# A SINGULAR ECTOR

# PERSPECTI E OF D AR

# FILTERING AND INTERPOLATING

C. Johnson [1], N. K. Nichols [2], B. J. Hoskins[1]

**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 O□ce through a CASE studentship.**

[1] Department of Meteorology, University of Reading, UK
[2] Department of Mathematics, University of Reading, UK

# A str ct

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

# List of Figures

# 1 Introduction

**In weather forecasting, data assimilation is used to genera**

is known in control theory as the observability matrix (Zou et al.

using the notation $\|\mathrm{z}\|_2^2 = \mathrm{z}^{\mathrm{T}}\mathrm{z}$, and where $\mathrm{d} = \mathrm{y} - \mathrm{H}\mathrm{x}^{\mathrm{b}}$ is the generalized innovation vector. $\rho_{\mathbf{R}}^{-1/2}$ is the symmetric square root of $\rho_{\mathbf{R}}^{-1}$

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

$f_j$ [illegible]

| Experiment | $\mu^2_{\text{actual}}$ | $\mu^2_{\text{specified}}$ | $l_{\text{actual}}(km)$ | $l_{\text{specified}}(km)$ | 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 | di�erent random seed for o |
| 5 | $4 \times 10^{-3}$ | $4 \times 10^{-3}$ | 1000 | 1000 | |
| 6 | $4 \times 10^{-3}$ | 1 | 1000 | 1000 | |

T le Summary of the parameters used in the six 4D-Var experiments. $\mu^2_{actual} = \sigma^2_{\mathbf{o}}/\sigma^2_{\mathbf{b}}$ is the actual variance ratio, $\mu^2_{specified}$ is the variance ratio that is used by the 4D-Var algorithm, $l_{actual}$ is the length-scale used to generate the background state, and $l_{specified}$ 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:

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

From hydrostatic balance, the boundary conditions are:

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

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

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

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 mode(e)4.4445v d;

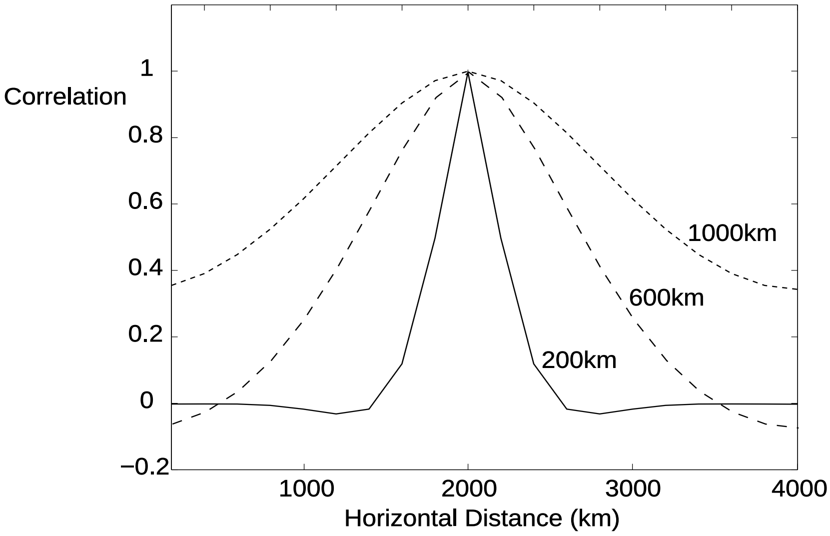
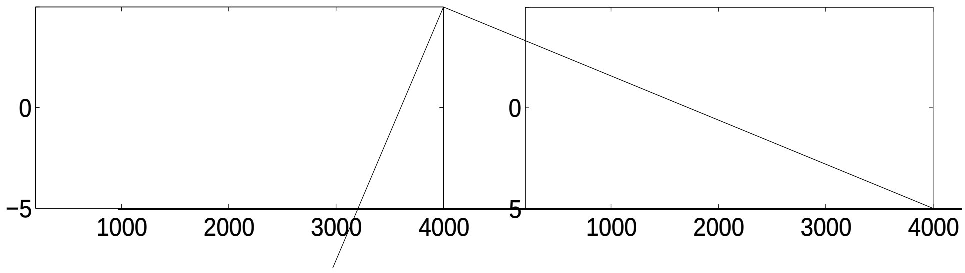


