TECHNICAL MEMORANDUM

445

The structure and realism of sensitivity perturbations and their interpretation as 'Key Analysis Errors'

L. Isaksen, M. Fisher and E. Andersson

Research Department

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

Abstract

Adjoint sensitivity-based perturbations (e.g. 'Key Analysis Errors') minimizing the two-day forecast error have been calculated for December 2001 and January 2002 using three different initial-time norms. The goal has been to investigate if the perturbations can justifiably be interpreted as analysis error. We perform a systematic comparison against available observations, at initial-time and as the perturbations evolve during the first 24 hours of forecasts. Ten-day forecasts have been run to verify that the medium-range forecasts from the perturbed analyses are better than the control forecasts. It is shown that the structure of the perturbations depends strongly on the initial-time norm, in experiments based on the energy norm, the background error covariance norm, and the Hessian norm. For all three norms it is found that forecasts starting from the perturbed analyses are further away from the observations than forecasts from control analyses during the first approximately 12 hours of forecasts. From 12 hours onwards the perturbed forecasts are closer to observations than the control forecasts. We conclude that the sensitivity perturbations cannot justifiably be interpreted as analysis error as far as their detailed structure is concerned. This result has implications in applications that rely on sensitivity-based or Singular Vector-based approaches for detailed characterization of analysis error, e.g. some observation targeting applications and the reduced-rank Kalman filter.

1 Introduction

Through detailed comparison with available observations we investigate whether adjoint sensitivity-based perturbations, which by construction reduce the two-day forecast error, can be interpreted as initial-condition error. The skill of numerical weather forecasts strongly depends on the accuracy of the initial conditions (Simmons and Hollingsworth 2002). The growth or decay of initial-condition errors depends on the atmospheric flow situation. Several studies have demonstrated that small and localized initial errors can grow very quickly in the first few days of forecasts, and ruin the prediction of important weather events. With good knowledge of the initial condition error ('analysis error', hereafter), or of its statistical characteristics, it is possible to predict the forecast error and make statements about forecast reliability at the time the forecast is made. Knowledge of analysis error is also a pre-requisite for observation targeting (Emanuel *et al.* 1995; Palmer *et al.* 1998), and for further development of data assimilation methods.

Characterization of analysis error has been an area of intense research in recent years, facilitated by the development of Singular Vector (SV) techniques (Molteni and Palmer 1993; Buizza and Palmer 1995), adjoint sensitivity (AS) diagnostics (Rabier *et al.* 1996), and ensemble data assimilation (EnDA) (e.g. Houtekamer and Mitchell 1998; 2001). Both SV and AS-based methods have been widely used to compute the fastest-growing components of analysis error. SVs are used operationally at the European Centre for Medium-Range Weather Forecasts (ECMWF) to produce initial perturbations for the Ensemble Prediction System (EPS) (Molteni *et al.* 1996).

Hansen and Smith (2000) investigated the validity of AS and SV approaches for observation targeting in the Lorenz (1995) 40-dimensional model. They found that linear-dynamics methods (such as AS and SV) work well compared to other targeting methods (such as EnDA) provided 1) the tangent-linear hypothesis is valid; 2) the model error is small over the optimisation period; 3) the initial-time linearisation state is accurately known, i.e. the analysis error is small. If these conditions are met and if the initial-time norm provides an accurate representation of the analysis error covariance matrix, then initial-time SVs will evolve into the leading eigenvectors of the forecast error covariance matrix (Palmer *et al.* 1998).

In data assimilation, there is ongoing research on how to incorporate knowledge about the flow-dependent nature of analysis and short-range forecast error into the analysis procedure, so as to give the appropriate higher weight to valuable data in the sensitive regions. The approach at ECMWF (first outlined by Courtier 1993) has been to explicitly incorporate, within the background error covariance matrix of 4D-Var (Rabier *et al.* 2000),

the leading SVs of the tangent linear model. The method can be described as a reduced-rank approximation of the Kalman filter (the RRKF) (Fisher 1998; Beck and Ehrendorfer 2004). After several years of development and experimentation with the RRKF system the results were positive, but not positive enough to warrant the extra computational expense in an operational context (Fisher and Andersson 2001). It was clear that the RRKF provided strongly flow-dependent background-error correlation structures within the sensitive regions, but it was less clear whether the observations had a systematic projection onto those structures. Small RRKF analysis impact did appear in the sensitive regions, but the resulting forecast impact was negative nearly as often as it was positive. RRKF experiments were carried out with SVs computed using different initial norms (total energy, the approximate analysis Hessian, the background error covariance), and with adjoint sensitivity structures - with essentially similar near-neutral results (Fisher and Andersson 2001). The questions were raised: How well does a low-dimensional subspace of SVs and sensitivity gradients represent the true structure of analysis and short-range forecast error? Are the structures obtained through these methods observable with the current observing network?

In the current study we carry out a systematic investigation over two one-month periods comparing the sensitivity perturbations against observations available in ECMWF's data assimilation system, at initial time and as the perturbations evolve over the first 24 hours of forecasts. We investigate the so-called 'Key Analysis Errors' obtained from adjoint sensitivity gradients of 48 hour forecast errors with respect to the initial conditions (Klinker *et al.* 1998). The structures are strongly dependent on the choice of initial norm; we compare results for total energy, the approximate analysis Hessian and the background error covariance norms. Gelaro *et al.* (1998) showed that the 'Key Analysis Errors' project strongly on SVs, to the extent that a linear combination of the leading 30 SVs describes a large fraction of their variance. Klinker *et al.* (1998) demonstrated that adding the 'Key Analysis Errors' to the unperturbed control analysis improves not only the 48-hour forecast (which it must do by construction) but that the improvement is retained into the medium-range and beyond. The improvement of the forecast was in many cases spectacular, which fostered the perception that the perturbed analysis was indeed a corrected analysis.

In this paper we test if the sensitivity perturbations correct the analysis, and thus whether or not they represent a component of the true analysis error. Central to our work is the assumption that analyses are improved in the sensitive regions if analyses and very short-range forecasts are closer to the observations than the control analysis. This is a reasonable assumption as we are comparing with a control analysis from an assimilation system that has a lack of flow-dependence in the background error covariance. It is expected that the lack of flow dependence leads to background errors that tend to be too small in the sensitive regions, and a corrected analysis should tend to be closer to the observations. The opposite situation - that a better analyses should be further from the observations, would imply that performance improvement could be obtained by a general decrease of background error variances. This has been tested at ECMWF and found not to be the case. In Klinker *et al.* (1998) the realism of the adjoint sensitivity-based structures was investigated in this manner for one case study over North America. The method has also been applied by Langland *et al.* (2002) with a more localized area to investigate an intense synoptic event. Our study covered a two-month period, so a large number of observations were available for a statistically significant evaluation over a range of atmospheric flow regimes.

In Section 2 we review the concepts of sensitivity gradients and 'Key Analysis Errors'. In Section 3 the design of the assimilation experiment and the method for validating sensitivity perturbations against observations are described. Section 4 investigates the impact of the three initial-time norms on the general structure of sensitivity perturbations. In Section 5 the sensitivity perturbations are compared with observations from a number of observing systems. Discussion and conclusions are presented in Section 6.

2 Methodology

2.1 Sensitivity of forecast errors to initial condition

Rabier *et al.* (1996) investigated the sensitivity of 48 hour forecast errors to the initial conditions using a tangent linear and adjoint technique. The aim was to minimize a diagnostic function *J* that provides a scalar measure of forecast errors based on the difference between the two-day forecast and the verifying analysis (available two days after the forecast was made).

The diagnostic function chosen to be minimised is

$$J_t = \frac{1}{2}(\mathbf{x}_t^f - \mathbf{x}_t^a)^T \mathbf{P}^T \mathbf{A} \mathbf{P}(\mathbf{x}_t^f - \mathbf{x}_t^a) = \frac{1}{2}\|\mathbf{P}(\mathbf{x}_t^f - \mathbf{x}_t^a)\|_a^2, \tag{1}$$

where $\mathbf{A}$ is the matrix defining a final-time norm ($\|\mathbf{x}\|_a^2 = \mathbf{x}^T \mathbf{A} \mathbf{x}$), and where $\mathbf{P}$ represents a projection on latitudes north of $30°$N. The time t is the final time, i.e. 48 hours. Like Rabier *et al.* (1996) we use as final-time norm the dry energy norm (defined in subsection (c)).

The forecast $\mathbf{x}_t^f$ is given by a 48 hour non-linear model integration, i.e.

$$\mathbf{x}_t^f = \mathbf{M}_t(\mathbf{x}_0) \tag{2}$$

and $\mathbf{x}_t^a$ is the verifying analysis valid at time t.

Rabier *et al.* (1996) showed that the initial-time gradient with respect to the inner product defined by the matrix $\mathbf{A}$ can be computed from the final-time gradient $\nabla_A J_t = \mathbf{P}(\mathbf{x}_t^f - \mathbf{x}_t^a)$, through the use of the adjoint $\mathbf{R}^* = \mathbf{A}^{-1}\mathbf{R}^T\mathbf{A}$ of the resolvent of the tangent linear model integration, $\mathbf{R}$, under the assumptions that the forecast model is perfect (i.e. no model error) and that the tangent-linear approximation is valid:

$$\nabla_A J_0 = \mathbf{R}^* \nabla_A J_t. \tag{3}$$

The gradient of the objective function with respect to an alternative initial-time inner product, defined by a matrix $\mathbf{C}$ is given by

$$\nabla_C J_0 = \mathbf{C}^{-1}\mathbf{A}\nabla_A J_0. \tag{4}$$

For infinitesimal perturbations of given 'size', as measured by the $\mathbf{C}$-norm, $\|\delta\mathbf{x}_0\|_c = \varepsilon$, the gradient may be interpreted as the direction that produces the largest change in the objective function.

The experiments performed by Rabier *et al.* (1996) showed that small sensitivity-determined adjustments of the initial fields could result in substantial improvement of the medium-range forecasts. They linked the sensitivity perturbations to analysis error, and showed the link for one case-study over North America. They recommended a more comprehensive objective verification against observations to diagnose cases of analysis shortcomings. One of the aims of this paper is to investigate that issue in detail.

Fig. 1(b) shows an example of the 48 hour temperature forecast error at approximately 650 hPa. Note the large forecast error (up to 12 K) east of Japan. Fig. 1(a) shows the energy norm sensitivity gradient two days earlier. It is seen that according to these sensitivity calculations the large forecast errors east of Japan can be corrected by modifying the analysed temperature field over Japan two days earlier. This specific case will be discussed further in section 4. An error in the middle of the North Atlantic can similarly be corrected by modifying the temperature field near the east coast of U.S.A.

