

## Singular vectors and estimates of the analysis-error covariance metric

By JAN BARKMEIJER<sup>1\*</sup>, MARTIN VAN GIJZEN<sup>2</sup> and FRANÇOIS BOUTTIER<sup>1</sup>

<sup>1</sup>European Centre for Medium-Range Weather Forecasts, UK

<sup>2</sup>TNO Physics and Electronics Laboratory, The Netherlands

(Received 27 February 1997; revised 15 September 1997)

### SUMMARY

An important ingredient of ensemble forecasting is the computation of initial perturbations. Various techniques exist to generate initial perturbations. All these aim to produce an ensemble that, at initial time, reflects the uncertainty in the initial condition. In this paper a method for computing singular vectors consistent with current estimates of the analysis-error statistics is proposed and studied. The singular-vector computation is constrained at initial time by the Hessian of the three-dimensional variational assimilation (3D-Var) cost function in a way which is consistent with the operational analysis procedure. The Hessian is affected by the approximations made in the implementation of 3D-Var; however, it provides a more objective representation of the analysis-error covariances than other metrics previously used to constrain singular vectors.

Experiments are performed with a T21L5 Primitive-Equation model. To compute the singular vectors we solve a generalized eigenvalue problem using a recently developed algorithm. It is shown that use of the Hessian of the cost function can significantly influence such properties of singular vectors as horizontal location, vertical structure and growth rate. The impact of using statistics of observational errors is clearly visible in that the amplitude of the singular vectors reduces in data-rich areas. Finally, the use of an approximation to the Hessian is discussed.

KEYWORDS: Cost function Ensemble forecasting Forecast skill Hessian singular vectors

### 1. INTRODUCTION

Ensemble prediction becomes more and more part of the daily routine at various numerical weather-prediction (NWP) centres in order to assess, a priori, whether a forecast will be skilful or unskilful. The development of the ensemble prediction systems (EPS) follows the pioneering work of Lorenz (1965), Epstein (1969) and Leith (1974). Central to it is the generation of different forecasts which at initial time somehow reflect the uncertainty in representing the initial conditions. The approaches followed by the various NWP centres to create perturbations to the initial conditions differ substantially. At the US National Center for Environmental Prediction the *breeding method* (Toth and Kalnay 1997) is used to obtain initial perturbations, and Houtekamer *et al.* (1996) advocate a generalization of this method as base for the EPS at the Canadian Meteorological Centre. In their *system simulated approach*, model errors are also taken into account by allowing different model configurations in the ensemble. At the European Centre for Medium-Range Weather Forecasts (ECMWF) *singular vectors* are used to generate initial perturbations for the EPS, e.g. Molteni *et al.* (1996).

An intercomparison of the different ensemble forecasting systems has not yet been made but is planned in the near future. One aspect of the ensemble forecast performance which is often observed is that the spread between the perturbed forecasts and the unperturbed forecast in the ensemble is smaller than the difference between the forecast and the validating analysis (e.g. Van den Dool and Rukhovets 1994; Buizza 1995). Our understanding of why this occurs is still fragmentary. It may be that the initial perturbations have the wrong amplitude or are not positioned correctly. Sensitivity patterns (Rabier *et al.* 1996) may provide some guidance on this. For the medium range (three days onwards), model errors certainly play an additional role in producing forecast errors. Therefore one may argue that an EPS must also account for perturbations in model parameters. In any case, a prerequisite for a well-defined EPS is that the initial perturbations reflect the uncertainty of the initial conditions.

\* Corresponding author: European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG2 9AX, UK.

In this paper we focus on the latter point and assume that the model is perfect. We will determine fast-growing perturbations which at initial time are constrained by what is known about analysis-error statistics as obtained from a variational data-assimilation system. More precisely, let  $\mathbf{A}$  be the analysis-error covariance matrix and  $\mathbf{M}$  the forward tangent operator of the forecast model. We are interested in perturbations which evolve through  $\mathbf{M}$  in the dominant eigenvectors of the error covariance matrix of the forecast error as given by

$$\mathbf{F} = \mathbf{MAM}^*, \quad (1)$$

where  $\mathbf{M}^*$  is the adjoint of  $\mathbf{M}$  for a particular inner product. Ehrendorfer and Tribbia (1997) show that determining the leading eigenvectors of  $\mathbf{F}$  provides an efficient way to describe most important components of the forecast-error variance (see also Houtekamer (1995)). However, in their calculations they make explicit use of the square root of  $\mathbf{A}$ . Since the dimension of covariance matrices in current numerical weather models is of the order  $10^6$ , this is not a feasible assumption in an operational context. Also the direct computation of eigenpairs of  $\mathbf{F}$  by, for instance, applying a Lanczos algorithm to  $\mathbf{MAM}^*$ , is not possible. This is because the analysis-error covariance matrix is not available in the variational assimilation system. However, its inverse  $\mathbf{A}^{-1}$  is known: it is equal to the Hessian of the cost function  $\mathcal{J}$  in the incremental variational data assimilation (see section 2). Moreover, in the incremental formulation the cost function is quadratic (Fisher and Courtier 1995), and we may write  $\mathbf{A}^{-1}\mathbf{x} = \nabla\mathcal{J}(\mathbf{x} + \mathbf{y}) - \nabla\mathcal{J}(\mathbf{y})$ . This enables us to find dominant eigenpairs of (1) by solving the generalized eigenvalue problem

$$\mathbf{M}^*\mathbf{M}\mathbf{x} = \lambda\mathbf{A}^{-1}\mathbf{x}. \quad (2)$$

Observe that the evolved vectors  $\mathbf{M}\mathbf{x}$  are eigenvectors of  $\mathbf{F}$ . Solutions of (2) are the singular vectors of  $\mathbf{M}$  using the analysis-error covariance metric at initial time. In other contexts the metric is also known as the Mahalanobis metric (e.g. Mardia *et al.* 1979; Stephenson 1997). Palmer *et al.* (1998) addresses the question to what extent simpler metrics approximate the analysis-error covariance metric. Use of these simpler metrics avoids the need of a generalized eigenvalue problem solver.

Davidson's method (Davidson 1975) and the recently proposed Jacobi–Davidson method (Sleijpen and Van der Vorst 1996) can solve the generalized eigenvalue problem without using explicit knowledge of the operators on both sides of the equation. In section 3 we make use of this algorithm to determine singular vectors using the three-dimensional variational assimilation (3D-Var) Hessian as constraint at initial time. We also consider the impact of a simpler form of the cost function, where the background covariance matrix is replaced by a diagonal matrix so that, without using covariance information of observations, the analysis-error covariance metric becomes the reciprocal of the total-energy metric which is currently used at ECMWF to compute singular vectors for the EPS. The impact of background- and observational-error statistics is clearly evident through the change of the unstable subspace formed by the leading singular vectors. Section 4 shows results on approximating the Hessian through its leading eigenpairs. This may provide an alternative to using the generalized eigenvalue solver. Some concluding remarks are given in section 5. In the appendix the algorithm to solve the generalized eigenvalue problem is described.

## 2. THE HESSIAN OF THE 3D-VAR COST FUNCTION

In 3D-Var the state of the atmosphere,  $\mathbf{x}$ , is estimated by minimizing a cost function  $\mathcal{J}$  consisting of the weighted sums of the distances of  $\mathbf{x}$  from an a priori estimate,  $\mathbf{x}^b$ , of

the field considered and from a set of observations,  $\mathbf{y}$  (Courtier *et al.* 1998). The solution  $\mathbf{x}^a$  is called the analysis and  $\mathbf{x}^a$  minimizes

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y})^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}), \quad (3)$$

where  $H$  stands for the operator which evaluates the observed variables for  $\mathbf{x}$  at the observation location. The matrices  $\mathbf{B}$  and  $\mathbf{R}$  are covariance matrices for background errors and observations errors, respectively. The symbol  $^T$  denotes the transpose.

In the incremental formulation of 3D-Var, the nonlinear operator  $H$  is linearized in the vicinity of the background  $\mathbf{x}^b$ :  $H(\mathbf{x}) = H(\mathbf{x}^b) + \mathbf{H}(\mathbf{x} - \mathbf{x}^b)$ , where  $\mathbf{H}$  is the linearized form of  $H$ . Consequently, the Hessian  $\nabla\nabla\mathcal{J}$  of the cost function may be written as

$$\nabla\nabla\mathcal{J} = \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}. \quad (4)$$

Furthermore, the cost function  $\mathcal{J}$  is quadratic and its gradient is zero at the unique minimum:

$$\mathbf{B}^{-1}\{\mathbf{x}^a - \mathbf{x}^t - (\mathbf{x}^b - \mathbf{x}^t)\} + \mathbf{H}^T \mathbf{R}^{-1}\{\mathbf{H}(\mathbf{x}^a - \mathbf{x}^t) + \mathbf{H}\mathbf{x}^t - \mathbf{y}\} = 0, \quad (5)$$

where  $\mathbf{x}^t$  is the true state of the atmosphere. Rewriting (5) gives

$$(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})(\mathbf{x}^a - \mathbf{x}^t) = \mathbf{B}^{-1}(\mathbf{x}^b - \mathbf{x}^t) - \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{x}^t - \mathbf{y}). \quad (6)$$

Under the additional assumption that the background and observation error are uncorrelated, the above equation implies that

$$(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})\mathbf{A}(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}) = \mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}, \quad (7)$$

where  $\mathbf{A}$  denotes the covariance matrix of errors  $(\mathbf{x}^a - \mathbf{x}^t)$  in the analysis. The above identity follows by multiplying each side of (6) to the right with its transpose, and by using the fact that  $\mathbf{B}$  and  $\mathbf{R}$  are the covariance matrices of  $(\mathbf{x}^b - \mathbf{x}^t)$  and  $(\mathbf{H}\mathbf{x}^t - \mathbf{y})$ , respectively, and  $\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$  is a symmetric matrix. Combining (4) and (7) we conclude that the Hessian of the cost function is equal to the inverse of the analysis-error covariance matrix (see also Rabier and Courtier 1992; Fisher and Courtier 1995).