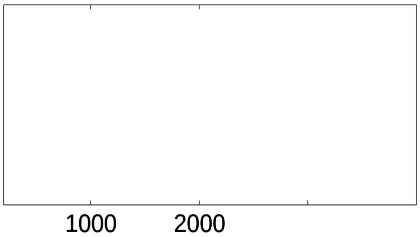

Figure **Horizontal auto-correlation of the background state error for correlation length-scales $l$ = 200$km$, 600$km$, and 1000$km$.**

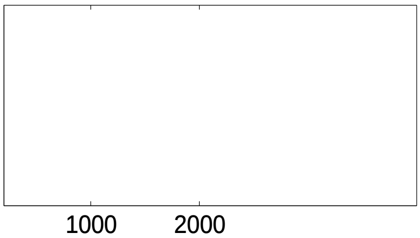
## B ckground St te

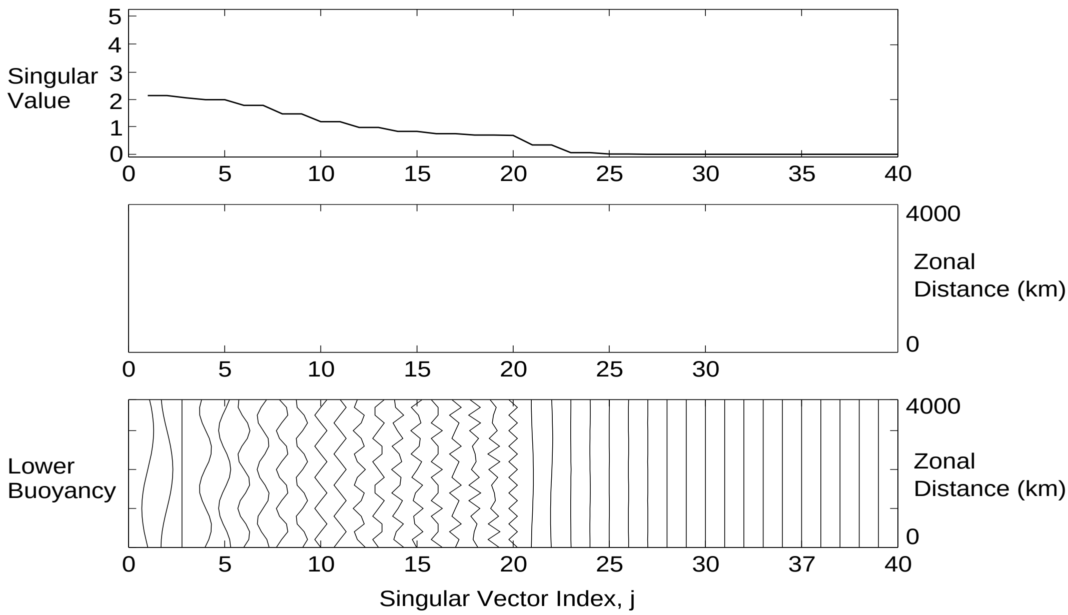
**The background state is defined by the true state with correlated random errors;**

