorithms also use the least-squares equations, and in such problems it is useful to assess the information content of observations. For example to select a subset of radiance channels it is useful to assess the information that is provided by each channel.

Many techniques have been developed to assess the information content of observations. One of these which will be used here is the singular value decomposition (SVD). This technique has been used in the context of satellite retrievals (Mateer 1965; Thépaut and Moll (1990); Prunet *et al.* 1998; and Rabier *et al.* 2002).

Guided by their importance in the filtering and interpolation in 3D-Var, we first consider the error covariance matrices by defining the pre-conditioned control variable,

$$\boldsymbol{\chi} = \boldsymbol{\rho}_B^{-1/2} (\mathbf{x}_0 - \mathbf{x}^b), \quad (5)$$

where  $\boldsymbol{\rho}_B^{-1/2}$ , which satisfies  $(\boldsymbol{\rho}_B^{-1/2})^T (\boldsymbol{\rho}_B^{-1/2}) = \boldsymbol{\rho}_B^{-1}$ , denotes the symmetric square root of  $\boldsymbol{\rho}_B^{-1}$ . A similar transformation is often used both to define the  $\mathbf{B}$  matrix and to precondition the problem (see for example, Courtier *et al.* 1998). The transformation is useful here as it gives variables that are uncorrelated.

The 4D-Var uncorrelated analysis increments then minimize the cost function,

$$2\sigma_o^2 J(\boldsymbol{\chi}) = \mu^2 \|\boldsymbol{\chi}\|_2^2 + \|\boldsymbol{\rho}_R^{-1/2} \hat{\mathbf{H}} \boldsymbol{\rho}_B^{1/2} \boldsymbol{\chi} - \boldsymbol{\rho}_R^{-1/2} \hat{\mathbf{d}}\|_2^2, \quad (6)$$

using the notation  $\|\mathbf{z}\|_2^2 = \mathbf{z}^T \mathbf{z}$ , and where  $\hat{\mathbf{d}} = \hat{\mathbf{y}} - \hat{\mathbf{H}}\mathbf{x}^b$  is the generalized innovation vector.  $\boldsymbol{\rho}_R^{-1/2}$  is the symmetric square root of  $\boldsymbol{\rho}_R^{-1}$ . The 4D-Var algorithm has now been written in a form similar to that in Tikhonov Regularization, which is used to solve discrete ill-posed inverse problems (for example, Winkler 1997; Hansen 2001).

The matrix  $\boldsymbol{\rho}_R^{-1/2} \hat{\mathbf{H}} \boldsymbol{\rho}_B^{1/2}$  will be called the normalized observability matrix as the correlation matrices can be considered as normalizing factors and also to be consistent with the terminology used by Rabier *et al.* (2002). We let the generalized observation vector,  $\hat{\mathbf{y}}$ , be of dimension  $m$  and the state vector,  $\mathbf{x}$ , be of dimension  $n$  so that the normalized observability matrix is an  $m \times n$  matrix with its rank denoted by  $r \leq \min\{m, n\}$ . We then consider the SVD of this matrix:

$$\boldsymbol{\rho}_R^{-1/2} \hat{\mathbf{H}} \boldsymbol{\rho}_B^{1/2} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{V}^T, \quad (7)$$

where  $\boldsymbol{\Lambda}$  is a diagonal matrix with  $r$  positive singular values,  $\lambda_j$ , arranged in decreasing order on the diagonal, where  $j$  is the singular vector index. The  $m$  columns  $\mathbf{u}_j$  of  $\mathbf{U}$  are known as the left singular vectors (LSVs) and the  $n$  columns  $\mathbf{v}_j$  of  $\mathbf{V}$  are known as the right singular vectors (RSVs). The LSVs form an orthonormal basis for the observation space and the RSVs form an orthonormal basis for the state space. Thus, the SVD can be used to identify the structures in state space that can be determined from the observations and also to identify the observations that are important in this determination.

If the SVD (7) is substituted into the cost function (6) then it is easily shown (Wunsch 1996; Rodgers 2000, p.109; Hansen 2001) that the analysis increments can be written as:

$$\boldsymbol{\chi} \equiv \boldsymbol{\rho}_B^{-1/2} (\mathbf{x}^a - \mathbf{x}^b) = \sum_{j=1}^r f_j c_j \mathbf{v}_j, \quad (8a)$$

where

$$f_j = \frac{\lambda_j^2}{\mu^2 + \lambda_j^2}, \quad (8b)$$

$$c_j = \frac{\mathbf{u}_j^T \boldsymbol{\rho}_R^{-1/2} \hat{\mathbf{d}}}{\lambda_j}. \quad (8c)$$

The uncorrelated analysis increments are now written as a linear combination of the RSVs of the normalized observability matrix, where the weight given to each RSV is determined by the product of two terms:  $f_j$  and  $c_j$ . This formulation illustrates the interpolating and filtering aspects of the 4D-Var algorithm.

The interpolating aspect is more clearly illustrated by considering the case with  $\mu^2 = 0$ . This is equivalent to considering the 4D-Var algorithm with no  $J^b$  term. It is assumed that such a problem is well-posed, although this is not true in general. When this parameter is zero,  $f_j = 1$  for all  $j$  and so the weight given to the RSVs is only determined by the coefficient  $c_j$ . This coefficient has a relatively large value for those LSVs almost in the direction of the normalized generalized innovation vector. Thus, the analysis increment is composed of the combination of RSVs needed for the analysis to fit exactly through the observations. In this way, the RSVs can be considered as basis functions for the interpolation, where the interpolating coefficients are determined by the innovation vector.

The filtering aspect is then illustrated by considering the case with  $\mu^2 > 0$ . This only changes the value of what are known as the Tikhonov filter factors,  $f_j$  (Hansen 2001). The filter

factors damp all the contributions to the analysis increment which have small singular values,  $\lambda_j$ , as:

$$f_j \cong \begin{cases} 1 & \lambda_j \gg \mu \\ 1/2 & \lambda_j = \mu \\ \lambda_j^2/\mu^2 & \lambda_j \ll \mu. \end{cases} \quad (9)$$

For example, consider a fixed value of  $\lambda_j$ . Then, if the observations are relatively accurate so that the variance ratio between the observation and background state errors is smaller than the squared singular value ( $\mu^2$  is smaller than  $\lambda_j^2$ ), the filter factor has a value close to one so that the corresponding RSV,  $\mathbf{v}_j$ , is incorporated into the analysis increment and the analysis is close to the observations. On the other hand, if the observations are relatively inaccurate, so that the variance ratio between the observation and background state errors is larger than the squared singular value ( $\mu^2$  is larger than  $\lambda_j^2$ ), the filter factor has a value close to zero so that the corresponding RSV,  $\mathbf{v}_j$ , is strongly damped and hence the corresponding observational information is rejected and the analysis is close to the background state. This filtering is a vital part of the 4D-Var algorithm as both the background state and the observations always have errors. The choice of the specified value of  $\mu^2$  is crucial to enable the algorithm to extract the signal whilst filtering the observational noise.

The filter factors also imply that if we wish 4D-Var to include observational information that has a large projection onto a particular RSV,  $\mathbf{v}_j$ , the ratio of the standard deviation of the observational and background state errors,  $\mu$ , must be comparable to or smaller than the corresponding singular value,  $\lambda_j$ .