2.2 Definition of 'Key Analysis Errors'

The adjoint sensitivity method described above identifies the optimal direction for perturbations of infinitesimal size. For finite perturbations, the gradient is not optimal. It is nevertheless possible to define a step size, α, that maximizes the change in objective function for perturbations in the direction of the gradient. Rabier *et al.* (1996) determined the optimal step size on a trial and error basis, though linked to the squared value of the fastest growing Singular Vector over 48 hours. Klinker *et al.* (1998) extended the sensitivity method by developing a way to determine a direction, and corresponding step size, that allowed larger improvements of the objective function than could be achieved by adding perturbations proportional to the sensitivity gradient.

Subject to the validity of the tangent linear and perfect-model approximations, the diagnostic function (1) is quadratic. For perturbations $\delta\mathbf{x}_0 = \alpha\mathbf{d}$ in a given direction $\mathbf{d}$, the optimal step size may be determined from (1), and is given by

$$\alpha_{opt} = \frac{\mathbf{d}^T\mathbf{R}^T\mathbf{P}^T\mathbf{AP}(\mathbf{M}(\mathbf{x}_0) - \mathbf{x}_t^a)}{\mathbf{d}^T\mathbf{R}^T\mathbf{P}^T\mathbf{APRd}} \tag{5}$$

The optimal step (5) locates the minimum of the objective function along the line $\mathbf{x}_0 + \alpha\mathbf{d}$. Further reduction of the objective function is possible by taking a second step in a new direction, which may be determined by calculating the gradient of the objective function at $\mathbf{x}_0 + \alpha_{opt}\mathbf{d}$. The process may be iterated a number of times in a manner reminiscent of a 4D-Var minimization. The choice of how many iterations to perform requires a balance to be struck between reduction of the objective function, and the realism of the perturbations. Klinker *et al.* (1998) determined the optimal number of iterations by comparing the obtained perturbations against observations. They stated that this 'optimal' perturbation represents the 'Key Analysis Error'. In their case study, three energy-norm-based iterations, using the quasi-Newton minimization algorithm M1QN3 (Gilbert and Lémarchal 1989), moved the analysis significantly closer to observations than the control analysis. In Klinker *et al.* (1998) it is discussed why more iterations do not necessarily provide a better description of analysis error. The main reason is that increasing the number of iterations amplifies noise in the initial perturbation due to contracting directions of the tangent-linear model. Reynolds and Palmer (1998) investigated this problem in more detail. A similar result was obtained by Thépaut and Courtier (1991) where they performed 4D-Var assimilation without a background constraint.

Klinker *et al.* (1998) also showed that enstrophy-norm sensitivity perturbations minimize the two day forecast error but have a very different structure than energy-norm perturbations. The enstrophy-norm sensitivity perturbations fit the observations significantly worse than the energy norm perturbations. In this paper we have further investigated this dependence of the chosen norm on the structure of sensitivity perturbations. It is seen from (4) and (5) that the initial-time norm plays an important role in determining both the structure of the perturbations and their optimal step-size.

Fig. 1(c) shows the temperature perturbation after one iteration, based on the above method, for the case presented in Fig. 1(a) and (b). It is seen that Fig. 1(a) and Fig. 1(c) have the same structure but different signs and amplitudes. Fig. 1(a) gives the gradient whereas Fig. 1(c) gives the perturbation size returned by the minimization algorithm. This step is not optimal, but is guaranteed by the minimization algorithm to result in a reduction of the objective function. (The step-size depends in a somewhat complicated way on the line-search algorithm of the minimization algorithm and on an initial guess at the optimal step.) Fig. 1(d) shows the perturbation after three iterations, the so-called 'Key Analysis Errors'. The additional iterations maintain the overall structure of the sensitivity gradient but also add finer scale structures and new larger scale structures, i.e. over South-East Europe. The large (up to 12 K) errors east of Japan are corrected by modifying the temperature field over Japan by 1.2 K, i.e. a ten-fold error growth in 48 hours.

2.3 Calculation of 'Key Analysis Errors' for various initial norms

As discussed in the previous section, the 'Key Analysis Errors' are initial time perturbations calculated by applying a few steps of minimization to an objective cost function that measures the difference between a two-day forecast and a verifying analysis. In this study we have investigated three initial-time norms. They will be described here.

The dry energy norm is defined as:

$$\|\mathbf{x}\|_c^2 = 1/2 \int_\eta \int_D (u^2 + v^2 + R_d T_r (\ln P_s)^2 + (C_p/T_r) T^2) dD (\delta P_r / \delta \eta) d\eta \tag{6}$$

where u, v, $\ln P_s$, and T are the wind, logarithm of surface pressure and temperature components of $\mathbf{x}$, respectively. D represents the horizontal domain and η represents the vertical domain (the whole model atmosphere). The components are scaled according to energy weighting, which involves a reference temperature, T_r, and reference surface pressure, P_r. C_p is specific heat at constant pressure for dry air and R_d is the gas constant for dry air. The energy norm is a simple operator that is easy to implement. This norm has been extensively used for many applications, i.e. singular vector calculation and for diagnostics (Palmer *et al.* 1998)

The Hessian ($\mathbf{H}$) norm ought to be the best choice of norm. The Hessian (i.e. the matrix of second derivatives) of the variational cost function yields an estimate of the inverse of the analysis error covariance matrix (Rabier and Courtier 1992), so that perturbations in the direction of the gradient maximize the change in the objective function for all perturbations on a surface of equal probability. However, the $\mathbf{H}^{-1}$ represents the true analysis error only on the assumption that the statistical information provided to the assimilation scheme is accurate. In particular it is assumed that the observation and background error covariances ($\mathbf{R}$ and $\mathbf{B}$) are correctly specified. The Hessian norm is more complex to apply and requires routines to multiply vectors by the symmetric square root of the matrix that defines the norm, and by the symmetric square-root of its inverse (Barkmeijer *et al.* 1999). In practice, this necessitates the use of an approximate Hessian. We use the same approximation as is used to precondition the minimization (see appendix B of Fisher and Andersson 2001). That is, the 4D-Var Hessian norm ($\mathbf{C} = \mathbf{H}$) is approximated by:

$$\tilde{\mathbf{H}} = \mathbf{B}^{-1/2} \tilde{\mathbf{Y}} \mathbf{B}^{-1/2} \tag{7}$$

where:

$$\tilde{\mathbf{Y}} = \mathbf{I} + \sum_{k=1}^{K} (\lambda_k - 1) \mathbf{v}_k \mathbf{v}_k^T \tag{8}$$

Here, λ_k and $\mathbf{v}_k$ are the leading eigenvalues and eigenvectors of the Hessian in the space of the control vector of the 4D-Var minimization. This approximate 4D-Var Hessian ($\tilde{\mathbf{H}}$) is easy to invert and to square root. In this study, we use K=100. A more detailed description of the $\tilde{\mathbf{H}}$ can be found in Barkmeijer *et al.* (1998; 1999). In order to study the influence of observations on the $\tilde{\mathbf{H}}$-norm 'Key Analysis Errors' have also been calculated using the background-error covariance norm. That is:

$$\mathbf{C} = \mathbf{B}^{-1} \tag{9}$$

This $\mathbf{B}$-norm has no influence from observations. Similarly, the effect of truncating the sum in (8) is to make the 'Key Analysis Errors' calculated using the $\tilde{\mathbf{H}}$-norm more similar to the $\mathbf{B}$-norm, especially in data sparse regions (Fisher and Andersson 2001). This is because the leading eigenvectors of the Hessian tend to be concentrated in regions of dense observations.

3 Design of assimilation experiments and method for validation against observations

ECMWF's 4D-Var assimilation system (Rabier *et al.* 2000) version 25r1, operational from April 2002 to January 2003, was used for the experimentation. The incremental version (Courtier *et al.* 1994) performed both outer-loop and inner-loop calculations at $T159L60$ (spectral triangular truncation at wavenumber 159, linear Gaussian grid and 60 vertical levels). Forecast error calculations were performed at $T42L60$ (quadratic Gaussian grid). The setup is similar to ECMWF's operational assimilation system in 2002, except that in the operational system the outer-loop calculations were performed at $T511L60$.

4D-Var assimilation experiments were performed for the period 1-31 December 2001 and 1-31 January 2002. For both periods, daily runs were performed to calculate 'Key Analysis Errors' at T63L60 resolution from the 48-hour forecast error using the method described in section 2(b) with the initial-time norms described in section 2(c). The energy norm calculations used 1200 UTC analyses as the starting point, whereas the **B**-norm and $\tilde{\mathbf{H}}$-norm calculations used three hour forecasts valid at 0300 UTC as the starting point, (the latter because the $\tilde{\mathbf{H}}$-norm calculations were influenced by observations during the 4D-Var assimilation window from 0300 UTC to 1500 UTC (Bouttier 2001). It should be noted that the initial state only depends on observations before 0300 UTC so observations after 0300 UTC can be used as an independent data source for evaluating the quality of analyses and forecasts, as will discussed below. The realism of the sensitivity perturbations can be evaluated without problems despite the difference in starting point because control forecasts were run starting from 0300 UTC and 1200 UTC, respectively.

Forecast verification scores (Fig. 2) confirm that modifying the initial conditions using any of the three 'Key Analysis Error' perturbations, corresponding to the norms described above, improve the 2-7 day forecasts to a similar degree. This is despite the fact that the initial perturbations are very different, as will be shown in section 4. The medium-range forecast improvement is sustained because the objective function reduces the two day error substantially and therefore delays the error growth on subsequent forecast days. This delayed error growth could be obtained in many different ways depending on the norm used.

A methodology similar to the one used in Klinker *et al.* (1998) was applied to investigate the realism of the various sensitivity perturbations. As in Klinker *et al.* (1998) we judge the quality of the analysis by its ability to fit quality observations in sensitivity perturbation areas. In addition, 24-hour forecasts were performed from the control analyses and from analyses modified by the three types of 'Key Analysis Error' perturbations. The forecasts were compared to observations using the operational 4D-Var observation operators. The model-observation comparison is performed at proper time (within 30 minutes), avoiding interpolation of model fields in the time domain. Forecasts were run every day for the two-month assimilation experiments, making it possible to make statistical evaluations based on large numbers of observations, covering many different synoptic situations.