$$x^{b} = x_0^{t} + \rho_{\mathbf{a}}^{1/2} \rho$$

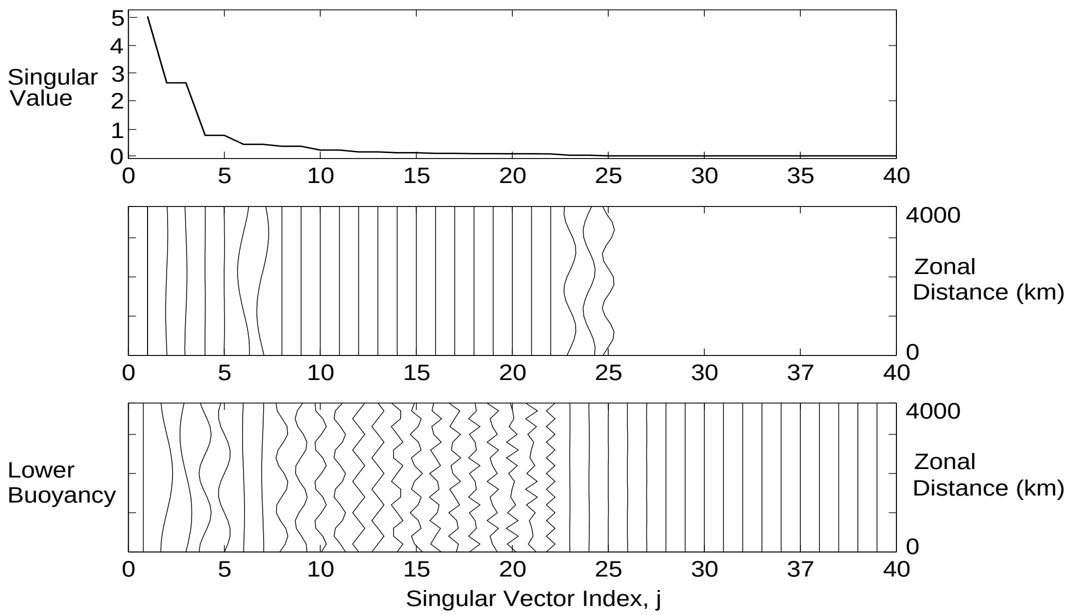

0
−5
1000
2000
3000
4000
0
5
1000
2000
3000
4000


1000
2000


1000
2000



Singular Value
5
4
3
2
1
0
0
5
10
15
20
25
30
35
40
4000
Zonal Distance (km)
0
0
5
10
15
20
25
30
Lower Buoyancy
4000
Zonal Distance (km)
0
0
5
10
15
20
25
30
37
40
Singular Vector Index, j


Singular Value
Zonal Distance (km)
Lower Buoyancy
Zonal Distance (km)
Singular Vector Index, j

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, $\mu^2_{\text{specified}}$, is smaller than the actual variance ratio, $\mu^2_{\text{actual}}$, 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 $\mu^2_{\text{specified}}$ 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 $\mu^2_{\text{specified}}$ is smaller than $\mu^2_{\text{actual}}$, the analysis contains unrealistic structures. Alternatively, if $\mu^2_{\text{specified}}$ is larger than $\mu^2_{\text{actual}}$, the filter factors damp the 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.

# 8 References


| | | |
|---|---|---|
| Badger, J. and Hoskins, B. J. | 2001 | Simple initial value problems and mechanisms for baroclinic growth. J. Atmos. Sci., , 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, A, 80–102 |

| | | |
|---|---|---|
| Laroche, S. and Gauthier, P. | 1998 | A validation of the incremental formulation of 4D variational data assimilation in a nonlinear barotropic flow. Tellus, A, 557–572 |
| Le Dimet, F. X. and Talagrand, O. | 1986 | Variational algorithms for analysis and assimilation of meteorological observations. Tellus, A, 97–110 |
| Lorenc, A. C. | 1986 | Analysis methods for numerical weather prediction. Q. J. R. Meteorol. Soc.ric |

| | | |
|---|---|---|
| Thépaut, J.-N., Hoff-man, R. N. and Courtier, P. | 1993 | Interactions of dynamics and observations in a four-dimensional variational assimilation. Mon. Weather Rev., , 3393–3414 |
| Thompson, P. D. | 1961 | A dynamical method of analyzing meteorological data. Tellus, , 334–349 |
| Toumazou, V. and Cretaux, J. F. | 2001 | Using a Lanczos eigensolver in the computation of empirical orthogonal functions. Mon. Weather Rev., , 1243–1250 |
| Winkler, J. R. | 1997 | Tikhonov regularisation in standard form for polynomial basis conversion. Appl. Math. Modelling, , 651–662 |
| Wunsch, C. | 1977 | Determining the general circulation of the oceans: a preliminary discussion. Science, , 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, A, 273–296 |