SENSITIVITY OF FORECAST ERRORS TO INITIAL CONDITIONS

E Klinker, F Rabier and R Gelaro*
European Centre for Medium-Range Weather Forecasts
Reading, Berkshire, UK
*Naval Research Laboratory, Monterey, California 93943, USA

1. INTRODUCTION

The quality of medium-range forecasts has been shown to depend crucially on the initial conditions in a number of predictability studies (e.g. *Molteni et al*, 1996). The adjoint technique (*Le Dimet and Talagrand*, 1986) provides us with a tool to directly investigate the sensitivity of a forecast aspect to the initial conditions. Adjoint equations have been used for a wide range of sensitivity problems. The sensitivity to model parameters has been investigated, for instance by *Hall* (1986), *Marais and Musson-Genon* (1992), *Rinne and Järvinen* (1993). The sensitivity of one aspect of the forecast to initial conditions has been the subject of various studies (*Errico and Vukicevic*, 1992; *Rabier et al*, 1992; *Errico et al*, 1993a; *Langland et al*, 1995; *Oortwijn and Barkmeijer*, 1995).

In a previous paper (*Rabier et al*, 1996), the sensitivity of forecast errors with respect to the analysis was investigated using this adjoint technique. The gradient of the energy of the day 2 forecast errors provided a sensitivity pattern which could be interpreted as a sum of rapidly growing components of analysis errors. An analysis modified by subtracting an appropriately scaled vector, proportional to the gradient, provided initial conditions for a "sensitivity integration" which could be used to diagnose the effect of initial-data errors on forecast errors. Statistics for the month of April 1994 have characterized the sensitivity patterns as small-scale, middle or lower troposphere structures tilted in the vertical. The general pattern of these structures is known to be associated with the fastest possible growth of perturbation (*Farrell*, 1990). When used as initial perturbations, an improvement in the skill of not only the short but the medium-range forecast as well was obtained.

These sensitivity integrations were run with subjectively scaled gradient perturbations. Further experimentation was performed with an objective scaling factor optimizing the reduction of forecast errors for perturbations proportional to the gradient. However, the impact on the forecast performance was fairly small as both objectively and subjectively chosen factors were found to be very close.

More importantly, it might be argued that perturbations obtained from the gradient are by no means a unique combination of fast-growing components of analysis errors. The current study aims at broadening the relevant optimization problem. Instead of seeking the most efficient perturbation needed to cancel part of the day 2 forecast error, a range of "sensitive" perturbations of the analysis obtained by a minimization algorithm will be

investigated. Firstly, information gathered over several seasons of gradient computations is described. The basic properties of the sensitivity gradient and factors influencing its structure are discussed in section 2.

In section 3 variations on the way an "optimal" perturbation is obtained are investigated. Firstly, the influence of the choice of inner-product on the gradient of two-day forecast errors is studied. Secondly, the benefit of performing a few steps of a minimization algorithm is investigated. In section 4 the choice of three steps of minimization is justified on the basis of a comparison with observations. There are some limitations in deriving sensitivity perturbations that could correct analysis and subsequent forecast errors as shown in section 5. The performance of model integrations based on the revised operational sensitivity calculations, performing three steps in the minimization, is discussed in section 6 followed by concluding remarks in section 7.

2. SENSITIVITY GRADIENTS

2.1 **Seasonal variation of sensitivity gradients**

Following the definition of sensitivity gradients in *Rabier et al* (1996), the gradient of two day forecast errors depends on the forecast errors and on the stability of the flow through the adjoint operator R^*. The comparison of the sensitivity between summer (Fig 1a) and winter (Fig 2a) shows quite a large difference in the magnitude and the horizontal distribution. Whereas the day-2 rms-forecast errors for the two seasons (Figs 1b and 2b) exhibit only a modest increase in winter compared to summer, the sensitivity in winter is almost a factor of 3 larger than in summer. The obviously important relation of the sensitivity to the stability property of the troposphere can be inferred from the `Eady index'