4 The impact of initial-time norm on the general structure of sensitivity perturbations

We have investigated how the general structure of the 'Key Analysis Errors' depends on the chosen norm. Fig. 3, panels (b), (c) and (d) show the root-mean-square (r.m.s.) of January 2002 'Key Analysis Errors' for temperature at model level 42 (620 hPa) for the experiments described in the previous section. Fig. 3(a) shows the r.m.s. of the Eady index for January 2002 calculated for the 300 hPa to 850 hPa layer using the Hoskins and Valdes (1990) definition. It is seen that the energy-norm sensitivities contain more small-scale

features and shorter frequency wave-train patterns than the **B**- and $\tilde{\mathbf{H}}$-norm sensitivities. The average **B**- and $\tilde{\mathbf{H}}$-norm sensitivities have a very similar structure, the main difference being larger amplitudes for the $\tilde{\mathbf{H}}$-norm sensitivities.

The energy norm sensitivities are clearly located in the areas with high Eady index, i.e. where the troposphere is most baroclinic. This feature has been discussed by Buizza and Palmer (1995). The **B**-norm and the approximate Hessian norm sensitivities typically have largest amplitude in the regions just south or north of the most unstable regions. They seem more often to adjust the baroclinic instability by modifying the temperature of the air masses north and south of the frontal regions.

Figures 4(a) and (b) show the power spectra of the temperature and vorticity fields for sensitivity perturbations at model level 42 on 1 January 2002, respectively. For reference, the spectra of the temperature and vorticity analysis fields are also shown. The analysis fields are at $T159$ resolution and the perturbations are calculated at $T63$ resolution. For both temperature and vorticity, energy norm perturbations have largest amplitude at smaller scales than the perturbations calculated using the **B**-norm and the $\tilde{\mathbf{H}}$-norm. The **B**-norm and $\tilde{\mathbf{H}}$-norm perturbations peak at total wave number 5-10 for temperature, and at total wave number 15-25 for vorticity; similar to the power spectra for the total fields. From the vertical profiles shown in Fig. 4(c) and (d) it is seen that the **B**-norm and $\tilde{\mathbf{H}}$-norm perturbations have larger vorticity perturbations and smaller temperature perturbations than the energy norm perturbations. The relative size of temperature versus vorticity amplitudes for the energy norm is determined by the scaling constants T_r and P_r in the (6) definition of the norm. The amplitudes of $\tilde{\mathbf{H}}$-norm perturbations do not differ much from the **B**-norm perturbations in Fig. 4. This may be because significantly more than 100 eigenvectors may be required to account for the smaller-scale influence of observations in the minimization process. It also reflects the dominant filtering rôle of **B** in the definition of $\tilde{\mathbf{H}}$ (7), and in the expression for the analysis gain matrix. A recent case study by Leutbecher and Palmer (personal communication 2004) on the 26 December 1999 storm 'Lothar' suggests that SVs computed with the exact Hessian as initial time metric can differ significantly from SVs computed with the **B**-norm as initial time metric. In their case study, the leading SVs computed with the exact Hessian metric exhibit vertically tilted temperature perturbations that are reminiscent of total energy singular vector structures rather than the equivalent barotropic structures of the **B**-norm SVs. We recommend further research into these issues.

Fig. 5 shows the vertical and horizontal structure of energy norm, **B**-norm and $\tilde{\mathbf{H}}$-norm perturbations at 1200 UTC 1 January 2002 for a region near Japan. The results for this individual case, and other cases not shown, reflect the average statistical results described above. 'Key Analysis Errors' calculated using the **B**-norm and $\tilde{\mathbf{H}}$-norm have similar structure, but have larger amplitude for the $\tilde{\mathbf{H}}$-norm than for the **B**-norm. 'Key Analysis Errors' calculated using the energy norm have much finer scale, and show wave-train features, both vertically and horizontally. All three flavours of perturbation have vertical tilts, most pronounced for the energy norm sensitivities. It is evident that the energy norm perturbations may locally have the opposite sign to the **B**-norm or $\tilde{\mathbf{H}}$-norm perturbations (e.g. over Eastern Japan).

Figure 6 shows a vertical cross section of the potential temperature analysis field along the same east-west axis as used in Fig. 5. It is seen that the main perturbations and vertical tilt occur in the baroclinic region near the western boarder of Figs. 5 and 6. The vertical and horizontal structure of the **B**-norm perturbations and $\tilde{\mathbf{H}}$-norm perturbations look more like analysis increments, as expected, due to the importance of the background error covariances in the assimilation process. Likewise there are clear similarities between the structure of energy norm perturbations and the singular vectors; this has been investigated by Gelaro *et al.* (1998).

5 Comparison of 'Key Analysis Errors' with observations

5.1 Methodology

Short-range forecasts from an analysis closer to the true atmospheric state can be expected to show smaller departures from reliable independent observations. As explained in Section 3, observations have been used to compare the quality of forecasts from initial conditions modified by the three types of 'Key Analysis Error' perturbations described in Section 2(c) against the quality of forecasts from unperturbed analyses.

Figure 7(a)-(v) shows the results from this evaluation for the observing systems and observed variables specified in Table 1 during a two-month period. The layout of the panels is similar. The horizontal axis is forecast time in hours, from 0 to 24. The vertical axis shows observation minus forecast r.m.s. statistics for all the observations within each one-hourly time slot.

The observation minus forecast differences, $\mathbf{d}_i$, are calculated at $T159L60$ resolution for each hourly time-slots i:

$$\mathbf{d}_i = \mathbf{y}_i - H(M_i(\mathbf{x}_0)) \tag{10}$$

where $\mathbf{y}_i$ represents the observations at time slot i. H is the nonlinear observation operator and M_i is the nonlinear forecast model integrated from time 0 to i valid at discrete 60 minute time intervals throughout the 24 hour forecast. We have, as a rule of thumb:

$$< \mathbf{d}_i^2 > = < \sigma_b^2 > + < \sigma_o^2 > \tag{11}$$

For all the observations used in this study it is reasonable to assume a constant observation error, σ_o. So the constant σ_o values specified in Table 1 have been subtracted from the statistics to get

$$< \mathbf{e}_i > = \sqrt{< \mathbf{d}_i^2 > - \sigma_o^2} \tag{12}$$

where $<>$ denotes the sample expectation and $\mathbf{e_i}$ is the 'error' shown as bars in Fig. 7. The white bars show statistics for forecasts performed from initial conditions modified using 'Key Analysis Error' perturbations and grey bars represent forecasts from unperturbed control analyses. Black bars show the difference multiplied by 25. Subtracting the observation error does not change the sign and structure of the difference plots; it is done in order to focus on the forecast error component. In these plots the left column is for energy-norm and right column is for the approximate Hessian ($\tilde{\mathbf{H}}$-norm) perturbations. The results for the $\mathbf{B}$-norm are quite similar to the $\tilde{\mathbf{H}}$-norm and are not shown. The name of the observing system and variable are shown on the plot. All plots are for high quality observing systems that (except for radiosondes) measure continuously or at hourly intervals. Only screened, quality-controlled observations are considered. Therefore it is possible to use the observations to evaluate the forecast quality every hour during the forecasts. By construction, the 'Key Analysis Error' perturbations improve the day-two forecast. So it is not surprising that towards the end of the 24 hour forecast window the perturbed forecasts are in better agreement with observations. This confirms the results from Fig. 2: that medium-range forecasts are better for both energy norm and $\tilde{\mathbf{H}}$-norm sensitivity perturbed analyses. However, during the first 12 hours or so, the sensitivity forecasts from both energy-norm and $\tilde{\mathbf{H}}$-norm based sensitivity-perturbed analyses are further away from observations than the control analyses.

5.2 Evaluation of results

Even though the differences in Fig. 7 between control forecasts (grey bars) and perturbed forecasts (white bars) are small, most of them are statistically significant due the large sample size of the order 100,000 observations (see Table 1) per hourly time slot. A statistical test is performed for each time slot during the first 12 forecast hours to investigate if the variance of the departure distribution for the two forecasts can be assumed to be identical. In Table 1 the hours that are more than 99% significant are listed for each observing system and variable. All significant results, except for one hour for radiosonde winds, show the same signal. E.g for American profilers (Fig. 7(a) and (b)) the 99% significance level corresponds to 0.25 on the left-hand vertical scale for the black bars, which is achieved for the entire initial 12 forecast hours. The ten other observing systems/variables in Fig. 7(c)-(v) that represent a large variety of observations and sample different parts of the atmosphere confirm the results from American profilers. The American profilers mainly measure winds at mid-tropospheric levels, where sensitivity perturbations according to Fig. 4(d) have largest amplitude. The ACARS aircraft measurements are predominantly from the 150-250hPa layer where sensitivity perturbations are smaller. The American profiler results are the most significant, showing that in the region where sensitivity perturbations are largest they are most significantly further away from observations. Both wind and temperature difference statistics for European AMDAR aircraft measurements (Fig. 7(c), (d), (k) and (l)) are in agreement with American ACARS statistics ((Fig. 7(e), (f), (m) and (n)). The aircraft results are significant to 99 % level for hours 4-8. The results are most robust for temperature measurements, reflecting the upper troposphere spectra in Fig. 4(c) and (d). The large Hessian norm departures during the first few hours of AMDAR measurements (Fig. 4(d) and (l)) are not significant because there are very few European aircraft from 0300-0500 UTC.

The SATOB wind departure statistics (Fig. 7(g) and (h)) are not monotonically increasing as a function of forecast length, unlike what is seen for most other observing systems. This is because the observations sample the globe (predominantly tropical and sub-tropical ocean regions), with a larger sample size and area every third hour. This is not a problem for this study because we are interested mainly in difference statistics (black bars) that within each time slot will be based on the same sample for the forecasts from control analyses and from perturbed analyses. The differences for SATOB are smaller than seen for American profilers but still significant for hourly time slots 4-10, as detailed in Table 1, due to the large sample size. QuikSCAT (Fig. 7(i) and (j)) estimates 10 metre winds over the ocean from a solar-synchronous polar orbiting satellite and thus does not measure the globe uniformly within each hourly time slot. Like SATOB measurements, they show a non-monotonic behaviour, but the difference statistics nevertheless contain useful information. Due to the very large sample size the QuikSCAT are significant at level 99% for hours 5-8. The smaller absolute difference values are due to the smaller amplitude on perturbations near the surface, partly because the surface fields are not perturbed.