Indeed, this identity is only true insofar as the background  $\mathbf{B}$  and observation  $\mathbf{R}$  error covariance matrix have been correctly specified. In practice, many approximations are made in 3D-Var and the analysis is not optimal, so that in principle the estimation of the genuine  $\mathbf{A}$  can only be done by a much more complicated Kalman-filter-like algorithm as presented by Bouttier (1994). It is hoped that this will be implemented in the future in a simplified form of Kalman filter. In this paper we use the 3D-Var Hessian as a surrogate for the inverse analysis-error covariance matrix. The main problem with doing so is in the approximations made in the  $\mathbf{B}$  matrix, and it will be demonstrated in what follows that the singular vectors are indeed sensitive to the specification of  $\mathbf{B}$ .

Currently the  $\mathbf{B}$  matrix is based upon error statistics derived from the difference between two-day and one-day forecasts verifying at the same day (commonly known as the NMC method, Parrish and Derber (1992)). The statistics obtained for a period of 90 consecutive days extending from December 1992 to February 1993 are documented in Rabier *et al.* (1998). Power spectra of the error statistics indicate that the energy spectrum peaks around total wave number 10. Thépaut *et al.* (1996) have compared power spectra of 3D-Var and 4D-Var analysis increments in the vicinity of an extratropical storm. The 4D-Var approach clearly allowed for more energy in total wave numbers larger than 10. The NMC method is very efficient in estimating time-averaged global background-error

covariances which are well suited for data assimilation. Some cross-validation with other covariance estimation methods shows no evidence of any mismatch between the average structures of the forecast-error differences and of the short-range forecast errors themselves (Rabier *et al.* 1998). However, it is likely that the average covariance structures are not optimal in dynamically unstable areas. This is going to affect the Hessian itself, as the numerical results of section 3(a) suggest. It might be possible to optimize the **B** matrix for the specific needs of the singular-vector computation; however, it is not yet clear how to do so.

The central point made in this paper is that, despite these caveats, the 3D-Var Hessian provides a more realistic metric than simpler norms like total energy, or like the metric implied by **B** itself. This is because the Hessian contains a term,  $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$ , which is related to the observing network. Some low-resolution experiments (Bouttier 1994) suggest that the main effect of observations is to decrease the error variances and to sharpen the analysis-error correlations in data-rich areas. Hence one expects the Hessian metric to penalize the singular vectors more realistically than previously used norms, at least in data-rich areas.

### 3. THE ANALYSIS-ERROR COVARIANCE METRIC

In the computation of singular vectors it is necessary to specify a norm at initial time  $t_0$  and at optimization time  $t_1$ . In this section, singular vectors  $\boldsymbol{\epsilon}$  are considered which maximize the ratio

$$\frac{(\mathbf{P}\boldsymbol{\epsilon}(t_1), \mathbf{E}\mathbf{P}\boldsymbol{\epsilon}(t_1))}{(\boldsymbol{\epsilon}(t_0), \mathbf{C}\boldsymbol{\epsilon}(t_0))}. \quad (8)$$

Here  $(\cdot)$  denotes the Euclidean inner product  $(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \mathbf{y}$ . The positive definite and symmetric operators **C** and **E** induce a norm at initial and optimization time, respectively. The operator **P** is a projection operator setting a vector to zero outside a given domain, e.g. south of 30°N as in this paper. The first singular vector SV1 maximizes the ratio (8), the second singular vector SV2 maximizes (8) in the subspace **C**-orthogonal to SV1, and so forth. The evolved singular vectors form an **E**-orthogonal set at optimization time. Alternatively, these singular vectors are solutions of the following generalized eigenvalue problem,

$$\mathbf{M}^* \mathbf{P}^* \mathbf{E} \mathbf{P} \mathbf{M} \mathbf{x} = \lambda \mathbf{C} \mathbf{x}. \quad (9)$$

The adjoint operators  $\mathbf{M}^*$  and  $\mathbf{P}^*$  are determined with respect to the Euclidean inner product. In the computation of the singular vectors for the ECMWF EPS the total energy metric is used at initial and optimization time, i.e. **E** and **C** are identical and

$$\begin{aligned} (\mathbf{x}, \mathbf{E}\mathbf{y}) = & \frac{1}{2} \int_0^1 \int_{\Sigma} \left( \nabla \Delta^{-1} \zeta_x \cdot \nabla \Delta^{-1} \zeta_y + \nabla \Delta^{-1} D_x \cdot \nabla \Delta^{-1} D_y \right. \\ & \left. + \frac{c_p}{T_r} T_x T_y \right) d\Sigma \left( \frac{\partial p}{\partial \eta} \right) d\eta + \frac{1}{2} \int_{\Sigma} R_d T_r \ln \pi_x \cdot \ln \pi_y d\Sigma, \end{aligned} \quad (10)$$

with  $(\zeta_x, D_x, T_x, \ln \pi_x)$  being the vorticity, divergence, temperature and logarithm of the surface pressure components of the state vector  $\mathbf{x}$ ,  $c_p$  the specific heat of dry air at constant pressure,  $p(\eta)$  the pressure at eta levels (0 = surface and 1 = top of atmosphere),  $R_d$  the gas constant for dry air,  $T_r = 300$  K a reference temperature and  $\Sigma$  denoting integration over a sphere.

In this case **C** ( $\equiv$  **E**) has the form of a diagonal matrix and the square root of **C** can be readily determined. Multiplying both sides of (9) to the left and right with the inverse of the square root of **C** yields an equation which can be solved using the Lanczos

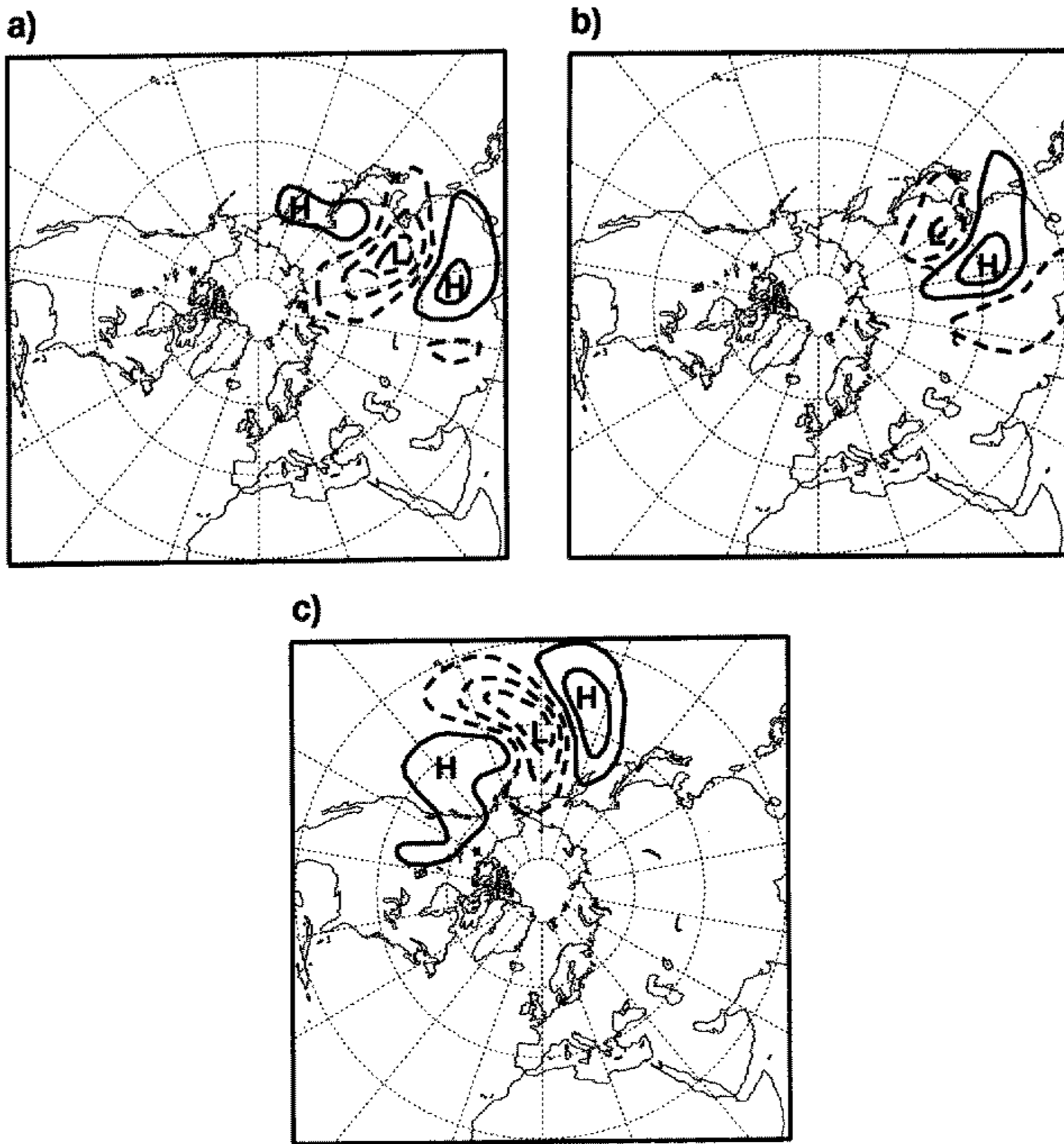


Figure 1. Stream-function fields of the three leading initial total-energy singular vectors at 500 hPa for 0000 UTC 1 November 1995 and an optimization time of one day. Solid (dashed) lines denote positive (negative) values and the contour interval is  $5 \times 10^{-8} \text{ m}^2 \text{ s}^{-1}$ .

algorithm (Strang 1986). Palmer *et al.* (1998) study the impact of choosing different simple metrics at initial time, keeping the total energy metric at optimization time. The spectra of the singular vectors can differ substantially for the various metrics. It turns out that, of the simple metrics considered, the total energy metric is the most consistent with the analysis-error statistics.