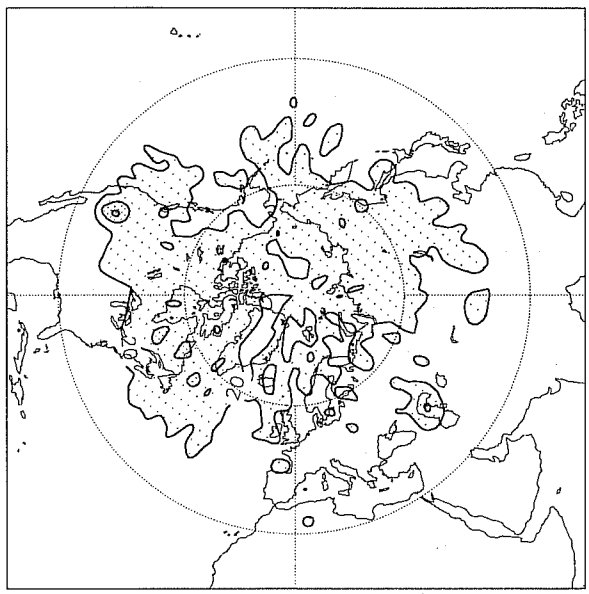
$$\sigma = 0.31 \frac{f}{N} \frac{du}{dz}$$

where f is the Coriolis parameter, N is the buoyancy frequency and $\frac{du}{dz}$ is the vertical wind-shear (300/1000 hPa).

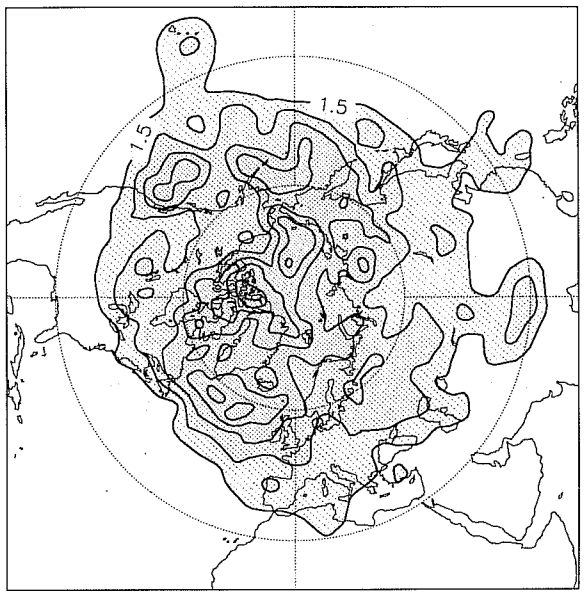
As the Eady index approximates the maximum growth rate of baroclinic waves, it provides a suitable measure for the baroclinicity of the flow, as shown for example by *Hoskins and Valdes* (1990). In winter (Fig 2c) the areas of maximum instabilities are found over the mid-latitude oceanic regions of the Northern Hemisphere. In the Pacific both the maximum of instability of the flow and the maximum values of sensitivity are located in the Western Pacific where cyclones are mostly in their early developing phase. Consequently the day-2 errors downstream in the exit area of the mid-latitude Pacific-jet are related to the sensitivity in the centre of the jet and further upstream, with maximum sensitivity in the entrance area of the jet.

In the Atlantic the displacement of errors relative to the major instability zone is smaller than in the Pacific, mainly as a consequence of weaker westerly flow. Therefore the spatial separation of forecast errors and

a) JJA 1994 RMS of Temperature Sensitivity (~730 hPa)