It is proposed here that this perspective of 4D-Var, summarized mathematically in (8), is a useful technique for identifying which components of the state vector are related to the observations for particular forecast models, observing systems, and error covariances. The RSVs do not depend on the relative weight,  $\mu^2$ , given to the background state or the innovation vector,  $\hat{\mathbf{d}}$ . Therefore, this singular vector perspective may be used to give general conclusions concerning 4D-Var.

### 3 Eady Model Experiments: Description

The following experiments explore the behaviour of 4D-Var using idealized case-studies with the 2D Eady model (Eady 1949). This model is able to capture quite realistic baroclinic wave growth and decay but its dynamics are sufficiently simple that it should allow a clear understanding of the mechanisms of 4D-Var. The first part of the study considers the results given by 4D-Var analyses and the second part of the study considers the SVD of the corresponding normalized observability matrix. The model, identical twin experiments, 4D-Var algorithm, and SVD computations are now described.

#### 3.1 2D Eady Model

The 2D Eady model is a simple linear, quasi-geostrophic (QG) model of baroclinic instability. The model equations describe the evolution of perturbations to a basic state. It should be noted that although the equations are linear, they have not been linearized as there are no terms that are quadratic in perturbation quantities. However, only the perturbations to the

Experiment	$\mu_{\text{actual}}^2$	$\mu_{\text{specified}}^2$	$l_{\text{actual}}(km)$	$l_{\text{specified}}(km)$	Note
1	1	1	1000	1000	
2	1	1	1000	200	
3	1	$4 \times 10^{-3}$	1000	1000	
4	1	$4 \times 10^{-3}$	1000	1000	different random seed for $\epsilon^o$
5	$4 \times 10^{-3}$	$4 \times 10^{-3}$	1000	1000	
6	$4 \times 10^{-3}$	1	1000	1000	

**Table 1:** Summary of the parameters used in the six 4D-Var experiments.  $\mu_{\text{actual}}^2 = \sigma_o^2/\sigma_b^2$  is the actual variance ratio,  $\mu_{\text{specified}}^2$  is the variance ratio that is used by the 4D-Var algorithm,  $l_{\text{actual}}$  is the length-scale used to generate the background state, and  $l_{\text{specified}}$  is the length-scale used by the 4D-Var algorithm. The parameter values are shown to an accuracy of one significant figure.

basic state are taken as control variables in the data assimilation, and the basic state flow is assumed to be correct. The non-dimensional equations are now given.

The basic state is assumed to be dependent on the meridional direction,  $y$ , through a linear temperature gradient. The perturbations are independent of  $y$ . The basic state is given by a linear zonal wind shear with height,  $z$ , that is associated with the uniform meridional temperature gradient in a domain between two rigid horizontal boundaries,  $z = \pm 1/2$ . The density, static stability and Coriolis parameter are all taken to be constants. It is also assumed that the interior quasi-geostrophic potential vorticity is zero.

The initial state is given by the perturbation buoyancy,  $b = b(x, z, t)$ , on the boundaries,  $z = \pm 1/2$ , at time  $t = 0$ . This is used to calculate the corresponding perturbation geostrophic streamfunction,  $\psi = \psi(x, z, t)$ , which satisfies:

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial z^2} = 0, \quad \text{in } z \in \left[-\frac{1}{2}, \frac{1}{2}\right] \quad x \in [0, X]. \quad (10a)$$

From hydrostatic balance, the boundary conditions are:

$$\frac{\partial \psi}{\partial z} = b, \quad \text{on } z = \pm \frac{1}{2}, \quad x \in [0, X]. \quad (10b)$$

Perturbations to the basic state are advected zonally by the basic shear flow as described by the non-dimensional QG thermodynamic equation:

$$\left(\frac{\partial}{\partial t} + z \frac{\partial}{\partial x}\right) b = \frac{\partial \psi}{\partial x}, \quad \text{on } z = \pm \frac{1}{2}, \quad x \in [0, X]. \quad (10c)$$

The spatial boundary conditions are taken to be periodic such that at any time,  $t$ , and height,  $z$ ,  $b(0, z, t) = b(X, z, t)$  and  $\psi(0, z, t) = \psi(X, z, t)$ .

The Eady model is discretized using 11 vertical levels for streamfunction. There are 20 grid points in the horizontal, giving 40 degrees of freedom. The advection equations are discretized using a leap-frog advection scheme, and the NAG routine known as nag-gen-lin-sys is used to perform an LU factorization to solve the elliptic equation. The discrete model with non-zero QGPV has previously been used by Badger and Hoskins (2001) to investigate the nature of optimal perturbations.

## 3.2 Identical Twin Experiments

The Eady model is used within 4D-Var identical twin experiments. The numerical model is first integrated from the initial conditions over a six hour assimilation window to give the true state. The background state is then defined from the true state with random errors that are consistent with a specified background error covariance with horizontal length-scale,  $l_{\text{actual}}$ , and variance,  $\sigma_b^2$ . The observations are also defined from the true state. Uncorrelated random errors are added with variance  $\sigma_o^2$  so that the actual error variance ratio is  $\mu_{\text{actual}}^2 = \sigma_o^2/\sigma_b^2$ . The variance ratio,  $\mu_{\text{specified}}^2$ , and the background error length-scale,  $l_{\text{specified}}$ , that are used in the 4D-Var algorithm are not necessarily the same as the actual parameter values. The following six experiments are designed to investigate the impact of different parameter settings and are summarized in Table 1.

### 3.2.1 True State.

The true state initial conditions,  $\mathbf{x}_0^t$ , are given by the analytical solution for the most unstable normal mode structure;

$$\psi(x, z) = \cosh(kz) \cos(kx) - \alpha \sinh(kz) \sin(kx), \quad (11)$$

where the non-dimensional wave number  $k = 1.6$  and  $\alpha = 2.59$ . This solution exhibits a westward tilt with height of the streamfunction field and an eastward tilt with height of the buoyancy field. This tilt with height is associated with the vertical coupling between the upper and lower waves that leads to exponential growth with the non-dimensional growth rate  $\omega = 0.31$ . The true state at the end of the assimilation window,  $t_N$ , has the same structure as in (11), but the amplitude has increased by a factor  $e^{\omega t_N}$ .