To study the impact of the analysis-error covariance metric on the singular vectors we used a five-level version of the ECMWF Integrating Forecast System model (Simmons *et al.* 1989; Courtier *et al.* 1991) with a triangular truncation at wave number 21 and a simple vertical diffusion scheme (Buizza 1994). The model levels are at 100, 300, 500, 700 and 900 hPa. The initial time for the singular vector computation is 0000 UTC 1 November 1995 and the optimization time is one day.

Figure 1 shows the stream-function fields of the first three total-energy singular vectors (TESV) at 500 hPa, i.e. the total-energy metric is used at both initial and optimization time. The two leading TESVs (Figs. 1(a) and (b)) are over the Asian area and the third SV

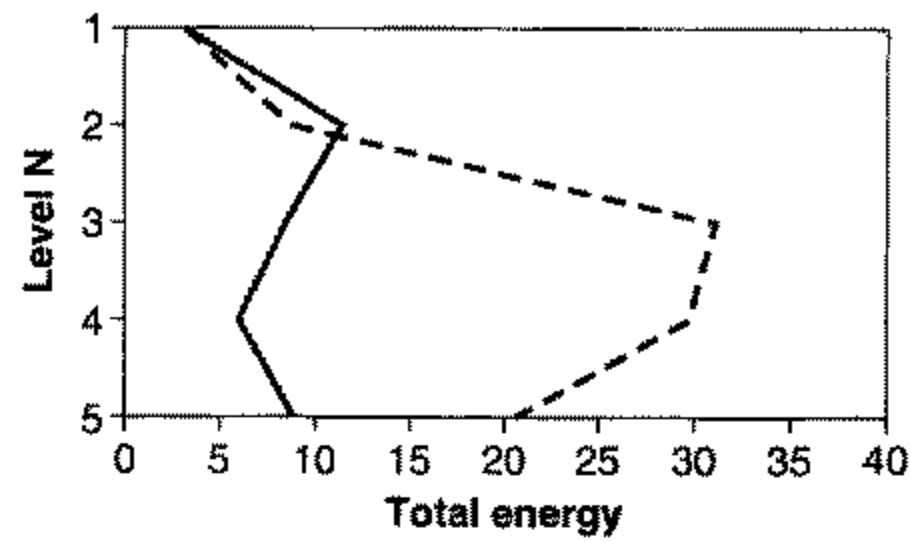


Figure 2. Vertical cross-section of the total-energy distribution of the ten leading total-energy singular vectors with respect to the five model levels. The dashed line represents the initial profile and the solid line represents the energy after 24 hours of linear integration. The energy in  $\text{m}^2\text{s}^{-1}$  at initial and final time has been multiplied by factors of 100 and 5 respectively.

is located over the Pacific. The vertical distributions of energy in the singular vectors at initial and optimization time are given in Fig. 2. It shows that the energy at initial time is confined to near the baroclinic steering level. During time evolution the energy propagates upwards and downwards, indicating error growth (Buizza and Palmer 1995).

When  $\mathbf{C}$  is specified to be equal to the full Hessian of the 3D-Var cost function, the operator  $\mathbf{C}$  is not known in matrix form, and determining its square root is not feasible. In order to solve (9), a generalized eigenvalue problem solver, called the generalized Davidson method, is used (see appendix). This algorithm can solve (9) efficiently and requires only the ability to calculate  $\mathbf{y} = \mathbf{S}\mathbf{x}$ , where  $\mathbf{S}$  is any of the operators  $\mathbf{M}$ ,  $\mathbf{P}$ ,  $\mathbf{E}$  and  $\mathbf{C}$ . No explicit knowledge of any operator is needed. In the following we assume that the total energy metric is always used at optimization time.

To improve the performance of the generalized Davidson algorithm, a coordinate transformation  $\boldsymbol{\chi} = \mathbf{L}^{-1}\mathbf{x}$  was carried out with  $\mathbf{L}\mathbf{L}^T = \mathbf{B}$ . Applying the transformation  $\mathbf{L}$ , the Hessian becomes equal to the sum of the identity and a matrix of rank less than or equal to the dimension of the vector of observations (see also Fisher and Courtier (1995)). Thus, when no observations are used in the cost function, the transformed operator  $(\mathbf{L}^{-1})^T\mathbf{C}\mathbf{L}^{-1}$  is the identity, and the proposed algorithm is equivalent to the Lanczos algorithm (see appendix). Observe that a similar coordinate transformation was used to compute the TESVs. Defining singular vectors using the total energy metric at initial time may be interpreted as using a background-error covariance matrix  $\mathbf{B}$  with the assumption of no horizontal and vertical correlation of forecast error.

#### (a) *Univariate background-error covariance matrix*

In the following, we refer to solutions of (9) using the full Hessian of the 3D-Var cost function as Hessian singular vectors HSV. Figure 3 shows the stream-function fields of the first three HSVs at 500 hPa when no error statistics of observations are used. The horizontal structure is more large-scale compared to the TESVs and the location of the amplitude maxima has changed completely. Also the vertical structure of the HSVs is quite different from the TESVs (see Fig. 4). Most of the HSV total energy is initially confined to the upper levels. At optimization time the vertical energy distribution of the TESVs and HSVs is comparable, although the total-energy amplification in the case of the HSVs is clearly smaller (see Fig. 5). It appears that the TESVs grow approximately twice as fast as the HSVs in terms of total energy. The large-scale structure and energy distribution of the HSVs is to a large extent determined by the formulation of the background covariance matrix. As explained in section 2, the distribution of the background-error variance (estimated using

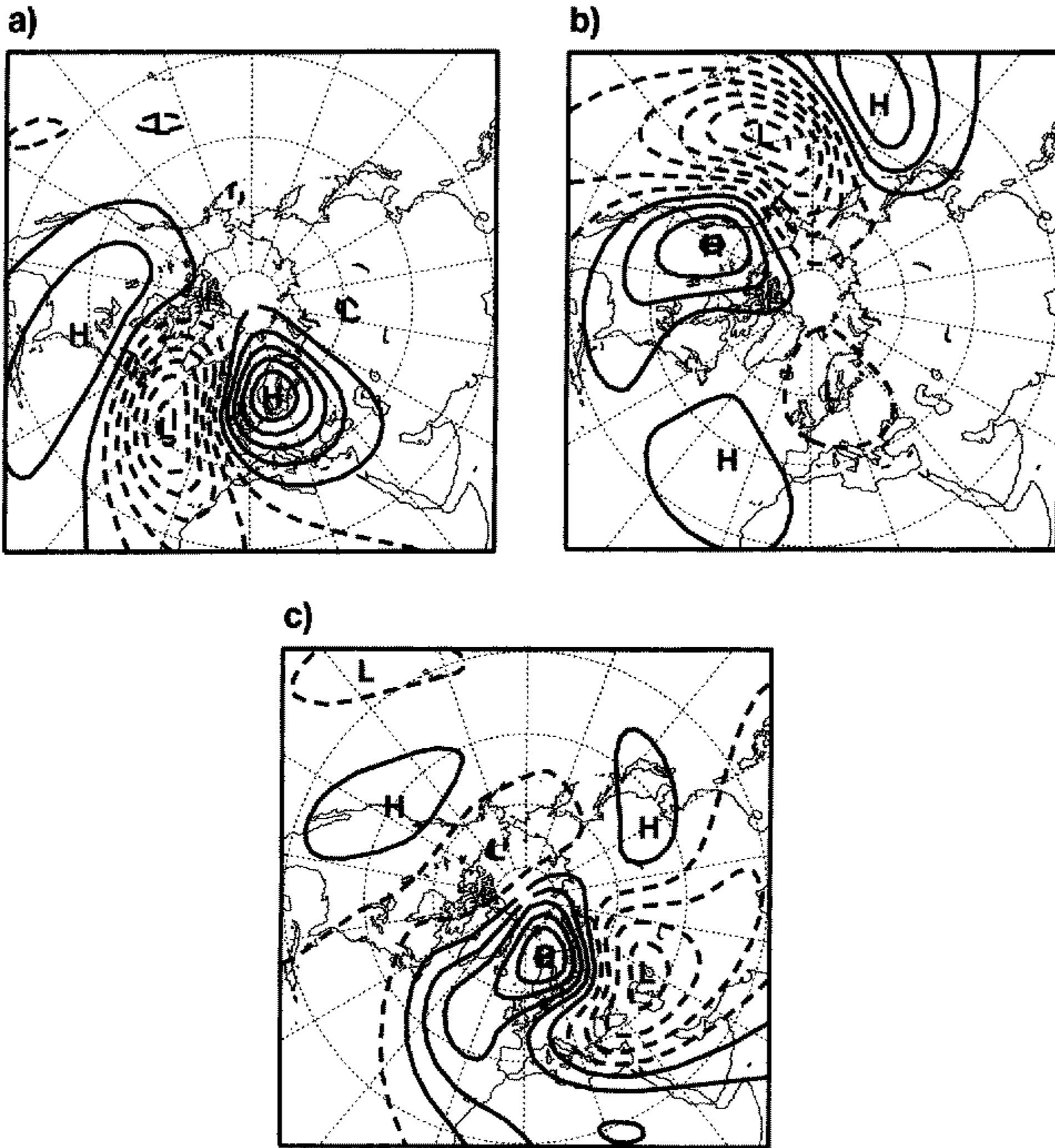


Figure 3. Same as Fig. 1 but for the Hessian singular vectors without observations.