b) D2 RMS-Error JJA 1994 500 hPa



c) JJA 1994 EADY-INDEX

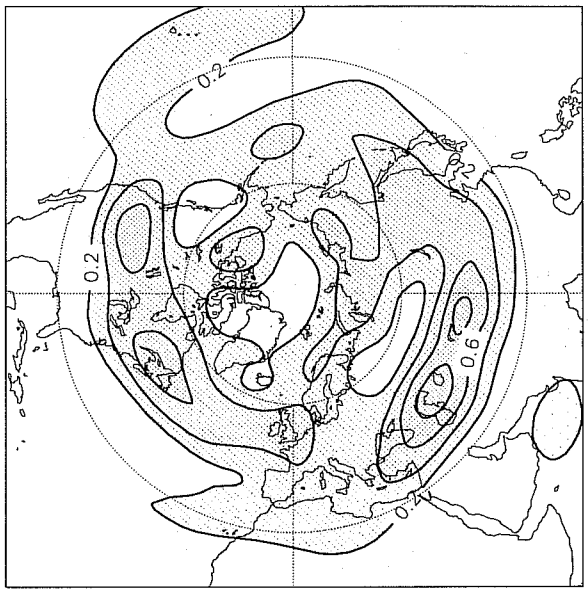
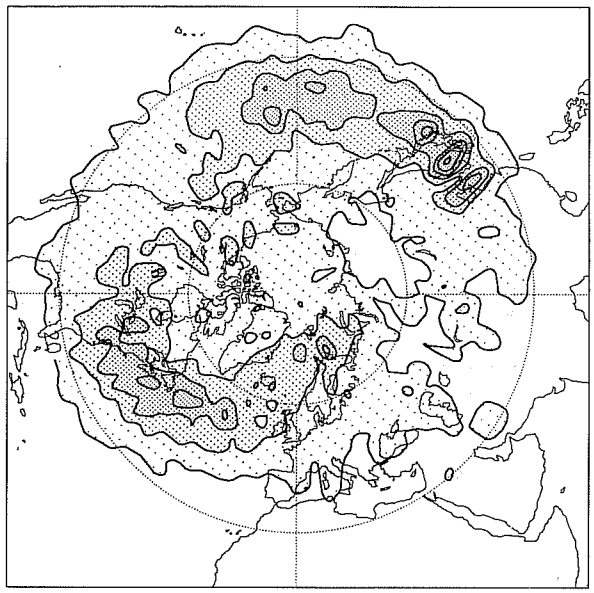
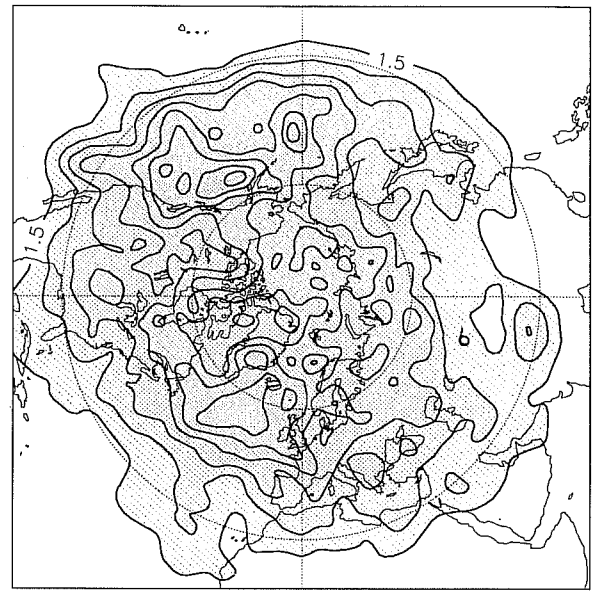


Fig 1 For summer 1994: Root mean square sensitivity gradients of temperature at model level 23 (~730 hPa) (a), root mean forecast errors for day 2 at 500 hPa (b) and the Eady index (see text for details) (c). Contour intervals are K (a), dam (b), day^{-1} (c).

a) DJF 1994/95 RMS of Temperature Sensitivity (~730 hPa)



b) D2 RMS-Error DJF 1994/95 500 hPa



c) DJF 1994/95 EADY-INDEX

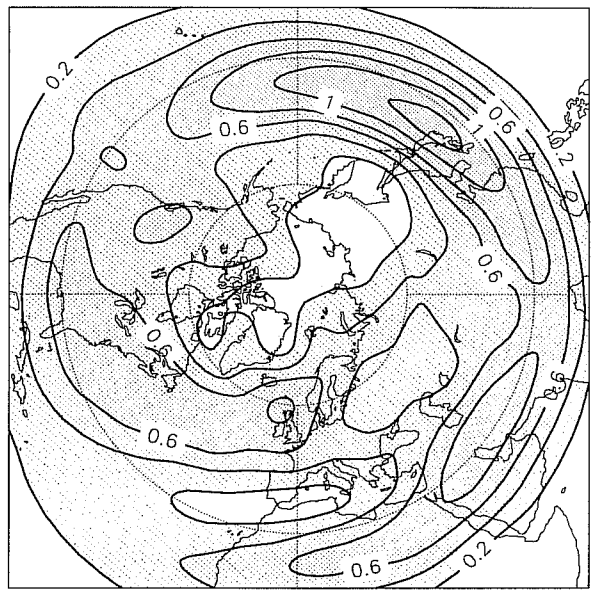


Fig 2 The same as Fig 1 except for winter 1994/95.

sensitivity is not as clear as in the Pacific. However, a large part of the sensitivity is still found well upstream of the day-2 forecast errors.

There are two reasons why only a small sensitivity can be seen in relation to the subtropical jet in winter. The atmospheric activity and day-2 rms-forecast errors are certainly fairly small here in the sub-tropics and the sensitivity calculations have been based on the day-2 forecast errors of the Northern Hemisphere limited by a southern boundary of 30 degrees north that cuts right through the middle of the subtropical jet in winter.