### 3.2.2 Background Error Correlations.

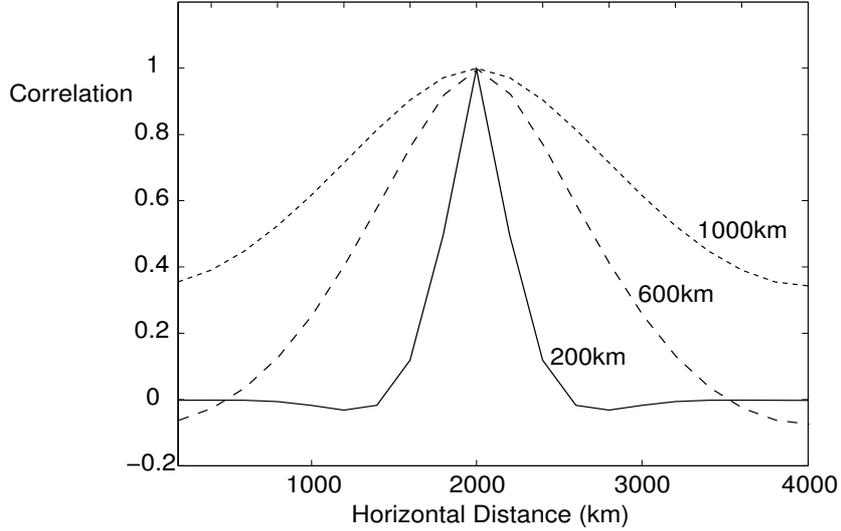
The background error correlation matrix is specified by defining the inverse matrix using second derivatives. It is assumed that there are no correlations between the upper and lower buoyancy errors, so the  $\boldsymbol{\rho}_B$  matrix is block diagonal and is defined by

$$\boldsymbol{\rho}_B^{-1} = \begin{pmatrix} \boldsymbol{\rho}_{TT}^{-1} & 0 \\ 0 & \boldsymbol{\rho}_{TT}^{-1} \end{pmatrix}, \quad (12a)$$

where

$$\boldsymbol{\rho}_{TT}^{-1} = \gamma \left( \mathbf{I} + \frac{l^4}{2} (\mathbf{L}_{xx})^2 \right). \quad (12b)$$

$\mathbf{L}_{xx}$  is a finite difference second derivative matrix in the  $x$  direction, incorporating periodic boundary conditions,  $l$  is the correlation length-scale, and  $\gamma$  is a scalar parameter that is specified so that the trace of  $\boldsymbol{\rho}_B$  is equal to 40 (the number of unknowns). This is similar to the second derivative smoothness constraints used by Sasaki (1970) and Schröter *et. al* (1993). Correlations with various length-scales are shown in Fig. 1.



**Figure 1:** Horizontal auto-correlation of the background state error for correlation length-scales  $l = 200km$ ,  $600km$ , and  $1000km$ .

### 3.2.3 Background State.

The background state is defined by the true state with correlated random errors;

$$\mathbf{x}^b = \mathbf{x}_0^t + \boldsymbol{\rho}_B^{1/2} \boldsymbol{\varepsilon}^b. \quad (13)$$

$\boldsymbol{\varepsilon}^b$  is a vector of unbiased random numbers with a Gaussian distribution of variance  $\sigma_b^2 = 1.5^2$  and  $\boldsymbol{\rho}_B$  is a correlation matrix, of the form (12), with length-scale  $l_{\text{actual}} = 1000km$ .

### 3.2.4 Observations.

As the position and amplitude of the upper level wave is vital for the growth of the lower level wave, we consider the case where only the lower level wave is observed and 4D-Var seeks to infer the unobserved upper level wave. It should be noted that due to the symmetry of the Eady model, the mathematical problem is equivalent to observing the upper level wave and reconstructing the lower level wave.

The synthetic observations,

$$\mathbf{y}_i = \mathbf{H}\mathbf{x}_i^t + \boldsymbol{\varepsilon}_i^o \quad \text{for } i = 0 \text{ and } N, \quad (14)$$

are taken from the lower level buoyancy field of the true state. The errors,  $\boldsymbol{\varepsilon}_i^o$ , are unbiased and random with a Gaussian distribution of variance  $\sigma_o^2$ , which takes the value of  $1.5^2$  in Exps. 1 to 4 and  $0.1^2$  in Exps. 5 and 6. The observational errors are assumed to be uncorrelated, so that  $\boldsymbol{\rho}_R = \mathbf{I}$ .

### 3.3 4D-Var Algorithm.

The first part of the study considers the 4D-Var analyses. From (4), the 4D-Var cost function may be written as:

$$J(\mathbf{x}_0) = \mu_{\text{specified}}^2 (\mathbf{x}_0 - \mathbf{x}^b) \boldsymbol{\rho}_B^{-1} (\mathbf{x}_0 - \mathbf{x}^b) + (\mathbf{y}_0 - \mathbf{H}\mathbf{x}_0)^T (\mathbf{y}_0 - \mathbf{H}\mathbf{x}_0) + (\mathbf{y}_N - \mathbf{HM}(t_N, t_0)\mathbf{x}_0)^T (\mathbf{y}_N - \mathbf{HM}(t_N, t_0)\mathbf{x}_0), \quad (15)$$

where  $\boldsymbol{\rho}_B^{-1}$  is a function of  $l_{\text{specified}}$ .

The 4D-Var algorithm uses the forward Eady model to calculate the values of  $J$  and the adjoint Eady model to calculate the values of  $\nabla J$ . These values are used by the quasi-Newton minimization algorithm known as CONMIN or algorithm 500 from TOMS (Shanno and Phua 1976) to find the state that minimizes the cost function.

### 3.4 SVD Computations.

The second part of the study interprets the 4D-Var analysis results using the singular values and singular vectors of the corresponding normalized observability matrix (7). With observations at only the beginning and the end of the window and no observation error correlations, this is given by

$$\boldsymbol{\rho}_R^{-1/2} \hat{\mathbf{H}} \boldsymbol{\rho}_B^{1/2} = \begin{bmatrix} \mathbf{H} \\ \mathbf{HM}(t_N, t_0) \end{bmatrix} \boldsymbol{\rho}_B^{1/2}. \quad (16)$$

Note that we again assume that the observational errors are uncorrelated,  $\boldsymbol{\rho}_R = \mathbf{I}$ . There are various methods to calculate the singular vectors (Toumazou and Cretaux 2001). When only the first few singular vectors of an operator are required, the Lanczos strategy may be applied to the linear operator. Here, as all the singular vectors are required, the SVD strategy must be applied to the matrix form of the linear operator.

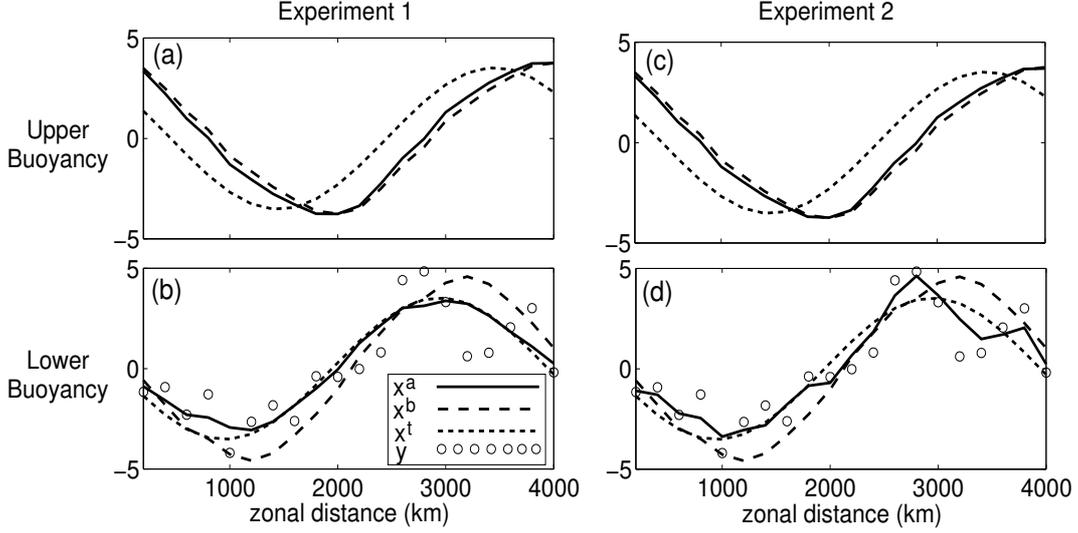
The real symmetric positive definite square root of the background error covariance is found by inverting  $\boldsymbol{\rho}_B^{-1}$  and then using the eigenvalues and eigenvectors. The Eady model is then applied to successive columns of  $\boldsymbol{\rho}_B^{1/2}$  to give the matrix  $\mathbf{M}(t_N, t_0)\boldsymbol{\rho}_B^{1/2}$ . The NAG routine known as nag-gen-svd, based on the SVD algorithm described by Golub and Van Loan (1996), is then used to find the RSVs, LSVs and singular values of the normalized observability matrix.