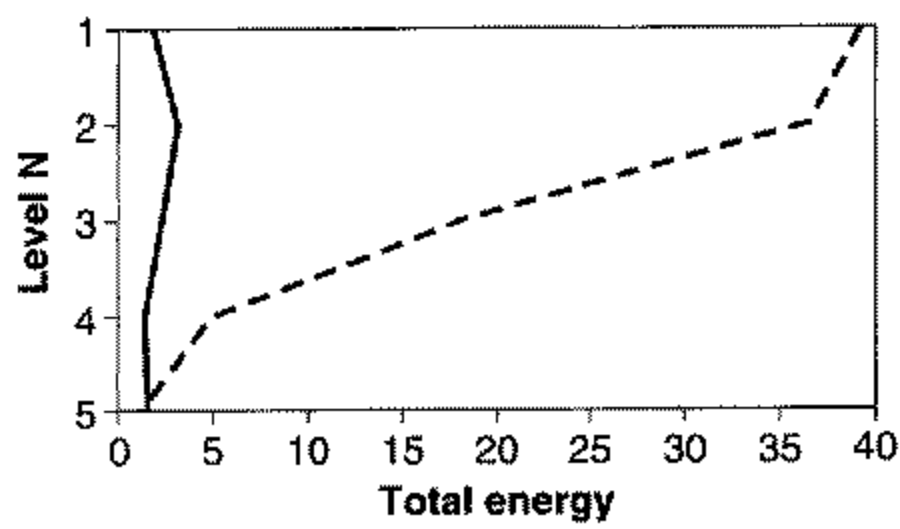


Figure 4. Same as Fig. 2 but for the Hessian singular vectors without observations.

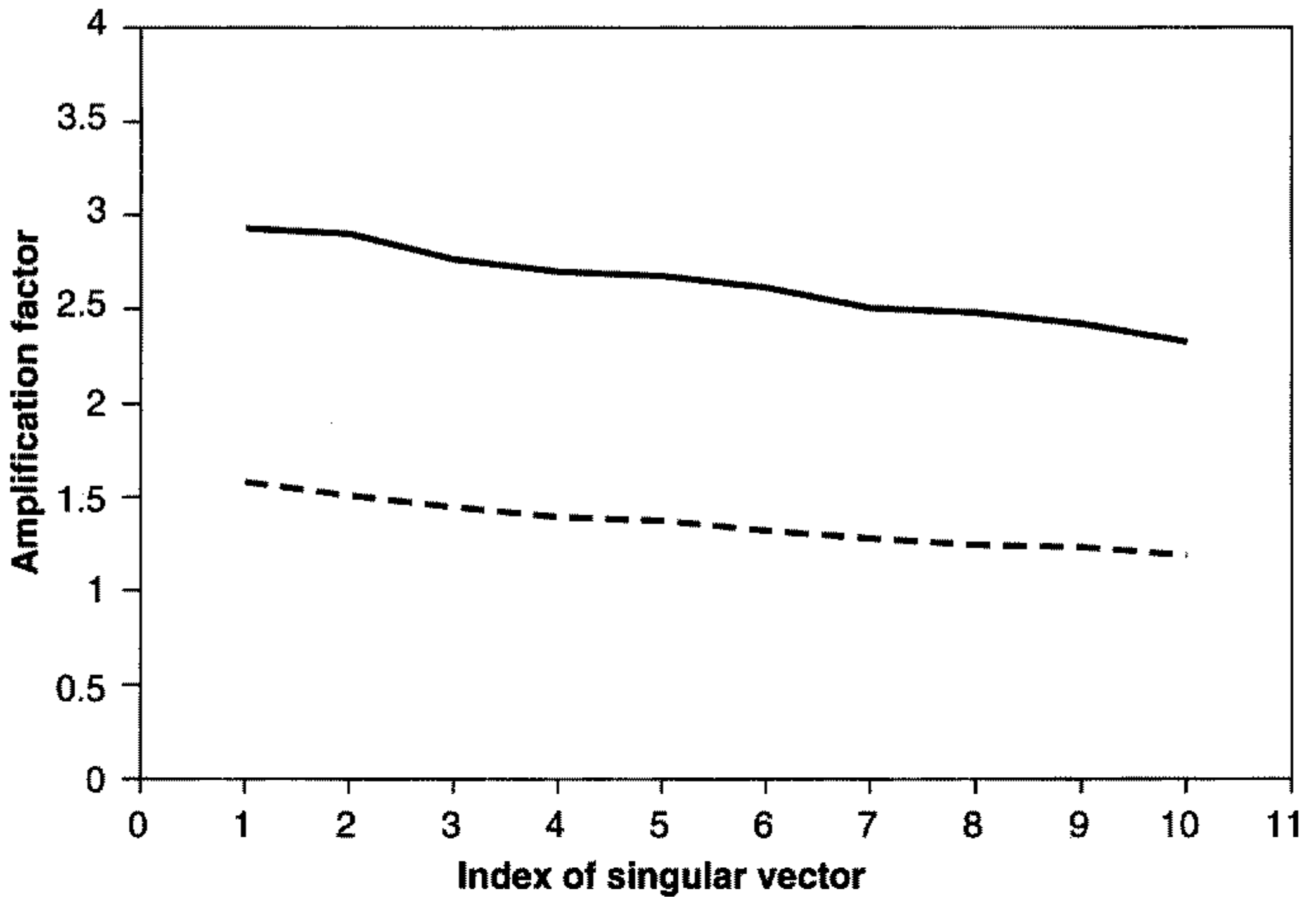


Figure 5. Amplification factor of the ten leading total-energy singular vectors (solid line) and Hessian singular vectors without observations (dashed line), as given by the square root of the ratio between the total energy at final and initial time.

forecast differences) reaches a maximum near the jet level, and this is reflected in the metric defined by the Hessian. Also, the background-error covariance matrix is specified to have broad horizontal and vertical correlations, and thus penalizes the occurrence of baroclinic structures in the analysis error. Hence we may suspect that the 3D-Var Hessian metric penalizes the sharp and baroclinic error patterns too much in the areas which are picked up by the singular vector computation.

Another way to highlight the differences between the TESVs and HSVs is to determine the similarity index of two unstable subspaces (Buizza 1994). To that end a Gram–Schmidt procedure is applied to the first ten TESVs and HSVs at initial time. Denote by  $\mathbf{W}^H$  and  $\mathbf{W}^{TE}$  the space thus obtained in the case of the HSVs and TESVs, respectively:

$$\mathbf{W}^{H(TE)} = \text{span}\{\mathbf{v}_j^{H(TE)}, j = 1, \dots, 10\}. \quad (11)$$

The similarity index  $\mathcal{S}$  of the two unstable subspaces  $\mathbf{W}^H$  and  $\mathbf{W}^{TE}$  is then defined by

$$\mathcal{S}(\mathbf{W}^H, \mathbf{W}^{TE}) = \frac{1}{10} \sum_{j,k=1}^{10} \langle \mathbf{v}_j^H, \mathbf{v}_k^{TE} \rangle^2, \quad (12)$$

where  $\langle \cdot, \cdot \rangle$  is the inner product associated with the total-energy norm. The similarity index  $\mathcal{S}$  goes from 1 for parallel unstable subspaces to 0 when the subspaces are orthogonal. The similarity index between the TESV and HSV unstable subspaces is only 0.06 indicating the different nature of the TESVs and HSVs. When subspaces were used spanned by the leading 15 singular vectors, the similarity index increased slightly to 0.11.



The impact of observational locations and errors on the HSVs was studied using covariance information from most of the conventional observations (SYNOP, AIREP, SATOB, DRIBU, TEMP, PILOT, SATEM and PAOB) from two areas: EUR (20°W–20°E; 30°N–80°N) and PAC (170°W–130°W; 30°N–80°N). Figures 6(a)–(c) show the first three HSVs at 500 hPa using observations from EUR. Compared to the HSVs without observations (Fig. 3), the amplitudes of the leading HSVs over EUR have decreased substantially. Including observational-error statistics left the vertical structure of the HSV basically unaltered (not shown). The similarity index of the two unstable subspaces, i.e. formed by HSVs computed with and without covariance information of observations from EUR, and spanned by the ten leading HSVs, is 0.79. Repeating the same for PAC gave similar results (see Figs. 6(d)–(f)). The amplitude of the SVs is now reduced over the PAC area at initial time. A comparison of the two unstable subspaces computed in the same way as the EUR yields a similarity index of 0.73.

(b) *Total-energy background covariance matrix with observations*

In order to investigate the impact of covariance information from observations on the TESVs, the 3D-Var background-error covariance matrix in the generalized eigenvalue calculation (9) was modified to a diagonal matrix whose diagonal contained the reciprocal of the total-energy weights (see (10)). The SVs are now computed in terms of the control variable  $\chi$  and we checked that, without observations, the same SVs resulted in terms of the model variables as defined by (9).

The SVs computed without observations do not depend on the scaling of the background covariance matrix. However, in order to account reasonably well for the way the observations are used in the 3D-Var analysis procedure, one needs to ensure that the relative magnitude of the background- and observation-error variances is balanced. Performing an analysis procedure with the modified background covariance matrix should give analysis increments which are consistent with the amplitude of the operational analysis increments. The appropriate magnitude of the total-energy background-error covariance matrix was determined by computing the analysis increments at a particular observation point for different scalings of the matrix and selecting one that implies sensible amplitudes for the increments.