The SYNOP MSL pressure observations are mainly from land areas but also from ships. This is a high quality data source with a large amount of observations. Many automatic stations report hourly data of very high quality, whereas every third hour less accurate (but still good quality) data from manual SYNOP observations are available. The departure statistics reflect this three-hourly pattern (Fig. 7(o) and (p)). The difference statistics show a clear and highly significant result during the first 6 hours for energy-norm perturbations, and for the first 12 hours (except hour 10) for $\tilde{\mathbf{H}}$-norm perturbations. For DRIBU (Fig. 7(q) and (r)) the results are similar to SYNOP results, but less robust due to a smaller sample size. Radiosonde observations are only available at synoptic time so it is only possible to do limited evaluation based on radiosonde data. But the results (Fig. 7(s), (t), (u) and (v)) from European radiosonde wind and temperature measurement confirm the finding from the other observing systems (see Table 1). It is also encouraging for the confidence in these results that the wind observation difference statistics are most pronounced for the $\tilde{\mathbf{H}}$-norm perturbation forecasts, as is to be expected from Fig. 4(d).

We can conclude that for almost all observing systems used in this study, the forecasts from the control analysis are more in agreement with observations during the first 12 hours. We take this as an indication that the 'Key Analysis Error' perturbations that improve the control forecasts do not represent observable analysis error, but rather artificial adjustments to the initial state that, due to the formulation of the objective cost function, are bound to improve the 2-5 day forecast. The fact that very different perturbations (viz. perturbations generated using the energy norm and the $\tilde{\mathbf{H}}$-norm) are capable of producing similar improvements to the two-day forecasts is a clear indication that 'Key Analysis Error' method is unable to unequivocally identify errors in the analysis for either of these two norms. It could be that both norms are poor approximations of the true analysis error covariance, or that the assumptions of linear dynamics and zero model error of the tangent-linear forecast are detrimental. Uncertainty of the final-time forecast error estimate will also lead to an additional uncertainty of 'key analysis errors'. The introduction of additional constraints at initial time might be one way to resolve these uncertainties. We would recommend further research into these issues.

6 Discussion and conclusions

The spectacular success of adjoint sensitivity-based (AS) perturbations in correcting medium-range forecast failures has fostered the perception that they represent actual analysis error. Through detailed comparison with hourly observations we have shown that the AS-perturbed analyses improve forecasts from 12 hours onwards, following an initial period of significant degradation. This result indicates that AS-perturbations (including the so-called 'Key Analysis Errors') do not unequivocally represent true analysis error. We have shown that the structure and magnitude of the AS-perturbations depend strongly on the initial norm used in the calculations, in a similar way that singular vectors (SV) depend on the initial-time norm (Kim and Morgan 2002). In both cases (AS and SV) the initial norm should represent the covariance of analysis error. In this study we have tested three different norms: the dry total energy, the Hessian of the analysis cost-function, and the background error covariance. The energy norm perturbations are more closely associated with baroclinic regions than the other two norms. The energy norm perturbations are very small-scale, whereas $\mathbf{B}$ and $\tilde{\mathbf{H}}$-norm perturbations are smooth larger-scale structures similar to analysis increments. There are significant differences between the three sets of AS-perturbations at initial time, but all three perform poorly in the initial 12 hours of forecast verification against observations.

In AS calculations, information about the short-range forecast errors (typically 48 hours) is used to deduce information about the analysis error at initial time (48 hours earlier). Real-time applications such as observation targeting, ensemble prediction and flow dependent data assimilation cannot utilize forecast verification at a future time. In several such applications SVs have been used instead, to gain information about the likely fastest-growing components of analysis error. Just like AS, SVs rely on an initial-time norm to provide the appropriate scaling of the various components of the initial-time perturbations. AS and SVs have many characteristic features in common: both tend to represent strongly tilting structures with maxima in the lower troposphere, located in the most baroclinic regions. Our findings about AS perturbations are therefore relevant also to SV structures. There is evidence from ECMWF's operational EPS that its SV-based initial perturbations are incompatible with available observations (Buizza *et al.* 2003).

In ensemble prediction applications for the medium-range, the precise location and structure of the initial-time perturbations are not of such crucial importance, as long as the perturbations evolve into structures that are relevant for the real forecast error growth (Reynolds *et al.* 2001). In the reduced-rank Kalman Filter (RRKF), on the other hand, precise knowledge of the structure is required for the definition of a flow-dependent sub-space within which the analysis modification is confined. In other words, the RRKF modification of the analysis increment is a linear combination of the SVs. So this is an important distinction between the use of SVs to determine EPS perturbations and the AS methods. In the latter to find a possible 'correctable' part of the

analysis error, it is crucial to have a high-quality representation of the analysis error covariance matrix. Fisher and Andersson (2001) showed that a successful RRKF is difficult to achieve in the context of operational 4D-Var NWP with its severe restrictions on the dimension of the SV subspace (of the order 25) due to computational costs, whereas Beck and Ehrendorfer (2004) demonstrated the validity of the RRKF approach in a 3-layer quasi-geostrophic model using relatively larger subspaces (100 SVs for a model with 1400 degrees of freedom). For observation targeting applications (e.g. Langland *et al.* 1999) it is questionable if AS perturbations or SV structures can determine the exact location of analysis error accurately enough to guide where to deploy a small number of dropsondes. Benefit from observation targeting may be expected if several additional observations are deployed more comprehensively over larger areas in the vicinity of sensitivity perturbations (Leutbecher *et al.* 2002; Leutbecher 2003).

Climatological maps of AS-perturbations (Marseille and Bouttier, 2001) have been used for observation network design and for making statements about the potential impact of future satellite instruments (e.g. McNally 2002; Tan and Andersson 2004). These climatologies tend to highlight the baroclinic regions of the atmosphere. At least the general conclusions from such studies may be valid in spite of some uncertainty due to the initial norm, but these results most likely do not add more information than could be obtained from simple baroclinic Eady index (Hoskins and Valdes 1990) calculations and climatologies. Information-content studies with respect to specific vertical structures of AS-perturbations (e.g. Rabier *et al.* 2002) are useful as illustrations, which, however, need to be interpreted with the knowledge that the structures themselves are highly uncertain and depend on the norm used. So studies based on e.g. energy norm perturbation may incorrectly suggest that it is important to measure temperature and wind accurately in shallow mid-tropospheric layers. In a MODIS wind impact study Bormann and Thépaut (2004) link the reduction of the sensitivity perturbations on a specific day to a reduction in analysis errors. Based on our results, that type of diagnostic is not recommendable in general! It should be noted that our results are not in contradiction to the positive results obtained by Hello *et al.* (2000) and Hello and Bouttier (2001) as in their studies they used an AS-technique to provide a 3D-Var analysis scheme with a means to modify the deepening rate of specific storm cases - a mechanism which is fully accounted for in 4D-Var (Järvinen *et al.* 1999).

Our results have had immediate implication on current ECMWF research plans. Research on the RRKF has been abandoned, and new research will focus more on ensemble-based methods. Ensemble-based methods are seen as a promising and fruitful way to gain knowledge about analysis and short-range forecast error.

It should not be forgotten that the 4D-Var algorithm itself relies on linear dynamics and the perfect model assumption. These assumptions are less severe in 4D-Var than in most AS and SV applications due to the shorter time window (typically 6 or 12 hours, rather than 24 or 48 hours). As a further parallel with the AS and SV issues investigated here, we note that the initial-time norm in 4D-Var is provided by the background error covariance matrix - the covariance evolution of forecast error within the 4D-Var assimilation window is thus dominated by **B**-norm SVs. We conclude that progress in modelling of all important error sources in data assimilation, including linearisation errors (Trémolet 2004; Janisková *et al.* 2002), observation biases and model error (Zupanski 1997; Trémolet 2003), is required for further understanding and improvement in a range of applications relying on linear dynamics to characterize initial condition error.

Acknowledgements

References

Barkmeijer, J., van Gijzen, M. and Bouttier, F., 1998: Singular vectors and estimates of the analysis-error covariance metric. *Q. J. R. Meteorol. Soc.*, **124,** 1695–1713

Barkmeijer, J., Buizza, R. and Palmer, T.N., 1999: 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, **125,** 2333–2351

Beck, A. and Ehrendorfer, M., 2004: Singular-Vector-Based Covariance Propagation in a Quasigeostrophic Assimilation System. Submitted to *Mon. Weather Rev.*

Buizza, R., Houtekamer, P.L., Toth, Z., Pellerin, G., Wei, M. and Zhu,Y., 2003: 'Assessment of the status of global ensemble prediction'. Proc. ECMWF Seminar on "Predictability of weather and climate", Reading, U. K., 9–13 September 2002, 127–154

Buizza, R. and Palmer, T.N., 1995: The Singular-Vector Structure of the Atmospheric Global Circulation. *J. Atmos. Sci.*, **52,** 1434–1456

Bormann, N. and Thépaut, J.-N., 2004: Impact of MODIS Polar Winds in ECMWF's 4DVAR Data Assimilation System. *Mon. Weather Rev.*, **132,** 929–940

Bouttier, F., 2001: 'The development of 12-hourly 4D-Var'. *ECMWF Tech. Memo.*, **348,** pp 21

Courtier, P., 1993: 'Introduction to Numerical Weather Prediction data assimilation methods'. Proc. ECMWF Seminar on "Developments in the Use of Satellite Data in Numerical Weather Prediction", Reading, U. K., 6–10 September 1993, 189–207.