## 4 Eady model experiments: Results

### 4.1 4D-Var Analyses

The 4D-Var analyses from the six experiments are now discussed. All six experiments consider the same true state and background state. The background state has correlated random errors, with  $l_{\text{actual}} = 1000km$  and  $\sigma_b^2 = 1.5^2$ , which gives mainly a displacement error, as shown in for example Fig. 2(a)-(b). The first four experiments consider observations with errors with the same variance as the background state errors so that  $\mu_{\text{actual}}^2 = 1$ . These errors may appear to be large, but they provide a clear illustration of the filtering by 4D-Var.

Experiment 1 considers an assimilation using the appropriate parameters so that  $\mu_{\text{specified}}^2 = 1$  and  $l_{\text{specified}} = 1000km$ . The analysis, shown in Fig. 2(a)-(b), is close to the true state on

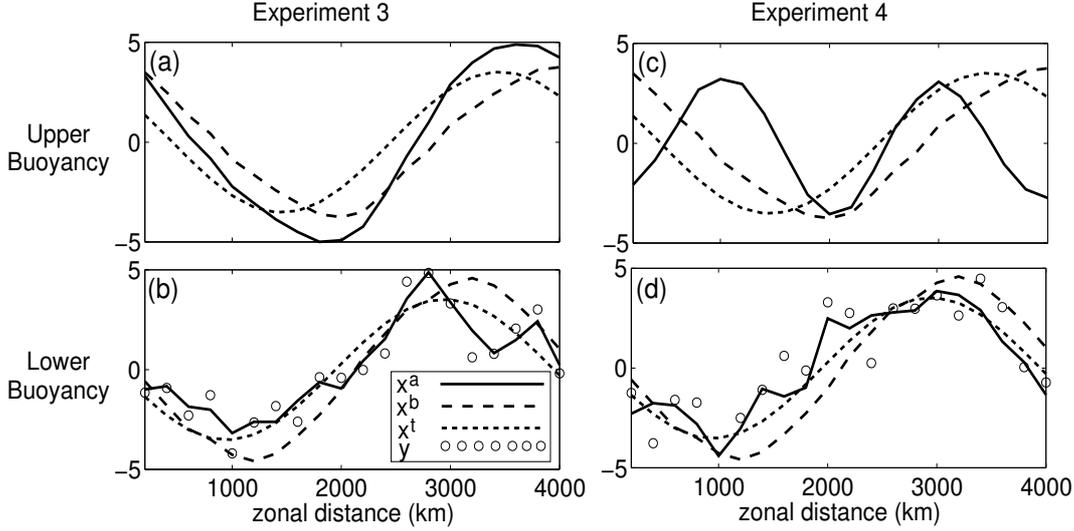


**Figure 2:** 4D-Var analyses ( $\mathbf{x}^a$ , solid), with (a)-(b) Exp. 1:  $l_{\text{specified}} = 1000\text{km}$  and (c)-(d) Exp. 2:  $l_{\text{specified}} = 200\text{km}$ . In both cases, the true state ( $\mathbf{x}_0^t$ , dotted) is given by the most unstable normal mode, and the background state ( $\mathbf{x}^b$ , dashed) has random correlated errors that are consistent with a covariance matrix with  $\sigma_b^2 = 1.5^2$  and  $l_{\text{actual}} = 1000\text{km}$ . Observations ( $\mathbf{y}$ , circles) of the lower level buoyancy are given at the beginning and the end of a 6 hour assimilation window. The observations have random uncorrelated errors with a Gaussian distribution and variance  $\sigma_o^2 = 1.5^2$ . The actual variance ratio is  $\mu_{\text{actual}}^2 = \sigma_o^2/\sigma_b^2 = 1$  and the specified variance ratio is also  $\mu_{\text{specified}}^2 = 1$ . All fields are shown at the beginning of the assimilation 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.

the lower boundary and close to the background state on the upper boundary. Thus, the algorithm uses the observations of the lower level to move the state from the background state closer to the true state. Although the observations are noisy, the analysis is smooth due to the filtering by the algorithm. In this case, the algorithm is unable to interpolate the observations through time to infer the state on the unobserved boundary.

Experiment 2 considers an assimilation using the appropriate value for  $\mu^2$ , but an inappropriate value for  $l$ . The background state errors are generated using  $l_{\text{actual}} = 1000\text{km}$ , but the 4D-Var algorithm uses a much shorter length-scale of  $l_{\text{specified}} = 200\text{km}$ . The analysis, shown in Fig. 2(c)-(d), is similar to that for Exp. 1, except that it is now noisy on the lower boundary. Thus, the specification of the correlation length-scale is vital in controlling the amount of filtering that is achieved by the algorithm. All the following experiments use the appropriate specified length-scale.

Experiment 3 considers an assimilation using the appropriate specification of the length-scale,  $l_{\text{specified}} = 1000\text{km}$ , but an inappropriate value for  $\mu^2$ . A much smaller specified value is used;  $\mu_{\text{specified}}^2 = 4 \times 10^{-3}$ . This may be interpreted as either specifying too large background state errors or specifying too small observational errors. The resulting analysis, shown in Fig. 3(a)-(b), is close to the noisy observations. Although the specified correlation length-scale is large, the  $J^b$  term is given a small weight, and therefore there is a smaller constraint on the smoothness of the analysis. Both the upper and the lower boundaries exhibit unrealistic solutions. The actual structure of these is associated with the actual noise on the observations, which is now demonstrated by using a different set of observations with the same variance. Experiment 4 is identical to Exp. 3, but the observational noise is generated using a different