In the 3D-Var analysis procedure, the analysis increments,  $\mathbf{x}^a - \mathbf{x}^b$ , are a linear combination of the observation increments,  $\mathbf{y} - \mathbf{H}\mathbf{x}^b$ , (Lorenz 1986; (A.9)):

$$\mathbf{x}^a - \mathbf{x}^b = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}^b), \quad (13)$$

where  $\mathbf{H}$  is the linearization of the observation operator  $H$  in the vicinity of the background  $\mathbf{x}^b$ . In the case of one observation,  $\mathbf{H}^T$  reduces to a vector  $\mathbf{h}^T$  of the same dimension as the model state vector, and the specified background error  $\sigma_b^2$  for the observed parameter satisfies  $\sigma_b^2 = \mathbf{h}\mathbf{B}\mathbf{h}^T$ . Denoting the observation variance by  $\sigma_o^2$ , (13) can be written as

$$\mathbf{x}^a - \mathbf{x}^b = \mathbf{B}\mathbf{h}^T \left( \frac{\mathbf{y} - \mathbf{H}\mathbf{x}^b}{\sigma_b^2 + \sigma_o^2} \right). \quad (14)$$

Applying  $\mathbf{h}$  to both sides of (14) yields for the analysis increments at the observation point

$$\mathbf{h}(\mathbf{x}^a - \mathbf{x}^b) = \mathbf{h}\mathbf{B}\mathbf{h}^T \left( \frac{\mathbf{y} - \mathbf{H}\mathbf{x}^b}{\sigma_b^2 + \sigma_o^2} \right) = \left( \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2} \right) (\mathbf{y} - \mathbf{H}\mathbf{x}^b). \quad (15)$$

Using the linearization  $\mathbf{H}$  of the observation operator in the vicinity of the background, we may write to first order

$$\mathbf{h}(\mathbf{x}^a - \mathbf{x}^b) = \mathbf{H}\mathbf{x}^a - \mathbf{H}\mathbf{x}^b. \quad (16)$$

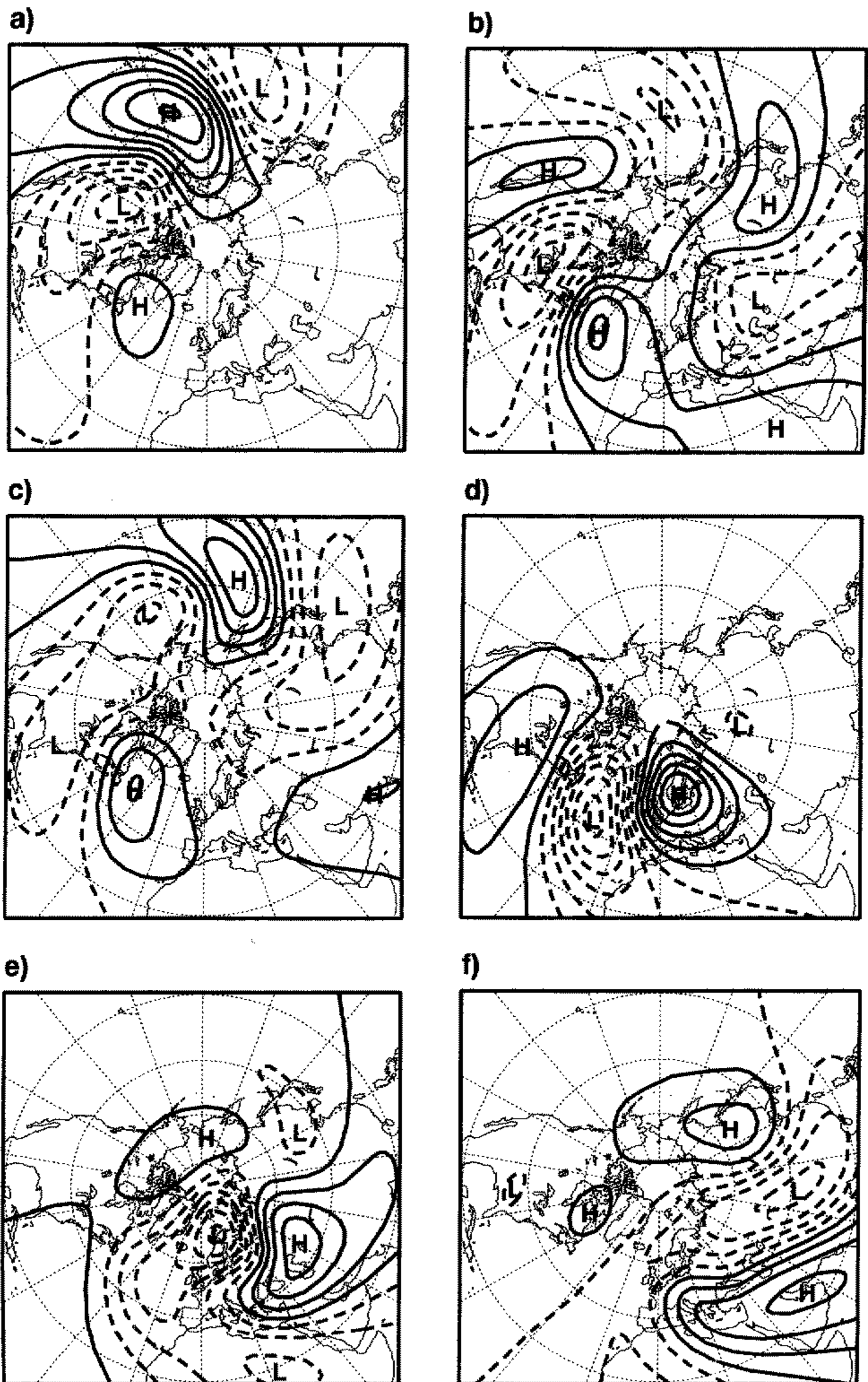


Figure 6. Same as Fig. 3 but for Hessian singular vectors with observations from (a)–(c) EUR (20°W–20°E; 30°N–80°N) and (d)–(f) PAC (170°W–130°W; 30°N–80°N).

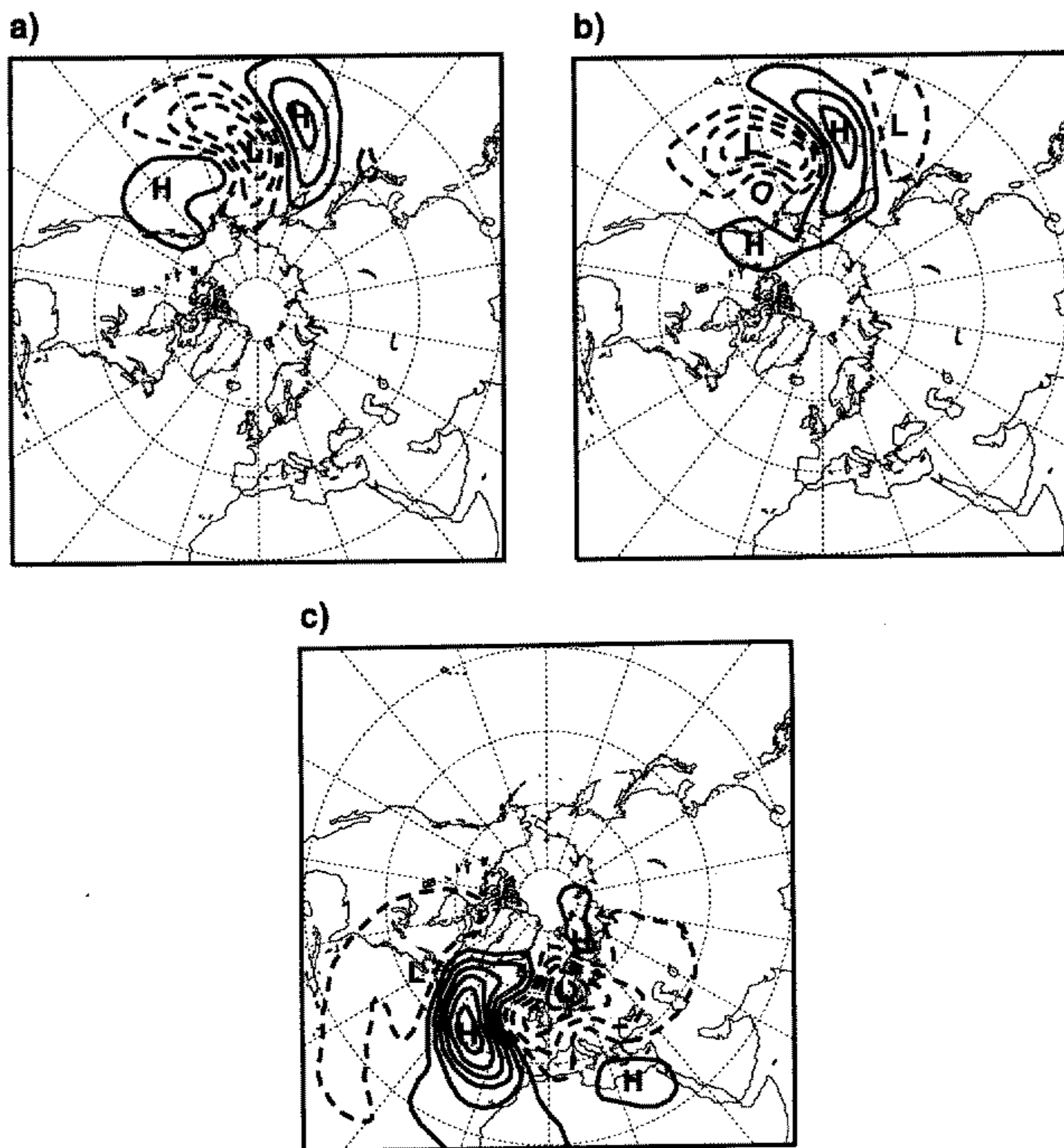


Figure 7. Same as Fig. 1 but now with observations from the northern hemisphere extratropics.

Combining (15) and (16) we obtain

$$\sigma_b^2 = \sigma_o^2 \left( \frac{r}{1-r} \right), \text{ with } r = \frac{\mathbf{Hx}^a - \mathbf{Hx}^b}{\mathbf{y} - \mathbf{Hx}^b}. \quad (17)$$

To determine the appropriate magnitude of the variances in  $\mathbf{B}$ , a 3D-Var minimization was performed using error statistics of only one observation. The scaling of  $\mathbf{B}$  was tuned so that the  $\sigma_b$  estimated from (17) is approximately equal to  $\sigma_o$  in the case of an AIREP observation of temperature at 500 hPa ( $\sigma_o = 2$  K). These values for  $\sigma_o$  and  $\sigma_b$  are assumed to be reasonable approximations of actual errors for temperature in the areas of interest for the SVs.

Figure 7 shows the three leading TESVs when covariance information of observations is included in the computation. The same type of observations are used as for PAC and EUR but now for the northern hemisphere extratropics ( $30^\circ\text{N}$ – $90^\circ\text{N}$ ). Observe that the leading SVs are located over the oceans. Most of the observational data over the oceans

are from satellites, which have less weight in the 3D-Var cost function than conventional data. Comparing these SVs with the TESVs when no observational information is used (see Fig. 1) yields a similarity index of 0.5. We conclude that incorporating statistics of observational errors in the SV computation may change the unstable subspace considerably.