Courtier, P., Thépaut, J.-N., and Hollingsworth, A., 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.*, **120,** 1367–1387

Emanuel, K., Raymond, D., Betts, A., Bosart, L., Bretherton, C., Droegemeir, K., Farrell, B., Fritsch, M., Houze, R., Le Mone, M., Lilly, D., Rotunno, R., Shapiro, M., Smith, R., and Thorpe, A., 1995: Report of the first Prospectus Development Team of the U.S. Weather Research Program to NOAA and the NSF. *Bull. Amer. Meteorol. Soc.*, **76,** 1194–1208

Fisher, M., 1998: 'Development of a simplified Kalman filter'. *ECMWF Tech. Memo.*, **260,** pp 16

Fisher, M. and Andersson, E., 2001: 'Developments in 4D-Var and Kalman Filtering'. *ECMWF Tech. Memo.*, **347,** pp 36

Gilbert, J.C. and Lémarechal, 1989: Some numerical experiments with variable storage quasi-Newton algorithms. *Linear Algebra Appl.*, **113**, 407–436

Gelaro, R., Buizza, R., Palmer, T.N. and Klinker, E., 1998: Sensitivity Analysis of Forecast Errors and the Construction of Optimal Perturbations Using Singular Vectors. *J. Atmos. Sci.*, **55,** 1012–1037

Hansen, J. A. and Smith, L. A., 2000: The Role of Operational Constraints in Selecting Supplementary Observations. *J. Atmos. Sci.*, **57,** 2859–2871

Hello, G., Lalaurette, F. and Thépaut, J.-N., 2000: Combined use of sensitivity information and observations to improve meteorological forecasts: A feasibility study applied to the "Christmas storm" case. *Q. J. R. Meteorol. Soc.*, **126,** 621–647

Hello, G. and Bouttier, F., 2001: Using adjoint sensitivity as a local structure function in variational data assimilation. *Nonlinear Processes in Geophysics*, **8,** 347–355

Hoskins, B. J. and Valdes, P. J., 1990: On the Existence of Storm-Tracks. *J. Atmos. Sci.*, **47,** 1854–1864

Houtekamer, P. L. and Mitchell, H. L., 1998: Data Assimilation Using an Ensemble Kalman Filter technique. *Mon. Weather Rev.*, **126,** 796–811

Houtekamer, P. L. and Mitchell, H. L., 2001: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Weather Rev.*, **129,** 123–137

Kim, H. M. and Morgan, M. C., 2002: Dependence of Singular Vector Structure and Evolution on the Choice of Norm. *J. Atmos. Sci.*, **59,** 3099–3116

Janisková, M., Mahfouf, J.-F., Morcrette, J.-J. and Chevallier, F., 2002: Linearized radiation and cloud schemes in the ECMWF model: Development and evaluation. *Q. J. R. Meteorol. Soc.*, **128,** 1505–1527

Järvinen, H., Andersson, E. and Bouttier, F., 1999: Variational assimilation of time sequences of surface observations with serially correlated errors. *Tellus*, **51A,** 469–488

Klinker, E., Rabier, F. and Gelaro, R., 1998: Estimation of key analysis errors using the adjoint technique. *Q. J. R. Meteorol. Soc.*, **124,** 1909–1933

Langland, R. H., Gelaro, R., Rohaly, G. D. and Shapiro, M. A. , 1999: Targeted observations in FASTEX: Adjoint-based targeting procedures and data impact experiments in IOP17 and IOP18. *Q. J. R. Meteorol. Soc.*, **125,** 3241–3270

Langland, R. H., Shapiro, M. A. and Gelaro, R., 2002: Initial Condition Sensitivity and Error Growth in Forecasts of the 25 January 2000 East Coast Snowstorm. *Mon. Weather Rev.*, **130,** 957–974

Leutbecher, M, 2003: A Reduced Rank Estimate of Forecast Error Variance Changes due to Intermittent Modifications of the Observing Network. *J. Atmos. Sci.*, **60,** 729–742

Leutbecher, M, Barkmeijer, J, Palmer, T. N., Thorpe, A. J., 2002: Potential improvement to forecasts of two severe storms using targeted observations. *Q. J. R. Meteorol. Soc.*, **128,** 1641–1670

Lorenz, E. N., 1995: 'Predictability - A problem partly solved'. Proc. ECMWF Seminar on "Predictability", Reading, U. K., 4–8 September 1995, 1–18.

Marseille, G.-J. and Bouttier, F., 2001: 'Climatologies of sensitive areas for short term forecast errors over Europe - EUMETNET-EUCOS study'. *ECMWF Tech. Memo.*, **334,** pp 50

McNally, A. P., 2002: A note on the occurrence of cloud in meteorologically sensitive areas and the implications for advanced infrared sounders. *Q. J. R. Meteorol. Soc.*, **128,** 2551–2556

Molteni, F. and Palmer, T.N., 1993: Predictability and finite-time instability of the northern winter circulation. *Q. J. R. Meteorol. Soc.*, **119,** 269–298

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., 1998: Singular Vectors, Metrics and Adaptive Observations. *J. Atmos. Sci.*, **55,** 633–653

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., Järvinen, H., Klinker, E., Mahfouf, J.-F. 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, Prunet, P., 2002: Channel selection methods for Infrared Atmospheric Sounding Interferometer radiances. *Q. J. R. Meteorol. Soc.*, **128,** 1011–1027

Reynolds, C.A., Gelaro, R. and Doyle, J.D., 2001: Relationship between singular vectors and transient features in the background flow. *Q. J. R. Meteorol. Soc.*, **127,** 1731–1760

Reynolds, C.A. and Palmer, T. N.., 1998: Decaying Singular Vectors and Their Impact on Analysis and Forecast Correction. *J. Atmos. Sci.*, **55,** 3005–3023

Simmons, A. J. and Hollingsworth, A., 2002: Some aspects of the improvements in skill of numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **128,** 647–677

Tan, D. G. H. and Andersson, E., 2004: 'Simulation of the yield and accuracy of wind profile measurements from the Atmospheric Dynamics Mission (ADM-Aeolus)'. *ECMWF Tech. Memo.*, **431,** pp 31

Thépaut, J.-N. and Courtier, P., 1991: Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. *Q. J. R. Meteorol. Soc.*, **117,** 1225-1254

Trémolet, Y., 2003: 'Model Error in Variational Data Assimilation'. Proc. ECMWF Seminar on "Recent developments in data assimilation for atmosphere and ocean", Reading, U. K., 8–12 September 2003, 361–367

Trémolet, Y., 2004: Diagnostics of linear and incremental approximations in 4D-Var. To appear in *Q. J. R. Meteorol. Soc.*, **130,**

Zupanski, D., 1997: A General Weak Constraint Applicable to Operational 4DVAR Data Assimilation Systems. *Mon. Weather Rev.*, **125,** 2274–2292

Table 1: ASSUMED OBSERVATION ERRORS, DATA COUNTS AND SIGNIFICANT HOURS

Observing system	Observation type	Obs. error	Data per hour	99% energy	99% $\tilde{\mathbf{H}}$	Figure
American Profilers	Wind 200-800 hPa	1.8 m/s	78000-83000	1-12	1-12	7(a)-(b)
AMDAR Europe	Wind 200-800 hPa	2.6 m/s	2000-42000	3-7,10,12	3-6	7(c)-(d)
ACARS U.S.A.	Wind 200-800 hPa	2.6 m/s	61000-273000	none	1-9	7(e)-(f)
SATOB N.Hem.	Wind	2.2 m/s	36000-216000	1,3,5,8	1-6,8,9,11	7(g)-(h)
QuikSCAT	10 metre wind	1.3 m/s	217000-1514000	2,4,6-8,10-12	3-6	7(i)-(j)
AMDAR Europe	Temp. 200-800 hPa	0.8 K	1000-20000	1-8	3-6	7(k)-(l)
ACARS U.S.A.	Temp. 200-800 hPa	0.8 K	30000-137000	1-8	1-7	7(m)-(n)
SYNOP N.Hem.	MSL Pressure	0.5 hPa	40000-178000	1-5	1-12	7(o)-(p)
DRIBU N.Hem.	MSL Pressure	0.7 hPa	3400-5100	1-5	2-6	7(q)-(r)
TEMP/PILOT EU	Wind 200-800 hPa	2.85 m/s	11000-94000	1,6,7	3	7(s)-(t)
TEMP EU	Temp. 200-800 hPa	0.9 K	6000-51000	1,(6),7	none	7(u)-(v)