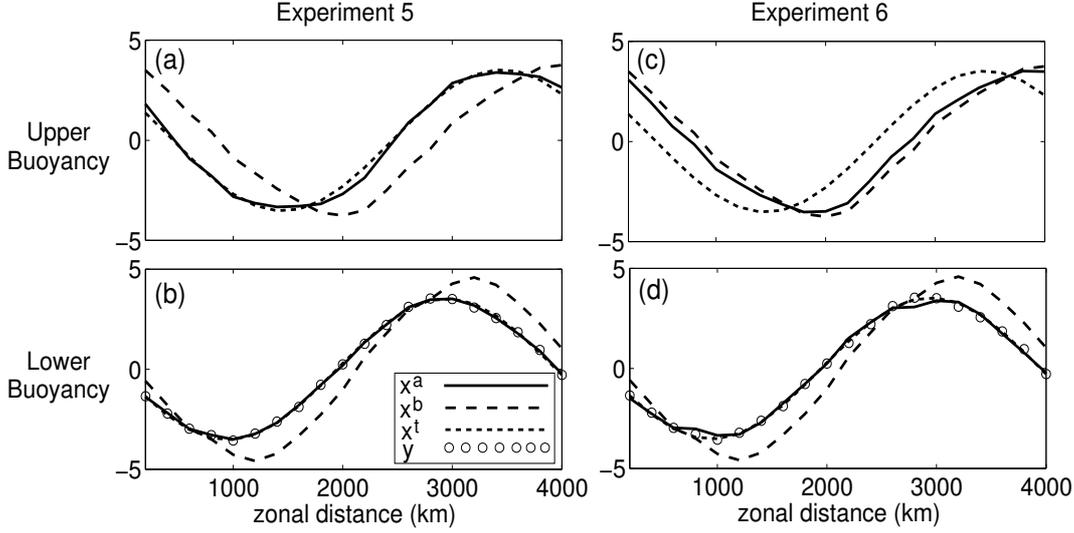
**Figure 3:** 4D-Var analyses where the actual variance is  $\mu_{actual}^2 = 1$  whilst the specified variance ratio is  $\mu_{specified}^2 = 4 \times 10^{-3}$ . The observations in (c)-(d) are generated using a different random seed to the observations in (a)-(b). Other parameter values are:  $\sigma_o^2 = 1.5^2$ ,  $\sigma_b^2 = 1.5^2$ ,  $l_{specified} = 1000km$ ,  $l_{actual} = 1000km$ . The details are as for Fig. 2.

random seed. The analysis, shown in Fig. 3(c)-(d), again exhibits unrealistic solutions on the upper and lower boundaries. They have very different structures to those in Experiment 3 and are therefore associated with the observational noise. The largest analysis errors are found on the unobserved upper boundary; this is consistent with the results found by Courtier and Talagrand (1987) and Laroche and Gauthier (1998).

In all the previous experiments, the 4D-Var algorithm is unable to infer the state on the unobserved upper boundary. This is because the observational noise is very large. The last two experiments consider observations with errors with variance  $\sigma_o^2 = 0.1^2$ , which is much smaller than the variance of the background state errors so that  $\mu_{actual}^2 = 4 \times 10^{-3}$ .

Experiment 5 considers an assimilation using the appropriate parameters;  $\mu_{specified}^2 = 4 \times 10^{-3}$  and  $l_{specified} = 1000km$ . The analysis, shown in Fig. 4(a)-(b), is close to the true state on both the lower and the upper boundary. The 4D-Var algorithm combines the valuable time-evolution information provided by the observations on the lower boundary with the model dynamics to correct the unobserved part of the state on the upper boundary. The analysed upper level wave is now in the correct position needed for the growth of the wave during the assimilation window. This clearly illustrates that 4D-Var interpolates through the observations in both space and time, enabling the state to be inferred in the unobserved regions. This reconstructive ability is a major benefit of 4D-Var and is one reason why 4D-Var has previously been shown to give better analyses than 3D-Var. It is therefore important that this benefit is maximized.

Experiment 6 considers an assimilation using the appropriate length-scale, but with a variance ratio that is too large. The specified value is  $\mu_{specified}^2 = 1$ , which may be interpreted as assuming that the background state errors are too small or that the observational errors are too large. The analysis, shown in Fig. 4(c)-(d), is similar to that for Exp. 5, except that the upper level wave is no longer near to the true state. The analysis is very similar to that in Exp. 1, as the same specified parameter values are used. Thus, although the information to the infer the state in the observed region is contained in the observations, it is not extracted by the



**Figure 4:** 4D-Var analyses where the specified variance ratio is (a)-(b)  $\mu_{\text{specified}}^2 = 4 \times 10^{-3}$  and (c)-(d)  $\mu_{\text{specified}}^2 = 1$ , whilst the actual variance ratio is  $\mu_{\text{actual}}^2 = 4 \times 10^{-3}$ . Parameter values are:  $\sigma_o^2 = 0.1^2$ ,  $\sigma_b^2 = 1.5^2$ ,  $l_{\text{specified}} = 1000\text{km}$ ,  $l_{\text{actual}} = 1000\text{km}$ . The details are as for Fig. 2.

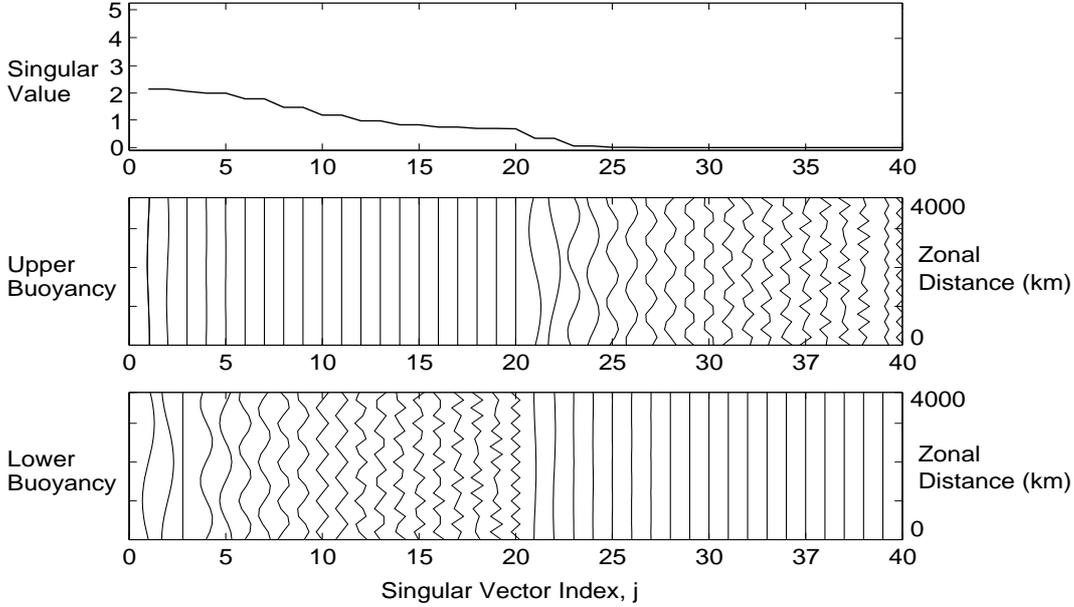
algorithm due to the inappropriate specification of  $\mu_{\text{specified}}^2$ . Thus, the reconstructive ability of 4D-Var is not maximised due to the inappropriate specification of the error variances.

The conclusions from these six experiments are now summarized. Experiment 1 shows that the correct specification of the variance ratio and length-scale gives a smooth analysis. The size of the observational noise is the same as the background state errors so that the analysis is close to the true state in the observed region, but close to the background state in the unobserved region. Experiment 2 shows that if the specified background error correlation length-scale is too small, the observational noise is not filtered enough so that the analysis is noisy. Experiments 3 and 4 show that if the specified variance ratio is too small, the algorithm draws too closely to the observations, giving a bad analysis, particularly in the unobserved regions. Experiment 5 shows that if the observational noise is smaller and the appropriate parameters are specified, the algorithm is able to draw closer to the observations to give an analysis that is close to the true state in both the observed and the unobserved regions. Experiment 6 shows that if the observational noise is still relatively small, but if the specified variance ratio is too large, the algorithm does not draw closely to the observations. Thus, there is little change in the unobserved regions and there is no benefit from the improved quality of the observations.