#### 4. APPROXIMATION OF THE HESSIAN

A possible alternative to using a generalized eigenvalue problem solver is to replace the full Hessian  $\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$  (see (4)) by an approximating operator which can easily be inverted. In this way, solving a generalized eigenvalue problem can be reduced to solving an ordinary eigenvalue problem. As noted in section 3, the Hessian in terms of the control vector  $\boldsymbol{\chi} = \mathbf{L}^{-1} \mathbf{x}$  is the sum of the identity  $\mathbf{I}$  and a matrix of rank less than or equal to the number of observations, and can be written as (see also Fisher and Courtier 1995):

$$\mathbf{I} + \sum_{i=1}^K (\lambda_i - 1) \mathbf{u}_i \mathbf{u}_i^T, \quad (18)$$

where  $\lambda_i$  and  $\mathbf{u}_i$  are eigenpairs of the Hessian and  $K$  denotes the number of observations. Observe that indeed  $\{\mathbf{I} + \sum_{i=1}^K (\lambda_i - 1) \mathbf{u}_i \mathbf{u}_i^T\} \mathbf{u}_k = \mathbf{u}_k + (\lambda_k - 1) \mathbf{u}_k = \lambda_k \mathbf{u}_k$ , for  $k = 1, \dots, K$ , because of the orthogonality of the eigenvectors  $\mathbf{u}_i$ . Since the Hessian can be written in the form (18), the square root of the inverse of the Hessian can be readily determined using the Sherman–Morrison and Woodbury formula (Wait 1979) and it is given by

$$\mathbf{I} + \sum_{i=1}^K \left( \frac{1}{\sqrt{\lambda_i}} - 1 \right) \mathbf{u}_i \mathbf{u}_i^T. \quad (19)$$

Multiplying both sides of (9) to the left and right with the above operator yields an eigenvalue problem which again can be solved with the Lanczos algorithm.

We consider two situations to see how well SVs with an approximated Hessian compared to SVs obtained without approximation of the Hessian. In the first experiment (E1) the full background covariance matrix  $\mathbf{B}$  was used together with covariance information for observations from EUR. In the second experiment (E2) a diagonal background-error covariance matrix was used with the inverse of the total-energy weights as variances and with observational statistics for observations from the northern hemisphere extratropics. The same observation sets are used as in section 3(a). Table 1 gives, for each experiment E1 and E2, the similarity index between the unstable subspace formed by the leading ten SVs thus obtained and the SVs in the case where the Hessian is approximated by 100 or 150 eigenvectors. Clearly, the leading singular vectors computed with an approximated

TABLE 1. SIMILARITY INDICES BETWEEN THE UNSTABLE SUBSPACES FOR TWO EXPERIMENTS, WITH THE HESSIAN APPROXIMATED USING 100 AND 150 EIGENPAIRS.

Number of eigenpairs used	100	150
Experiment E1	0.71	0.83
Experiment E2	0.63	0.74

See text for explanation.

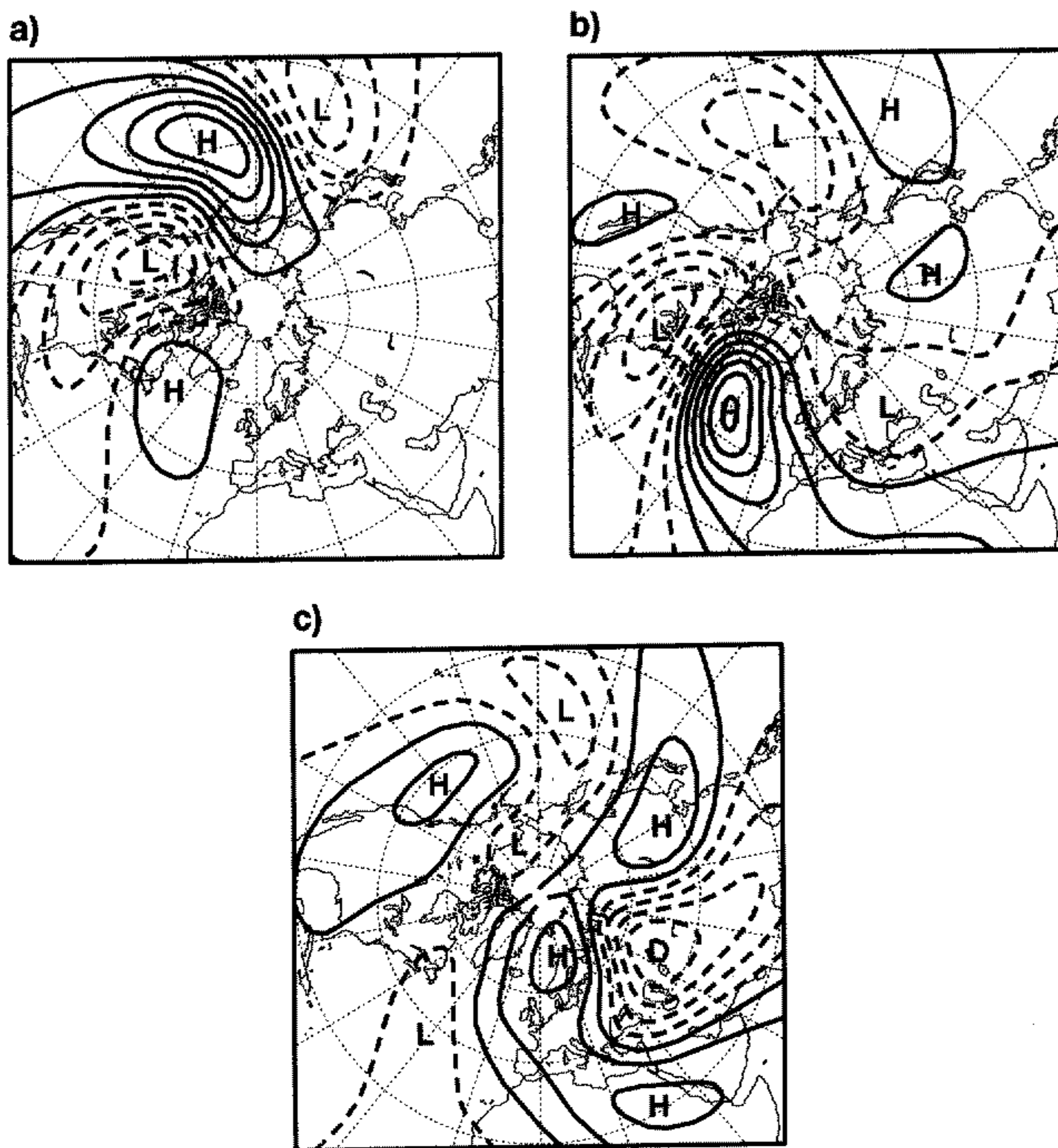


Figure 8. Stream-function fields at 500 hPa of the three leading Hessian singular vectors using 150 eigenpairs in the approximation of the Hessian and observations from EUR (20°W–20°E; 30°N–80°N). Solid (dashed) lines denote positive (negative) values and the contour interval is  $5 \times 10^{-8} \text{ m}^2 \text{ s}^{-1}$ .

Hessian, based on 150 eigenpairs, capture reasonably well the dominant part of the unstable subspace obtained with the full Hessian. Figure 8 shows the stream-function fields at 500 hPa of the first three singular vector for experiment E1. In particular, the first singular vector compares well with the corresponding HSV computed with the full Hessian (see Fig. 6(a)).

## 5. FINAL REMARKS

This paper considers the choice of the inner product to be used to define singular vectors for the ECMWF EPS. In the context of predictability experiments one would like to determine the leading eigenpairs of the forecast-error covariance matrix. The use of a metric which incorporates the second moment of the analysis-error probability distribution satisfies this requirement. A practical problem in using this metric (which we refer to as the

analysis-error covariance metric) to determine singular vectors is that it requires an efficient solution of a generalized eigenvalue problem. Because of the difficulty of solving such a system, the calculation of singular vectors up to now has been based on an energy metric, which may be interpreted as an approximation of the analysis-error statistics (Palmer *et al.* 1998). With this approximation, the singular vectors can be efficiently computed by applying the Lanczos algorithm to the propagators of the tangent and adjoint model.

The method described in this paper uses the full Hessian (or second derivative) of the cost function of the variational data assimilation as an approximation to the analysis-error covariance metric. In this way the calculation of singular vectors can be made consistent with the 3D/4D-Var calculation of the analysed state. The algorithm to solve the generalized eigenproblem is a generalization of the Davidson algorithm (Davidson 1975). It requires only that the propagators of the linear and adjoint model and the Hessian of the cost function are available in operator form, i.e.  $y = Sx$  can be computed, where  $S$  is any of these operators and  $x$  is an input vector.

We used the Hessian from different configurations of the 3D-Var cost function. When no observations are used in 3D-Var, the analysis-error metric is entirely based on the description of the first-guess error. Currently the first-guess error statistics are derived from the difference between the two-day and one-day forecast valid for the same day (the so-called NMC method, Parrish and Derber (1992)). The background covariance matrix defined in this way penalizes the occurrence of baroclinic structures in the analysis error and it lacks a realistic description of small-scale error structures. Experiments with a five-level PE model gave singular vectors which are of larger scale than the singular vectors defined in terms of the total-energy metric. Also, the vertical structure of the singular vectors revealed a difference with respect to the energy distribution. Most of the energy is confined to the upper levels when the Hessian was used, instead of peaking around the baroclinic steering level as in the case of total-energy singular vectors. Including covariance information of observations in the Hessian mainly changed the location of the maxima of the singular vectors. The amplitude of the singular vectors was reduced substantially over the areas where observations were present.

The impact of observational-error statistics on the singular vectors which are used in the ECMWF EPS was investigated by modifying the background-error covariance matrix. By setting this matrix to a diagonal matrix with the inverse of the total-energy weights on its diagonal, the total-energy singular vectors are retrieved when no observations are used. Using most of the conventional observations from the northern hemisphere extratropics leads to a significant change of the unstable subspace as spanned by the leading singular vectors. This may have consequences for the set-up of future observational experiments such as Fronts and Atlantic Storm Tracks Experiment. The areas to perform adaptive observations as indicated by the leading singular vectors may change when statistics of observational errors are included in the singular-vector computation.