The data per hour numbers show the range for the 25 hourly time slots. The 99% energy and 99% $\tilde{\mathbf{H}}$ columns show the hourly time slots during the first 12 hours for which it is more than 99% significant that short-range forecasts from sensitivity perturbed fields are further away from observations than forecasts from unperturbed analyses.
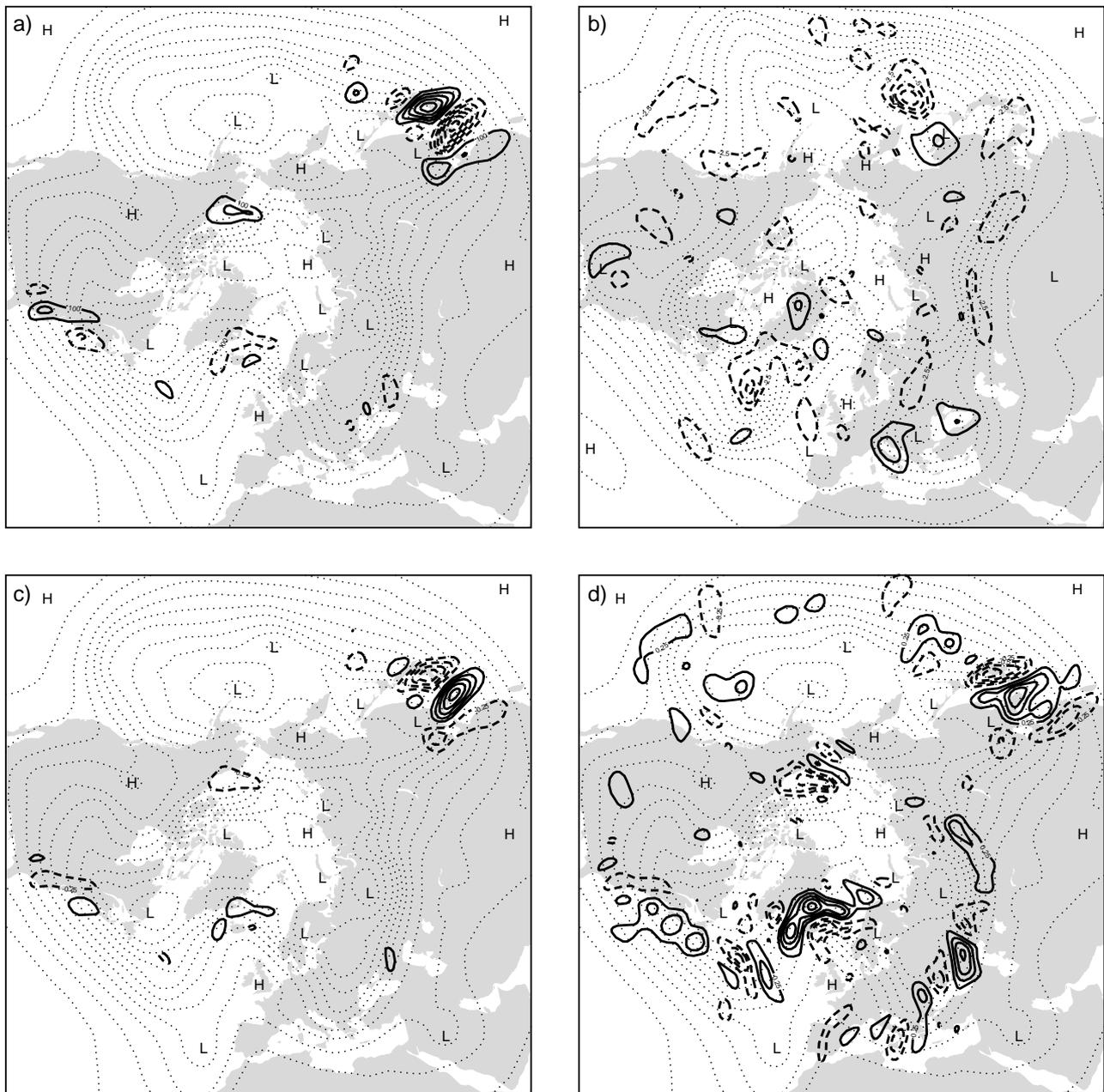
Figure 1: Panel a: Initial-time sensitivity gradient, with respect to temperature at model level 43 (650 hPa), of the 48-hour forecast error for a forecast starting at 1 January 2002, 1200 UTC, shown as thick dashed (negative values) and thick solid (positive values) contours. The forecast error diagnostic function J is restricted to the area to the north of 30°N. Panel b: The 48-hour temperature forecast error at model level 43 (650 hPa) in the area to the north of 30°N for a forecast starting at 1 January 2002, 1200 UTC. Negative values are shown as thick dashed contours and positive values are shown as thick solid contours with a 2.5 K interval. Panel c: 'Key Analysis Errors' (Klinker et al. 1998) after one iteration. Panel d: Like panel c but after three iterations. Temperatures are shown as thick dashed (negative) thick solid (positive) contours with a 0.25 K interval. The thin dotted contours (8 m interval) show the 500 hPa geopotential height analysis field valid at 1 January 2002, 1200 UTC on panels a, c, d and valid at 3 January 2002, 1200 UTC on panel b.
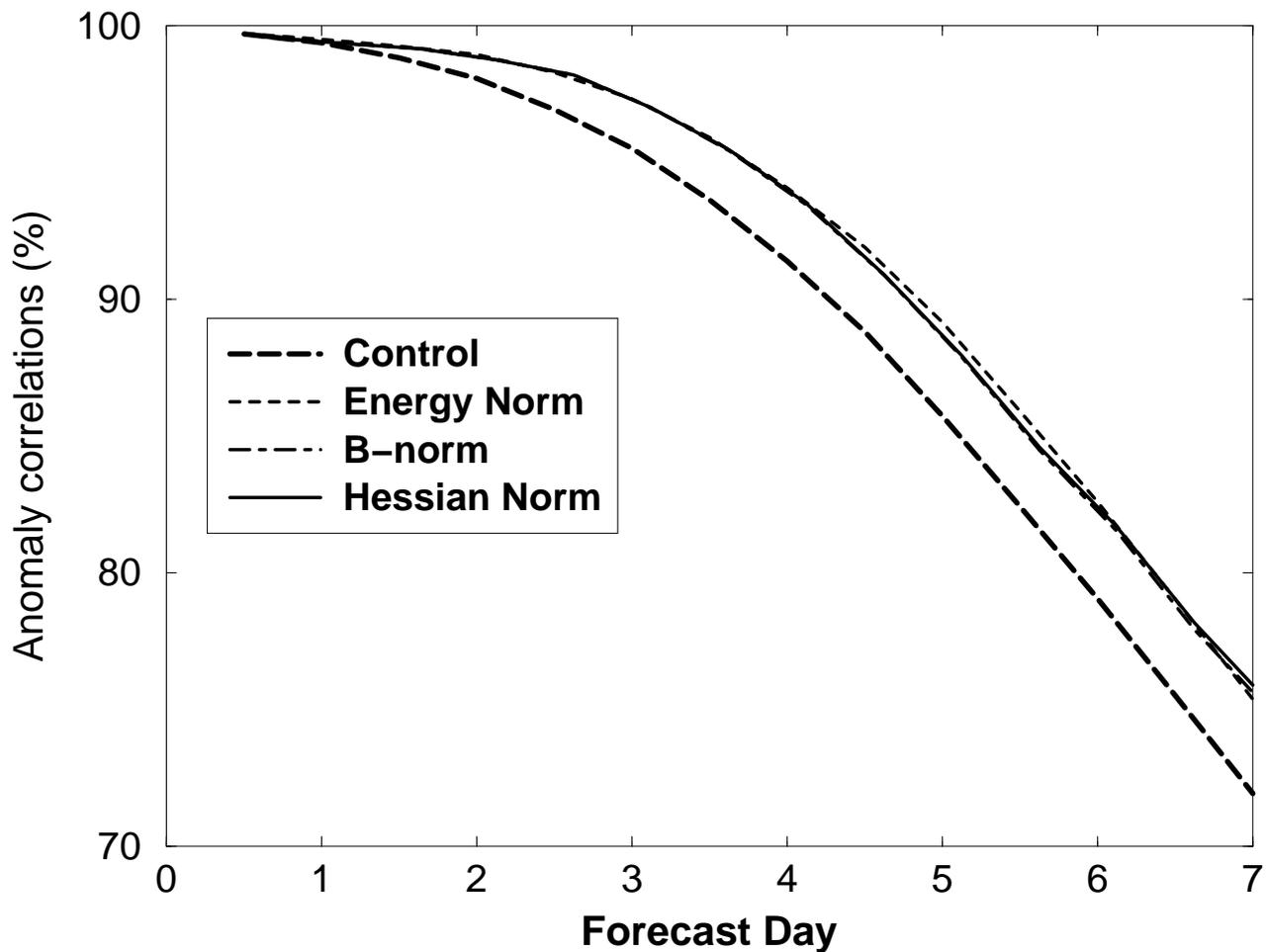