## 4.2 SVD Interpretation

A new interpretation of the 4D-Var analyses in the previous six experiments is now given through the use of the SVD technique, described in Section 2. First the singular values and the structure of the matrix of RSVs are examined to consider which information is filtered by 4D-Var. Then the coefficients,  $c_j$ , are examined to consider the effect of the size of the observational noise.

The SVD of the normalized observability matrix with  $l_{\text{specified}} = 200\text{km}$  is first considered. The singular values and RSVs are shown in Fig. 5. There are 20 observations at two time



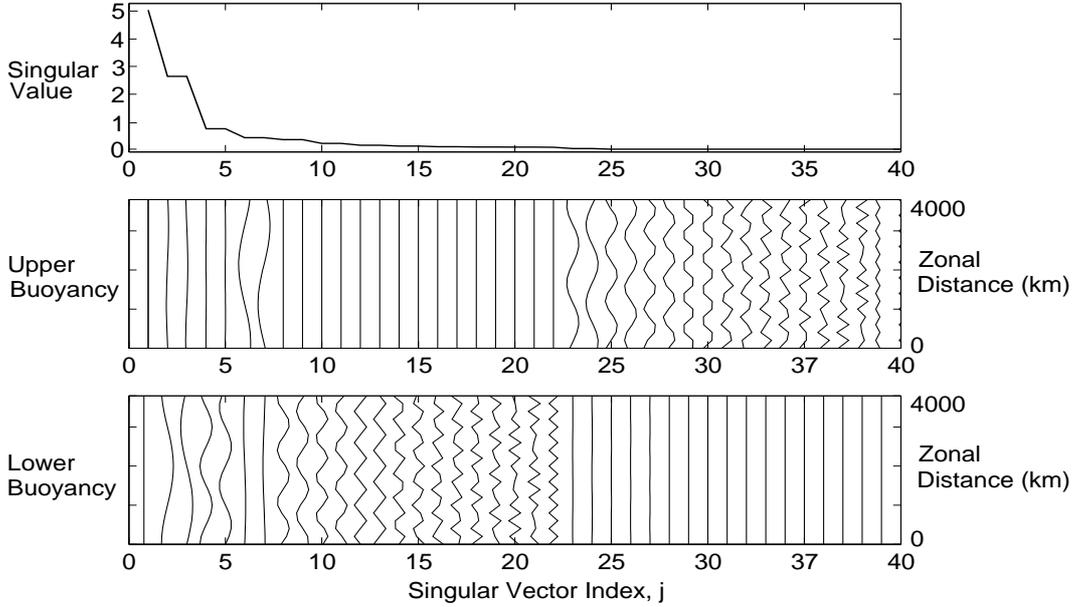
**Figure 5:** The singular values,  $\lambda_j$ , (upper panel) and the corresponding RSVs,  $\mathbf{v}_j$ , (lower panels) of the normalized observability matrix as a function of the singular vector index,  $j$ . The RSVs are shown by the amplitude of the upper and lower buoyancy fields as a function of horizontal distance  $x$ , with each RSV separated by one unit of (non-dimensional) buoyancy. Observations are given at the beginning and the end of a 6 hour window. The background error correlation has a length-scale of  $l_{\text{specified}} = 200\text{km}$ .

levels, giving a total of 40 observations. Although there are 40 non-zero singular values, the singular value spectrum decays so that many of the singular values are very small. The RSVs that correspond to large singular values have large amplitudes on the lower boundary, whereas the RSVs that correspond to small singular values have large amplitudes on the upper boundary. On both the lower and the upper boundaries, there tends to be an increase in the number of oscillations in the horizontal with increasing singular vector index.

Almost all the RSVs form pairs with the same singular value. For example, the first and second RSVs both correspond to a singular value of 2.15. The second RSV has the same structure as the first RSV, but is phase shifted in the horizontal by 1000km. This is due to the zonal symmetry of the Eady model. The only RSVs that are not in a pair either do not vary in the horizontal, or have a wavelength of 2 grid spaces.

The analysis from Exp. 2 (Fig. 2(c)-(d)) is close to the observations on the lower boundary but close to the background state on the upper boundary. This is now explained using the SVD. Equation 8 shows that the 4D-Var algorithm damps the RSVs with small singular values. From Fig. 5, these are the RSVs that correspond to the information that is needed to infer the state on the unobserved upper boundary and therefore the algorithm has filtered the structures in the unobserved region. The singular values of the first few RSVs are large enough that the corresponding observational information is largely drawn to in the analysis. These RSVs have long wavelengths and large amplitudes on the lower boundary.

The analyses from Exps. 1 and 2 (Fig. 2) show that specifying a longer length-scale gives a smoother analysis. Therefore, we now investigate the SVD of the normalized observability matrix with  $l_{\text{specified}} = 1000\text{km}$ . The RSVs, shown in Fig. 6, are now re-ordered so that, for example, RSVs 21&22 have moved to become RSVs 6&7 and the RSVs with small wavelengths



**Figure 6:** The singular values,  $\lambda_j$ , (upper panel) and the corresponding RSVs,  $\mathbf{v}_j$ , (lower panels) of the normalized observability matrix as a function of the singular vector index,  $j$ . The background error correlation has a length-scale of  $l_{\text{specified}} = 1000\text{km}$  and the details are as for Fig. 5.

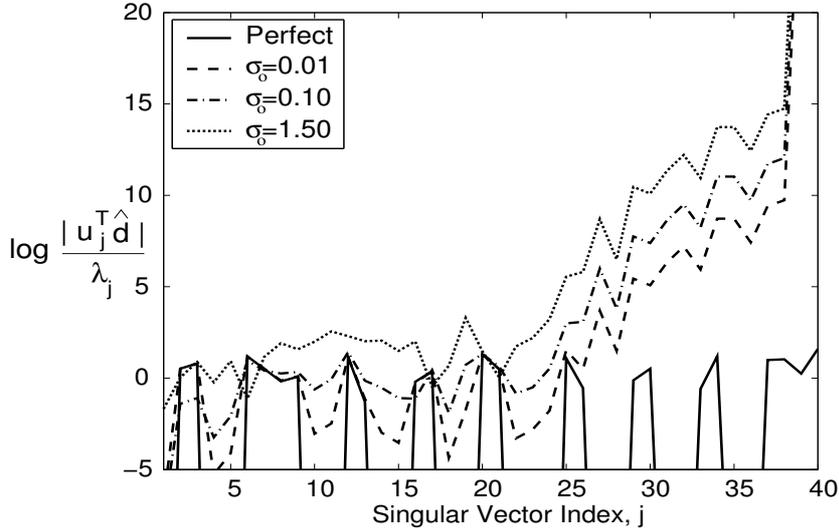
are now associated with relatively small singular values whilst the RSVs with large wavelengths are now associated with slightly larger singular values. This is because the background error covariance now has a larger variance at the large spatial scales and a smaller variance at the small spatial scales (see for example Rodgers 2000, p. 39). The RSVs with small spatial scales are therefore damped by the 4D-Var algorithm so that the analysis is more smooth. The RSVs needed to infer the large-scale component of the upper level wave still have small singular values and are therefore still damped by the algorithm when  $\mu_{\text{specified}}^2 = 1$ .