Although the proposed algorithm is of the order of three times more expensive computationally than the Lanczos algorithm, there is scope to improve upon this. One way is to reduce the number of observations without essentially changing the 3D-Var analysis procedure. Another possibility would be to approximate the Hessian through its leading eigenpairs. By doing so, the Lanczos algorithm can again be used to determine singular vectors. Preliminary experiments showed that by using a considerable number (150) of eigenpairs in the approximation of the Hessian, the dominant part of the unstable subspace could be reasonably well retrieved. The additional computation time needed to determine 150 eigenpairs of the Hessian is around 10% of the time to solve the Lanczos problem.

The background covariance matrix is an important part of the singular-vector computation when the Hessian is employed. The quite different results for the total-energy

and full-background covariance matrix indicate that some effort must be put into correctly modelling the background error. It seems that neither the 3D-Var nor the total-energy background-error metric are completely satisfactory for the calculation of singular vectors: the implied correlation structure seems respectively, either too broad or too sharp. A first improvement would be to optimize the 3D-Var background-error term for areas which are pointed to by the singular vectors. One can also use the 4D-Var Hessian. A definite answer to the problem will only come when some flow-dependent error covariances are estimated using a simplified Kalman filter. These approaches are being tried at ECMWF.

#### ACKNOWLEDGEMENTS

We thank R. Buizza, T. Palmer and M. Fisher for the stimulating discussions we had during the preparation of this paper. The advice from H. A. van der Vorst and G. L. G. Sleijpen on solving the generalized eigenvalue problem was very helpful. Also thanks are due to A. Hollingsworth and A. Simmons for carefully reading the manuscript. Finally the comments given by two reviewers were much appreciated.

#### APPENDIX

##### The generalized eigenproblem

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{B}\mathbf{x} \quad (\text{A.1})$$

is usually handled by bringing it back to a standard eigenproblem

$$\mathbf{B}^{-1}\mathbf{A}\mathbf{x} = \lambda\mathbf{x}. \quad (\text{A.2})$$

The matrix  $\mathbf{B}^{-1}\mathbf{A}$  is in general non-symmetric, even if both  $\mathbf{A}$  and  $\mathbf{B}$  are symmetric. However, if  $\mathbf{B}$  is symmetric and positive definite, the  $\mathbf{B}$  inner product is well defined. The matrix  $\mathbf{B}^{-1}\mathbf{A}$  is symmetric in this inner product if  $\mathbf{A}$  is symmetric:

$$(\mathbf{w}, \mathbf{B}^{-1}\mathbf{A}\mathbf{v})_{\mathbf{B}} = (\mathbf{B}\mathbf{w}, \mathbf{B}^{-1}\mathbf{A}\mathbf{v}) = (\mathbf{w}, \mathbf{A}\mathbf{v}) = (\mathbf{A}\mathbf{w}, \mathbf{v}) = (\mathbf{B}^{-1}\mathbf{A}\mathbf{w}, \mathbf{v})_{\mathbf{B}}, \quad (\text{A.3})$$

where  $\mathbf{w}$  and  $\mathbf{v}$  are arbitrary vectors. The proposed method to solve (A.1) constructs a set of basis vectors  $\mathbf{V}$  of a search space  $\mathcal{V}$ , cf. the Lanczos method (Lanczos 1950; Strang 1986). The approximate eigenvectors are linear combinations of the vectors  $\mathbf{V}$ . The classical and most natural choice for the search space  $\mathcal{V}$ , for instance utilized in the Lanczos method, is the so-called Krylov subspace, the space spanned by the vectors

$$\mathbf{v}, \mathbf{B}^{-1}\mathbf{A}\mathbf{v}, (\mathbf{B}^{-1}\mathbf{A})^2\mathbf{v}, \dots, (\mathbf{B}^{-1}\mathbf{A})^{i-1}\mathbf{v}. \quad (\text{A.4})$$

This  $i$ -dimensional subspace is denoted by  $\mathcal{K}^i(\mathbf{v}, \mathbf{B}^{-1}\mathbf{A})$ . The vector  $\mathbf{v}$  is a starting vector that has to be chosen. The Krylov subspace is well suited for computing dominant eigenpairs since the vector  $(\mathbf{B}^{-1}\mathbf{A})^{i-1}\mathbf{v}$  has a (much) larger component in the direction of the dominant eigenvector than  $\mathbf{v}$ .

In the next three subsections we will discuss how approximate eigenpairs can be computed from a given subspace, how a  $\mathbf{B}$ -orthogonal basis for the Krylov subspace can be constructed, and how operations with  $\mathbf{B}^{-1}$  can be avoided in the construction of the basis. This last subject is of particular importance in our application, since we have no explicit formulation of  $\mathbf{B}$  and we can only calculate  $\mathbf{y} = \mathbf{B}\mathbf{x}$  for an input vector  $\mathbf{x}$ .

*How to compute approximate eigenpairs*

Given a search space  $\mathcal{V}$  the approximate eigenpair  $(\theta, \mathbf{u})$  is a linear combination of the basis vectors of  $\mathcal{V}$ :

$$\mathbf{u} = \mathbf{V}\mathbf{y}. \quad (\text{A.5})$$

A suitable criterion for finding an optimal pair  $(\theta, \mathbf{u})$  is the Galerkin condition that the residual

$$\mathbf{r} = \mathbf{B}^{-1}\mathbf{A}\mathbf{u} - \theta\mathbf{u} = \mathbf{B}^{-1}\mathbf{A}\mathbf{V}\mathbf{y} - \theta\mathbf{V}\mathbf{y} \quad (\text{A.6})$$

is  $\mathbf{B}$ -orthogonal to the search space  $\mathcal{V}$ . Hence

$$\mathbf{V}^T\mathbf{B}\mathbf{r} = 0 \quad (\text{A.7})$$

and consequently, using (A.6),

$$\mathbf{V}^T\mathbf{A}\mathbf{V}\mathbf{y} - \theta\mathbf{V}^T\mathbf{B}\mathbf{V}\mathbf{y} = 0. \quad (\text{A.8})$$

Note that the resulting eigenproblem is of the dimension of the search space, which is generally much smaller than that of the original problem. The basis vectors are usually orthogonalized so that  $\mathbf{V}^T\mathbf{B}\mathbf{V} = \mathbf{I}$ . Approximate eigenpairs that adhere to the Galerkin condition are called Ritz pairs.

*Construction of a basis for the Krylov subspace*

If approximate eigenpairs are computed according to (A.8), then the residuals  $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_i$  form a  $\mathbf{B}$ -orthogonal basis for  $\mathcal{K}^i(\mathbf{r}_1, \mathbf{B}^{-1}\mathbf{A})$ . The  $\mathbf{B}$ -orthogonality

$$\mathbf{r}_i^T\mathbf{B}\mathbf{r}_j = 0 \quad (\text{A.9})$$

of the residuals  $\mathbf{r}_i$  follows immediately from their construction. According to (A.8) it is true for  $i < j$ . Since  $\mathbf{B}$  is symmetric the inner product is symmetric in  $i$  and  $j$ , and consequently it also holds for  $i > j$ .

It remains to show that the residuals form a basis for this Krylov subspace. Given a  $(i - 1)$ -dimensional search space, the Galerkin condition (A.8) yields  $i - 1$  approximate eigenpairs and residuals. The search space can be extended with any of these residuals. Each selected residual  $\mathbf{r}_i$  will be a new  $\mathbf{B}$ -orthogonal basis vector for  $\mathcal{K}^i(\mathbf{v}, \mathbf{B}^{-1}\mathbf{A})$ :

$$\text{span}\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_i\} = \mathcal{K}^i(\mathbf{r}_1, \mathbf{B}^{-1}\mathbf{A}). \quad (\text{A.10})$$

This can be seen by noting that the residual is given by (see (A.6))

$$\mathbf{r}_i = \mathbf{B}^{-1}\mathbf{A}\mathbf{u}_{i-1} - \theta_{i-1}\mathbf{u}_{i-1} = \mathbf{B}^{-1}\mathbf{A}\mathbf{V}_{i-1}\mathbf{y}_{i-1} - \theta_{i-1}\mathbf{V}_{i-1}\mathbf{y}_{i-1}, \quad (\text{A.11})$$

with  $(\theta_{i-1}, \mathbf{u}_{i-1})$  an eigenpair approximation in the space  $\mathcal{K}^{i-1}(\mathbf{r}_1, \mathbf{B}^{-1}\mathbf{A})$ . Since

$$\mathbf{B}^{-1}\mathbf{A}\mathbf{V}_{i-1}\mathbf{y}_{i-1} \in \mathcal{K}^i(\mathbf{r}_1, \mathbf{B}^{-1}\mathbf{A}), \quad (\text{A.12})$$

the residual  $\mathbf{r}_i$  also belongs to  $\mathcal{K}^i(\mathbf{r}_1, \mathbf{B}^{-1}\mathbf{A})$ . Because  $\mathbf{r}_i$  cannot be expressed as a linear combination of  $\mathbf{r}_k$  with  $k \leq i - 1$  we conclude that  $\mathbf{r}_i$  is a new basis vector of  $\mathcal{K}^i(\mathbf{r}_1, \mathbf{B}^{-1}\mathbf{A})$ . Note that the residual is zero only if there is an eigensolution.