Figure 2: Anomaly correlations of forecast error in terms of 500 hPa geopotential height (m), for the Northern Hemisphere extra-tropics averaged over 29 days in December 2001. Forecasts from analyses perturbed by the 'Key Analysis Errors' (dashed curve: Energy norm, dot-dashed curve: B-norm, solid curve: Hessian norm) clearly outperform forecasts from the control analyses (thick long dashed) irrespective of which of the three norms is used.
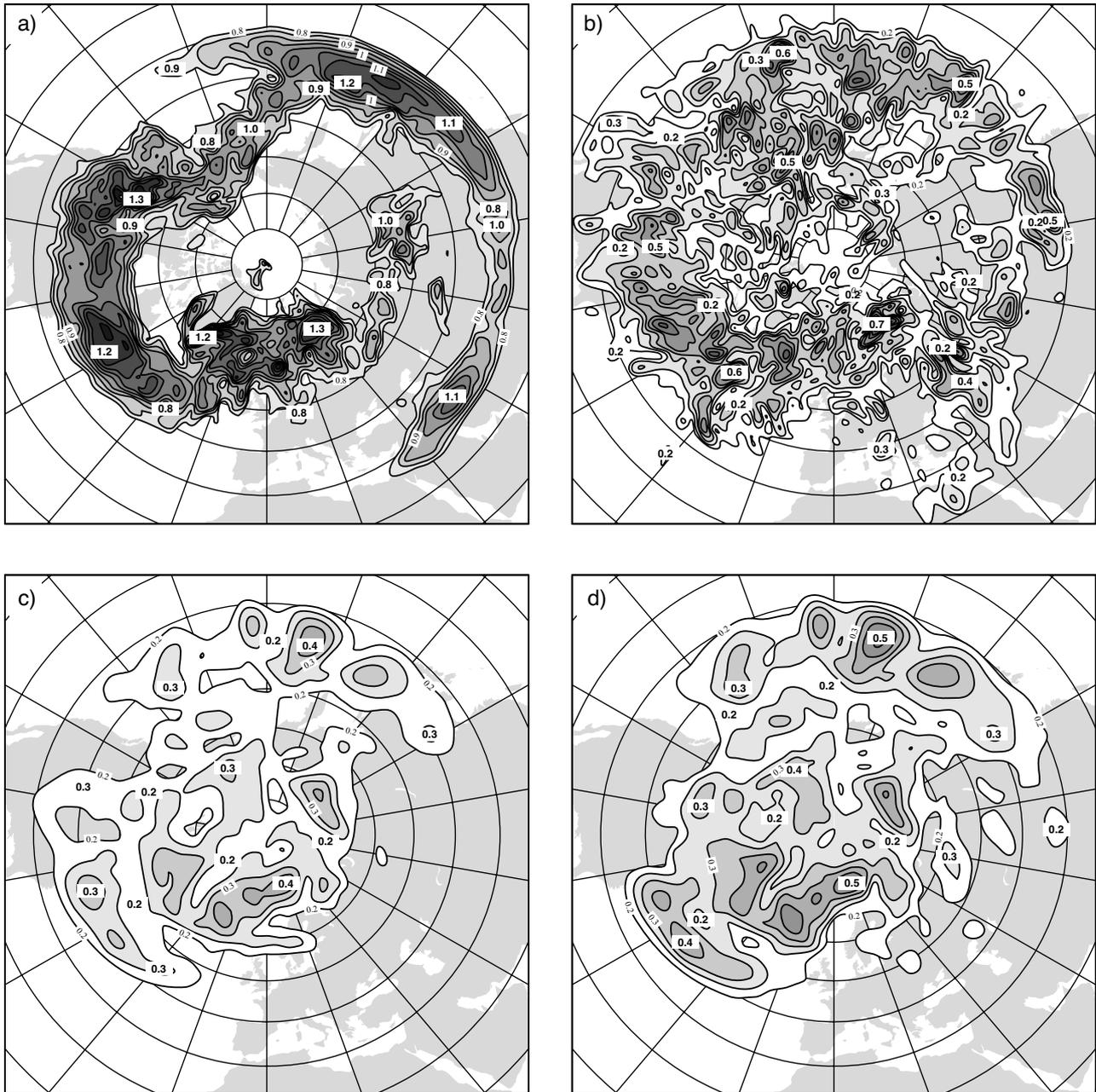
*Figure 3: Root-Mean-Square values based on 1200 UTC fields from 29 days in January 2002 for a) the Eady index calculated from 300 hPa and 850 hPa fields, and 'Key Analysis Errors' at model level 42 (620 hPa) for three different initial norms: b) the total energy norm, c) the background-error covariance norm (**B**-norm), and d) the approximate Hessian norm (**H̃**-norm).*
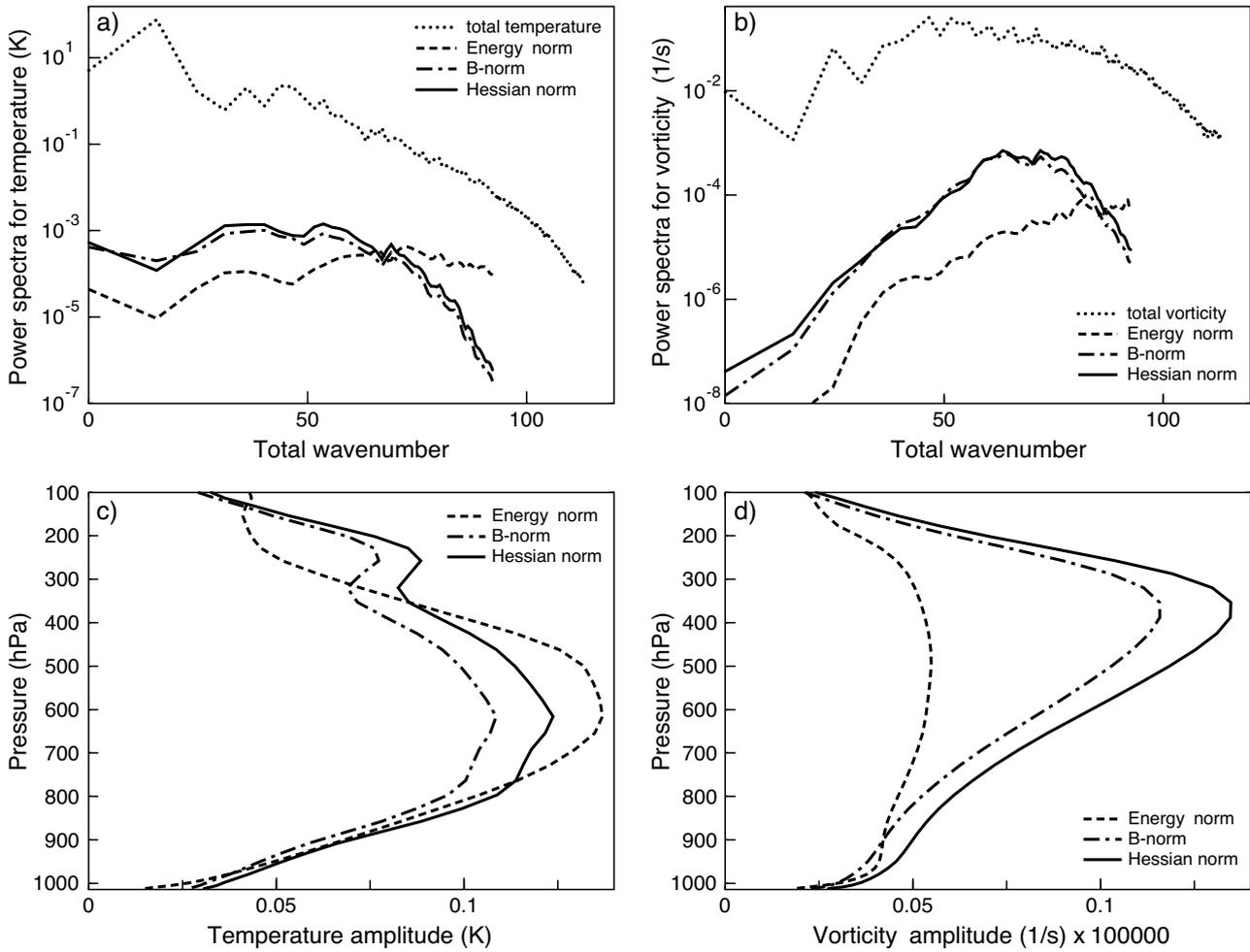
*Figure 4: Power spectra for temperature (K) (panel a) and vorticity (s^{-1}) (panel b) at model level 42 (near 620 hPa) valid at 12 UTC 1 January 2002. Dashed curves: energy- norm key analysis error. Dash-dotted curves: **B**-norm key analysis error. Solid curves: **H̃**-norm key analysis error. Also shown (as thin dotted curves) are the power spectra for the analysed temperature and vorticity fields. Panel c shows the amplitude of the temperature component of the 'Key Analysis Errors' for model levels 20 to 60 (from around 35 hPa to the surface). The colour coding of curves is the same as in panel a. Panel d: Similar to panel (c) but for the vorticity component (s^{-1}), scaled by a factor 10^5.*