It is the coefficients,  $c_j$ , that determine which RSVs contribute to the analysis increment. Since these values span several orders of magnitude, it is more convenient to examine what is known as the Picard ratio (Winkler 1997), which is given by

$$\log(c_j) = \log \left( \frac{|\mathbf{u}_j^T \boldsymbol{\rho}_R^{-1/2} \hat{\mathbf{d}}|}{\lambda_j} \right). \quad (17)$$

As there are no observation error correlations, we use  $\boldsymbol{\rho}_R^{-1/2} = \mathbf{I}$ . The Picard ratio values for perfect observations and observations with noise of standard deviation  $\sigma_o = 0.01, 0.1$  and  $1.5$  are shown in Fig. 7. With perfect observations, approximately half the RSVs contribute to the analysis increment and the coefficients of these are all of the same order of magnitude.

The analyses from Exps. 3 and 4 show that if  $\mu_{\text{specified}}^2$  is too small, unrealistic structures corresponding to the observational noise are generated and the analysis errors are largest in the unobserved region. This is now explained using the Picard ratios. When the observations are noisy ( $\sigma_o = 1.5$ ), many more RSVs with small singular values are given a large weight. These RSVs have small-scale structures and large amplitudes on the upper boundary. Note that this is similar to the results found by Callies and Eppel (1995) where the projections of the initial state onto the eigenvectors of the Hessian of the cost function are examined. In the absence of filtering ( $\mu_{\text{specified}}^2 \ll 1$ ) these RSVs dominate the analysis increment. This explains why the largest analysis errors are found on the upper boundary. When  $\mu_{\text{specified}}^2$  has



**Figure 7:** Values of the Picard ratio  $\log(c_j) = \log(|\mathbf{u}_j^T \boldsymbol{\rho}_R^{-1/2} \hat{\mathbf{d}}|/\lambda_j)$  for observations with no noise (solid), and noise with standard deviation  $\sigma_o = 0.01$  (dashed), 0.1 (dot-dashed) and 1.5 (dotted). The background error correlations have a length-scale of  $l_{\text{specified}} = 1000\text{km}$ . Only the values between  $-1$  and 20 are shown for clarity, but the Picard ratios have a maximum value of 40 at  $j = 40$  and the values for perfect observations reach values of log zero to machine accuracy.

the appropriate value, the filter factors,  $f_j$ , damp these RSVs to give a smooth and realistic analysis.

The analysis from Exp. 5 shows that if the observations are more accurate, the state on the upper boundary can be reconstructed. With smaller observational noise ( $\sigma = 0.01$ ), the Picard ratio values are closer to the values for perfect observations. Therefore less damping is required and hence it is appropriate to use a smaller value for  $\mu_{\text{specified}}^2$ . From Fig. 6, the pair of RSVs that are needed to reconstruct the large-scale structure on the upper boundary have a squared singular value of  $\lambda^2 = 0.18$ . The filter factors (9) show that  $\mu_{\text{specified}}^2$  needs to be smaller than  $\lambda^2$  for the corresponding RSV to be incorporated into the analysis increment. Therefore, when the algorithm uses  $\mu_{\text{specified}}^2 = 4 \times 10^{-3}$ , it is able to infer the state on the unobserved upper boundary. When  $\mu_{\text{specified}}^2 = 1$ , the algorithm is unable to infer the upper level wave, as the analysis from Exp. 6 shows. In this last experiment,  $\mu_{\text{specified}}^2 > \mu_{\text{actual}}^2$  so that not only do the filter factors damp the RSVs corresponding to the observational noise, but they damp the RSVs needed to reconstruct the upper level wave as well. Consequently the useful information that is contained in the observations is lost.

## 5 Conclusions

The filtering and interpolating properties of 4D-Var have been investigated using simple idealized case-studies with the 2D Eady model. The 4D-Var analyses have been interpreted using a new technique involving the SVD of the normalized observability matrix. The main results can be summarized as follows:

- The 4D-Var analysis increments can be written as a linear combination of the RSVs of the normalized observability matrix. The weight given to an RSV is determined by the projection of the innovation vector onto the corresponding LSV and the filter factor

which acts to damp the RSVs with singular values comparable to or smaller than the ratio of the standard deviation of the observational and background state errors.

- The 4D-Var algorithm optimally combines the information from the background state and the observations. Observations are noisy and therefore it is important to specify the appropriate background error correlations so that the algorithm is able to extract the signal whilst filtering the noise. With a longer correlation length-scale, the RSVs are re-ordered and the RSVs with small-scale structures have reduced singular values and are thus more heavily damped by the filter factor; this gives a smoother analysis.
- If the specified variance ratio,  $\mu_{\text{specified}}^2$ , is smaller than the actual variance ratio,  $\mu_{\text{actual}}^2$ , the algorithm draws too close to the observations so that the analysis contains unrealistic structures that have large amplitudes in the unobserved regions. From an SVD perspective, the RSVs with small singular values have small-scale structures and large amplitudes in the unobserved regions. The observational noise has a large projection onto these RSVs, and therefore they dominate the solution unless they are filtered from the solution through a value of  $\mu_{\text{specified}}^2$  greater than the square of their singular values.
- The specification of the variance ratio in the 4D-Var algorithm is critical so that all the useful information contained in the observations is extracted, whilst filtering the observational noise. If  $\mu_{\text{specified}}^2$  is smaller than  $\mu_{\text{actual}}^2$ , the analysis contains unrealistic structures. Alternatively, if  $\mu_{\text{specified}}^2$  is larger than  $\mu_{\text{actual}}^2$ , the filter factors damp the structures corresponding to the observational noise, but also damp the structures corresponding to the signal. In particular, the information that is needed to reconstruct the state in the unobserved regions is damped.
- 4D-Var is able to interpolate through observations distributed in time to infer the state in unobserved regions. This has been illustrated by observing the lower level of a baroclinic wave and reconstructing the upper level. We have also shown that the algorithm is only able to reconstruct the upper level wave if the growth of the true solution over the assimilation window can be detected from the noisy observations. This requires that the ratio of the standard deviation of the observational and background errors be less than the singular value of the RSV that has the relevant structure in the unobserved region.

## 6 Discussion

We have shown that the use of the SVD of the normalized observability matrix provides a useful framework to examine the behaviour of 4D-Var. The RSVs do not depend on either the innovation vector or the variance ratio and therefore more general conclusions may be obtained than from individual 4D-Var analyses. This framework provides insight into the spatial and temporal interpolation resulting from the interaction of a time-sequence of observations with the model dynamics, for example, to infer the state in unobserved regions and to provide the vertical structures that are necessary for baroclinic growth.

This paper has focussed on the filtering properties of the 4D-Var algorithm. The results have illustrated the importance of specifying the a priori error statistics so that the maximum amount of useful information, particularly that for the unobserved regions, may be extracted from the observations. If the relative magnitude of the observational and background state

errors is underestimated then useable information for the unobserved regions is rejected. If it is overestimated, then false structures may be analysed, particularly in the unobserved regions. To be able to maximize the benefits of 4D-Var, it is important to draw close to the true state, but this may give poor analyses if the observations are inaccurate. Alternatively, if the algorithm is tuned so that it does not draw so close to the observations, we can expect the observational noise to be filtered, but may not benefit from the reconstructive ability of 4D-Var.