*Solution method*

As was stated before, the natural search space for the generalized eigenvalue problem is the Krylov subspace  $\mathcal{K}^i(\mathbf{v}, \mathbf{B}^{-1}\mathbf{A})$ . This basis can be generated by expanding the basis by new residual vectors. The problem in the construction of a basis for this space is that operations with  $\mathbf{B}^{-1}$  are needed. Since in our application the inverse of  $\mathbf{B}$  is not known explicitly, we approximated its action by an iterative solution method, like the Conjugate Gradient method (CG) (Hestenes and Stiefel 1954) or the Conjugate Residual method (CR) (Cline 1976). To compute the vector

$$\mathbf{r} = \mathbf{B}^{-1}\tilde{\mathbf{r}}, \quad (\text{A.13})$$

one iteratively solves the system

$$\mathbf{B}\mathbf{r} = \tilde{\mathbf{r}}. \quad (\text{A.14})$$

Iterative solution methods require, apart from vector operations, only multiplications with  $\mathbf{B}$ . The vector  $\mathbf{r}$  can in principle be determined to high accuracy. This, however, may require many multiplications with  $\mathbf{B}$  and hence may be very expensive. We therefore propose to approximate the action of  $\mathbf{B}$  to low accuracy only, by performing only a few steps with an iterative solution method. The subspace generated in this way is not a Krylov subspace and the basis vectors are not the residuals (A.6) but only approximations to it. As a consequence they are not perfectly  $\mathbf{B}$ -orthogonal. Orthogonality has to be enforced explicitly.

The algorithm we use can be summarized as follows:

- Select number of CR steps, maximum dimension of  $\mathbf{V}$ , starting vector  $\mathbf{v}$ .
- Compute  $\mathbf{B}\mathbf{v}$ ,  $\mathbf{B}$ -normalize  $\mathbf{v}$ .
- Do  $i = 1$  until maximum dimension of  $\mathbf{V}$ :
  1. compute  $\mathbf{A}\mathbf{V}$ ,  $\mathbf{V}^T\mathbf{A}\mathbf{V}$ ,
  2. solve small eigenproblem  $\mathbf{V}^T\mathbf{A}\mathbf{V}\mathbf{y} = \theta\mathbf{y}$ ,
  3. select  $\mathbf{y}$ , Ritz value  $\theta$ ,
  4. compute Ritz vector  $\mathbf{u} = \mathbf{V}\mathbf{y}$  and residual  $\tilde{\mathbf{r}} = \mathbf{A}\mathbf{V}\mathbf{y} - \theta\mathbf{B}\mathbf{V}\mathbf{y}$ ,
  5. compute approximately  $\mathbf{v} = \mathbf{B}^{-1}\tilde{\mathbf{r}}$  with the CR method,
  6.  $\mathbf{B}$ -orthonormalize new  $\mathbf{v}$  against  $\mathbf{V}$ ,
  7. expand  $\mathbf{V}$  with the resulting vector.

The matrix  $\mathbf{V}$  contains the basis for the search space, the vector  $\mathbf{v}$  contains the new basis vector, and  $(\mathbf{u}, \theta)$  is an approximate eigenpair. In step 3 the pair with the largest  $\theta$  is selected since we are interested in the upper part of the spectrum. However, if the eigenpair approximation reaches a certain accuracy  $\varepsilon$ , i.e.  $\tilde{\mathbf{r}} = \mathbf{A}\mathbf{u} - \theta\mathbf{B}\mathbf{u} \leq \varepsilon$ , a smaller  $\theta$  is selected. In step 5 a few CR steps are performed to approximate the action of  $\mathbf{B}^{-1}$ . Another iterative solution, e.g. CG, could also be used. The optimal number of CR iterations depends mainly on the cost of operations with  $\mathbf{B}$ . If these are expensive the number of CR iterations should be kept small. In our experiments the number of CR steps per iteration is 15. Evaluation of  $\mathbf{A}$  and  $\mathbf{B}$  is then computationally equally expensive. In step 6 the modified Gram–Schmidt procedure is used for reasons of numerical stability.

*Relation with standard methods*

The above method is closely related to Davidson's method (Davidson 1975) for solving the standard eigenproblem. The difference is that we try to approximate  $\mathbf{B}^{-1}$  by means of an iterative solution method, whereas in Davidson's method the action of  $(\mathbf{A} - \theta\mathbf{I})^{-1}$  is approximated. The purpose of this is to speed up convergence. The way to construct the basis for the search space is the same, however. Both methods compute residuals based

on Galerkin approximations for the eigenpairs. Recently, an improvement over Davidson's method has been proposed by Sleijpen and Van der Vorst (1996), the so called Jacobi–Davidson method. The essential difference between the Davidson method and the Jacobi–Davidson method is that in the latter the action of  $(\mathbf{A} - \theta\mathbf{I})^{-1}$  is approximated in the space orthogonal to the Ritz vector  $\mathbf{u}$ . Sleijpen *et al.* (1996) give a framework for Davidson and Jacobi–Davidson type methods to solve generalized and polynomial eigenproblems. The method described above can be seen as a generalized Davidson method. If the action of  $\mathbf{B}^{-1}$  is approximated to machine accuracy, or if  $\mathbf{B} = \mathbf{I}$ , the eigenpair approximations should be the same as the ones obtained with the Lanczos method.

## REFERENCES

- Bouttier, F. 1994 A dynamical estimation of error covariances in an assimilation system. *Mon. Weather Rev.*, **122**, 2376–2390
- Buizza, R. 1994 Sensitivity of optimal unstable structures. *Q. J. R. Meteorol. Soc.*, **120**, 429–451
- 1995 Optimal perturbation time evolution and sensitivity of ensemble prediction to perturbation amplitude. *Q. J. R. Meteorol. Soc.*, **121**, 1705–1738
- Buizza, R. and Palmer, T. N. 1995 The singular vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52**, 1434–1456
- Cline, A. K. 1976 Several observations on the use of conjugate gradient methods. ICASE Report 76, NASA Langley Research Center, Hampton, Virginia, USA
- Courtier, P., Freyder, C., Geleyn, J. F., Rabier, F. and Rochas, M. 1991 'The Arpege project at Météo France'. Proceedings of ECMWF seminar on numerical methods in atmospheric models, 9–13 September 1991, Vol. 2. ECMWF, Reading, UK
- 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). Part I: Formulation. *Q. J. R. Meteorol. Soc.*, (in press)
- Davidson, E. R. 1975 The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real symmetric matrices. *J. Comp. Phys.*, **17**, 87–94
- Ehrendorfer, M. and Tribbia, J. J. 1997 Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.*, **54**, 286–313
- Epstein, E. S. 1969 Stochastic dynamic predictions. *Tellus*, **21**, 739–759
- Fisher, M. and Courtier, P. 1995 Estimating the covariance matrices of analysis and forecast error in variational data assimilation. ECMWF, Technical Memorandum No. 220
- Hestenes, M. R. and Stiefel, E. 1954 Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Stand.*, **49**, 409–436
- Houtekamer, P. L. 1995 The construction of optimal modes. *Mon. Weather Rev.*, **122**, 2179–2191
- Houtekamer, P. L., Lefaiivre, L., Derome, J., Ritchie, H. and Mitchell, H. L. 1996 A system approach to ensemble prediction. *Mon. Weather Rev.*, **124**, 1225–1242
- Lanczos, C. 1950 An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bur. Stand.*, **45**, 255–282
- Leith, C. E. 1974 Theoretical skill of Monte Carlo forecasts. *Mon. Weather Rev.*, **102**, 409–418
- Lorenç, A. C. 1986 Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177–1194
- Lorenz, E. N. 1965 A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17**, 321–333
- Mardia, K. V., Kent, J. T. and Bibby, J. M. 1979 *Multivariate analysis*. Academic Press, London
- Molteni, F., Buizza, R., Palmer, T. N. and Petroliagis, T. 1996 The ECMWF ensemble prediction system: methodology and validation. *Q. J. R. Meteorol. Soc.*, **122**, 73–119

- Palmer, T. N., Gelaro, R.,  
Barkmeijer, J. and Buizza, R.  
Parrish, D. I. and Derber, J. C. 1998 Singular vectors, metrics and adaptive observations. *J. Atmos. Sci.*, **54**, 633–653
- 1992 The National Meteorological Centre's spectral statistical interpolation analysis system. *Mon. Weather Rev.*, **120**, 1740–1763
- 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., Klinker, E., Courtier, P.  
and Hollingsworth, A. 1996 Sensitivity of forecast errors to initial conditions. *Q. J. R. Meteorol. Soc.*, **122**, 121–150
- Rabier, F., McNally, A.,  
Anderson, E., Courtier, P.,  
Undén, P., Eyre, J.,  
Hollingsworth, A. and  
Bouttier, F. 1998 The ECMWF implementation of three-dimensional variational assimilation (3D-Var). Part II: Structure functions. *Q. J. R. Meteorol. Soc.*, (in press)
- Simmons, A. J., Burridge, D. M.,  
Jarraud, M., Girard, M. and  
Wergen, W. 1989 The ECMWF medium-range prediction models development of the numerical formulations and the impact of increased resolution. *Meteorol. Atmos. Phys.*, **40**, 28–60
- Sleijpen, G. L. G. and van der  
Vorst, H. A. 1996 A Jacobi–Davidson iteration method for linear eigenvalue problems. *SIAM J. Matrix Anal. Appl.*, **17**, 401–425
- Sleijpen, G. L. G., Booten, J. G. L.,  
Fokkema, D. R. and van der  
Vorst, H. A. 1996 Jacobi–Davidson type methods for generalized eigenproblems and polynomial eigenproblems. *BIT*, **36**, 595–633
- Stephenson, D. B. 1997 Correlation of spatial climate/weather maps and the advantages of using the Mahalanobis metric in predictions. *Tellus*, **49A**, 513–527
- Strang, G. 1986 *Introduction to applied mathematics*. Wellesley–Cambridge Press
- Thépaut, J.-N., Courtier, P.,  
Belaud, G. and Lemaître, G. 1996 Dynamical structure functions in a four-dimensional variational assimilation: A case-study. *Q. J. R. Meteorol. Soc.*, **122**, 535–561
- Toth, Z. and Kalnay, E. 1997 Ensemble forecasting at NCEP and the breeding method. *Mon. Weather Rev.*, **125**, 3297–3319
- Van den Dool, H. M. and  
Rukhovets, L. 1994 On the weights for an ensemble-averaged 6–10-day forecast. *Weather Forecasting*, **9**, 457–565
- Wait, R. 1979 *The numerical solution of algebraic equations*. John Wiley & Sons, Chichester, UK