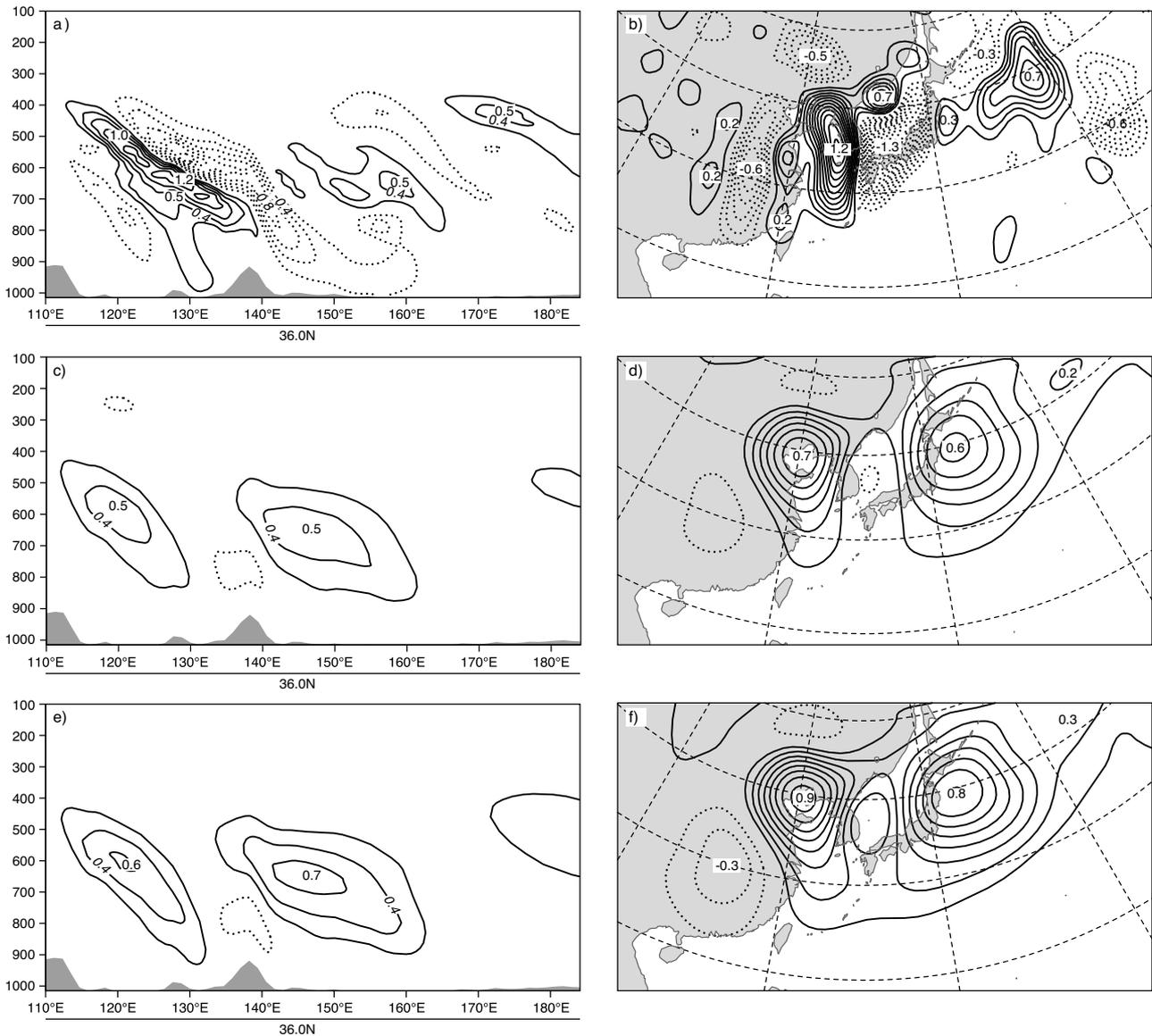
*Figure 5: The left column shows cross-sections of temperature perturbations ('Key Analysis Errors') near Japan along latitude 36°N valid 1200 UTC, 1 January 2002. Contour interval of 0.2 K used. The right column shows the corresponding temperature perturbations at model level 42 (620 hPa). Contour interval of 0.1 K used. Solid contours represent positive values and dashed contours represent negative values. Panels a and b: For energy norm, Panels c and d: For the background-error covariance norm (**B**-norm). Panels e and f: For the approximate Hessian norm (**H̃**-norm).*
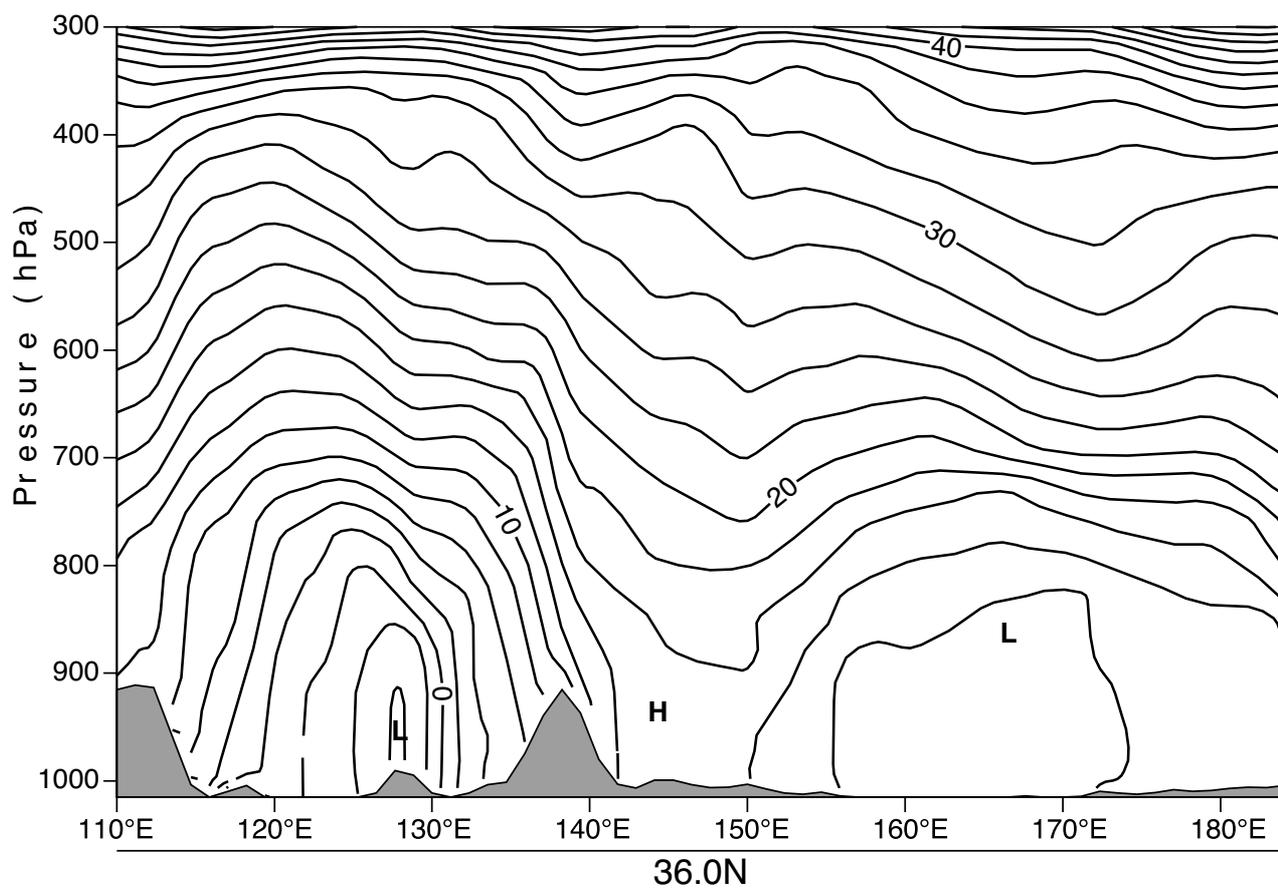
Figure 6: Potential temperature east-west cross section along the axis used in Fig. 5. Contour interval of 2.0 K used.
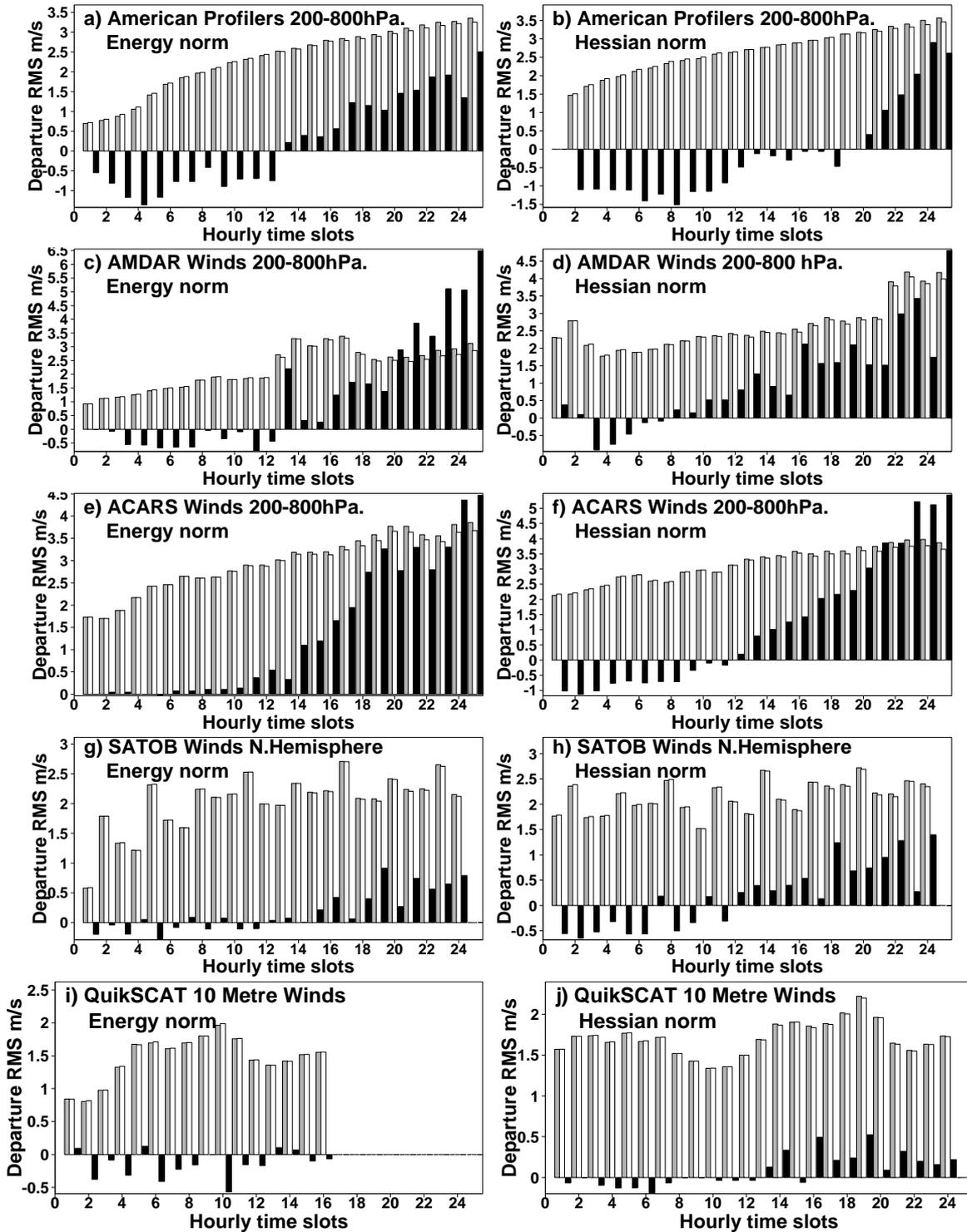
Figure 7: Observation minus forecast statistics as function of forecast time in hourly bins. The constant observation errors specified in Table 1 has been subtracted from the statistics (see section 5 for further details). The white bars show statistics for forecasts performed from initial conditions modified using key analysis error perturbations and grey bars represent statistics for forecasts from unperturbed control analyses. Black bars show the difference multiplied by 25. Left column is for energy-norm and right column is for the approximate Hessian perturbations. A variety of observing systems are shown, as indicated.
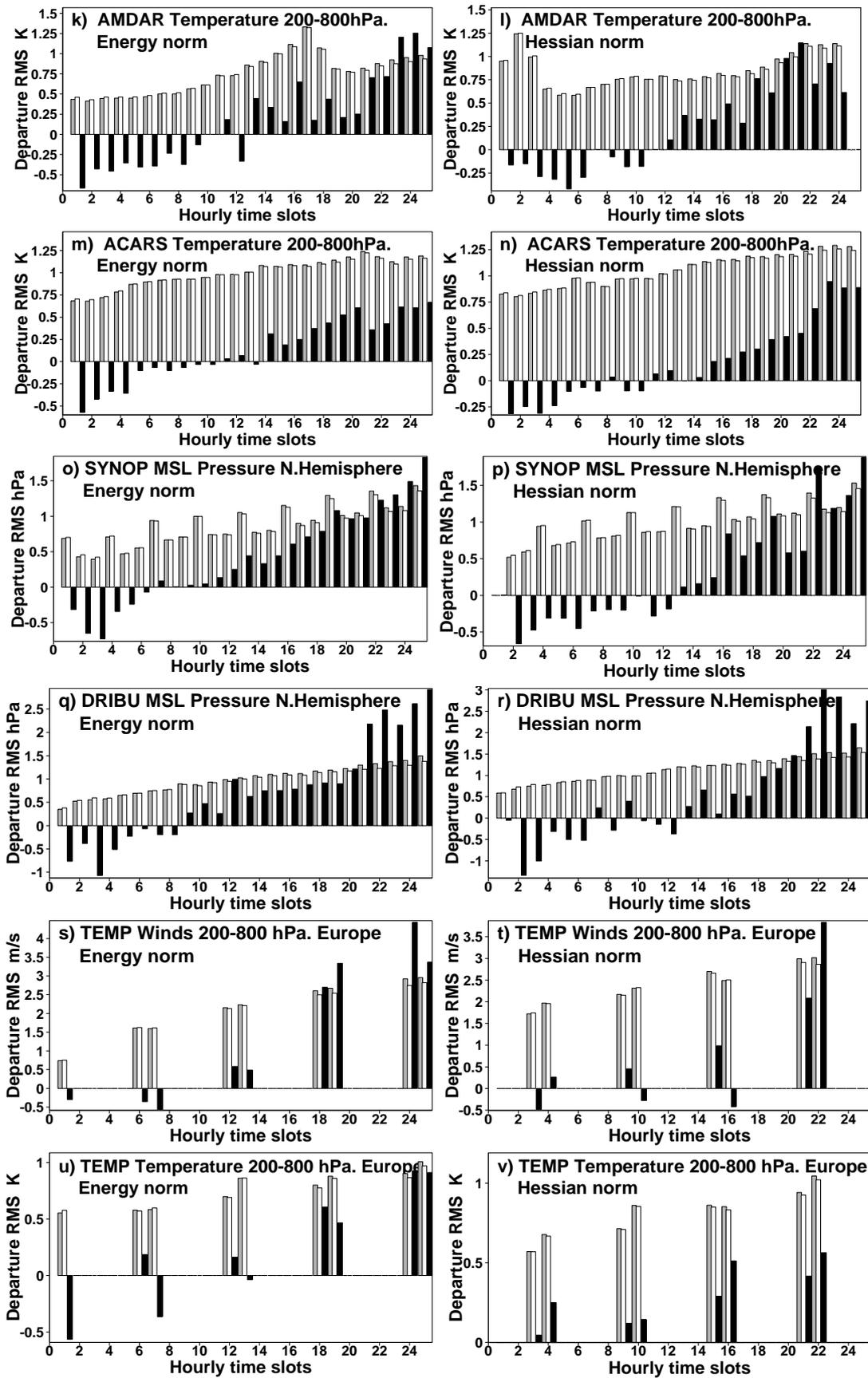
Figure 8: Figure 7 continued.