In operational data assimilation, the background error statistics are specified from climatological information. This could be improved by varying the statistics for each analysis so that they are consistent with the actual background state errors. An important question to address is whether it is possible to obtain the appropriate value for the variance ratio from the data itself. Various methods have been proposed, such as the L-Curve (Hansen 2001), Generalized Cross Validation (Gong *et. al* 1998), Maximum Likelihood techniques (Dee 1995), and Iterative Tuning (Desroziers and Ivanov 2001), but further investigation is required.

The SVD of the observability matrix has also been used to investigate the impacts of the variance ratio and the spatial and temporal position of the observations on the vertical structure of 4D-Var analysis increments (Johnson 2003). Results from this work will be reported in a future paper.

## 7 Acknowledgements

The authors thank S. P. Ballard from the Met Office and A. S. Lawless from the University of Reading for their input throughout the project. The authors are also grateful to A. Hollingsworth, from the European Centre for Medium Range Weather Forecasts, who provided many useful comments and suggestions for improvement. Christine Johnson was supported by a NERC CASE studentship with the Met Office as the cooperating organization.

## 8 References

- Badger, J. and Hoskins, B. J. 2001 Simple initial value problems and mechanisms for baroclinic growth. *J. Atmos. Sci.*, **58**, 38–49
- Callies, U. and Eppel, D. P. 1995 On the sensitivity of a least-squares fit of discretized linear hyperbolic equations to data. *Tellus*, **47A**, 80–102
- Courtier, P. and Talagrand, O. 1987 Variational assimilation of meteorological observations with the adjoint vorticity equation (II): Numerical Results. *Q. J. R. Meteorol. Soc.*, **113**, 1329–1368
- Courtier, P., Andersson, E., Heckley, W., Pailleux, J., Vasiljević, D., Hamrud, M., Hollingsworth, A., Rabier, F. and Fisher, M. 1998 The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Q. J. R. Meteorol. Soc.*, **124**, 1783–1807
- Dee, D. 1995 On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Weather Rev.*, **123**, 1128–1145
- Desroziers, G. and Ivanov, S. 2001 Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Q. J. R. Meteorol. Soc.*, **127**, 1433–1452
- Desroziers, G., Pouponneau, B., Thépaut, J.-N., Janisková, M. and Veersé, F. 1999 Four-dimensional variational analyses of FASTEX situations using special observations. *Q. J. R. Meteorol. Soc.*, **125**, 3339–3358
- Eady, E. T. 1949 Long waves and cyclone waves. *Tellus*, **1**, 33–52
- Golub, G. H. and Van Loan, C. F. 1996 *Matrix Computations*. The John Hopkins University Press
- Gong, J., Wahba, G., Johnson, D. R. and Tribbia, J. 1998 Adaptive tuning of numerical weather prediction models: Simultaneous estimation of weighting, smoothing, and physical parameters. *Mon. Weather Rev.*, **126**, 210–231
- Hansen, P. C. 2001 ‘The L-curve and its use in the numerical treatment of inverse problems’. Pp. 119–142 in *Computational Inverse Problems in Electrocardiology*, Ed. P. Johnstone, WIT Press, Southampton
- Hollingsworth, A. and Lönnberg, P. 1986 The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus.*, **38A**, 111–136
- Ide, K., Courtier, P., Ghil, M. and Lorenc, A. C. 1997 Unified notation for data assimilation: Operational, sequential and variational. *J. Meteorol. Soc. Jpn.*, **75**, 181–189
- Johnson, C. 2003 ‘Information content of observations in variational data assimilation’. PhD thesis, Department of Meteorology, University of Reading

- Laroche, S. and Gauthier, P. 1998 A validation of the incremental formulation of 4D variational data assimilation in a nonlinear barotropic flow. *Tellus*, **50A**, 557–572
- Le Dimet, F. X. and Talagrand, O. 1986 Variational algorithms for analysis and assimilation of meteorological observations. *Tellus*, **38A**, 97–110
- Lorenc, A. C. 1986 Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177–1194
- Mateer, C. L. 1965 On the information content of Umkehr observations. *J. Atmos. Sci.*, **22**, 370–381
- Prunet, P., Thépaut, J.-N. and Cassé, V. 1998 The information content of clear sky IASI radiances and their potential for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **124**, 211–241
- Rabier, F. and Courtier, P. 1992 Four-dimensional assimilation in the presence of baroclinic instability. *Q. J. R. Meteorol. Soc.*, **118**, 649–672
- Rabier, F., Thépaut, J.-N. and Courtier, P. 1998 Extended assimilation and forecast experiments with a four-dimensional variational data assimilation system. *Q. J. R. Meteorol. Soc.*, **124**, 1861–1887
- Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J. and Simmons, A. 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., Fourrié, N., Chafaï, D. and Prunet, P. 2002 Channel selection methods for Infrared Atmospheric Sounding Interferometer radiances. *Q. J. R. Meteorol. Soc.*, **128**, 1011–1027
- Rodgers, C. D. 2000 *Inverse Methods for Atmospheric Sounding, Theory and Practice*. Series on Atmospheric, Oceanic and Planetary Physics, World Scientific
- Sasaki, Y. 1970 Some basic formalisms in numerical variational analysis. *Mon. Weather. Rev.*, **98**, 875–883
- Schröter, J., Seiler, U. and Wenzel, M. 1993 Variational Assimilation of Geosat Data into an Eddy-resolving Model of the Gulf Stream Extension Area. *J. Phys. Oceanogr.*, **23**, 925–953
- Shanno, D. F. and Phua, K. H. 1976 Algorithm 500: minimisation of unconstrained multivariate functions. *ACM Trans. Math. Software*, **2**, 87–94
- Tanguay, M., Bartello, P. and Gauthier, P. 1995 Four-dimensional data assimilation with a wide range of scales. *Tellus*, **47A**, 974–997
- Thépaut, J.-N. and Moll, P. 1990 Variational inversion of simulated TOVS radiances using the adjoint technique. *Q. J. R. Meteorol. Soc.*, **116**, 1425–1448

- Thépaut, J.-N., Hoffman, R. N. and Courtier, P. 1993 Interactions of dynamics and observations in a four-dimensional variational assimilation. *Mon. Weather Rev.*, **121**, 3393–3414
- Thompson, P. D. 1961 A dynamical method of analyzing meteorological data. *Tellus*, **13**, 334–349
- Toumazou, V. and Cretaux, J. F. 2001 Using a Lanczos eigensolver in the computation of empirical orthogonal functions. *Mon. Weather Rev.*, **129**, 1243–1250
- Winkler, J. R. 1997 Tikhonov regularisation in standard form for polynomial basis conversion. *Appl. Math. Modelling*, **21**, 651–662
- Wunsch, C. 1977 Determining the general circulation of the oceans: a preliminary discussion. *Science*, **196**, 871–875
- Wunsch, C. 1996 *The Ocean Circulation Inverse Problem*. Cambridge University Press
- Zou, X., Navon, I. M. and Le Dimet, F. X. 1992 Incomplete observations and control of gravity waves in variational data assimilation. *Tellus*, **44A**, 273–296