In the summer (Fig 1c) the major zones of large wind-shear and related baroclinic growth are found over land areas and the implied growth rates are smaller than in winter. In a consistent way large values of sensitivity can be found mainly at high latitudes, with maxima over the Rocky Mountains and North-East Canada. Some secondary maxima of sensitivity can also be seen in association with the subtropical jet that has a more northerly position in summer compared to winter.

2.2 **Influence of model change on sensitivity gradients**

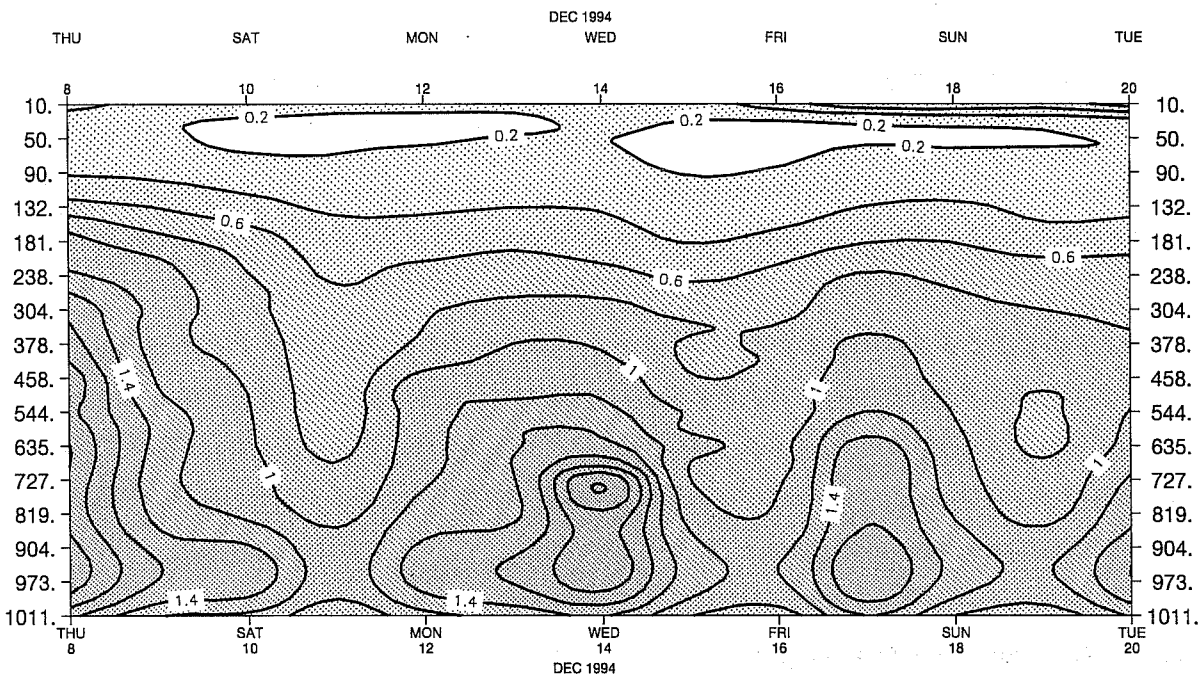
As the six hour forecast is used as a background field for analysing the observational data, it is not surprising that a change in error characteristics of the model employed in data assimilation has a noticeable influence on the final analysis. In December 1994 the prognostic cloud scheme and a modified gravity wave drag scheme were tested in parallel to the operational model version including a rerun of the data assimilation and medium-range forecasts. For this period the new system improved the forecast skill in many aspects. In particular, the short-range forecast errors for wavenumbers that describe the synoptic scale (total wavenumber larger than 8) were noticeably reduced.

The sensitivity calculations performed in parallel for the two systems showed a change in the sensitivity that is consistent with a reduction of errors in the baroclinic waves. Fig 3 a shows a time-pressure cross section for the Northern Hemisphere sensitivity (rms-values) of day-2 operational forecast errors to the initial vorticity. Days with maximum sensitivity in the operational suite were found on 8, 14, 17 and 20 December. The same maxima can be seen in the new system (Fig 3b), however there has been a general reduction of sensitivity throughout this period. In particular the low level sensitivity (below 700 hPa) is consistent with a reduction of the low level temperature errors in the short range forecast. It therefore seems that the analysis errors have been reduced by using a first guess of the low level flow that is more realistic by using the prognostic cloud scheme in the model.

3. OPTIMAL SENSITIVITY PERTURBATIONS

Results presented in the previous sections confirmed the usefulness of the sensitivity gradient. It helped to diagnose where analysis errors might produce large forecast errors. However, in *Rabier et al* (1996), it was

a) 9412 Control NH rms of vorticity grad



b) 9412 New Model NH rms of vorticity grad

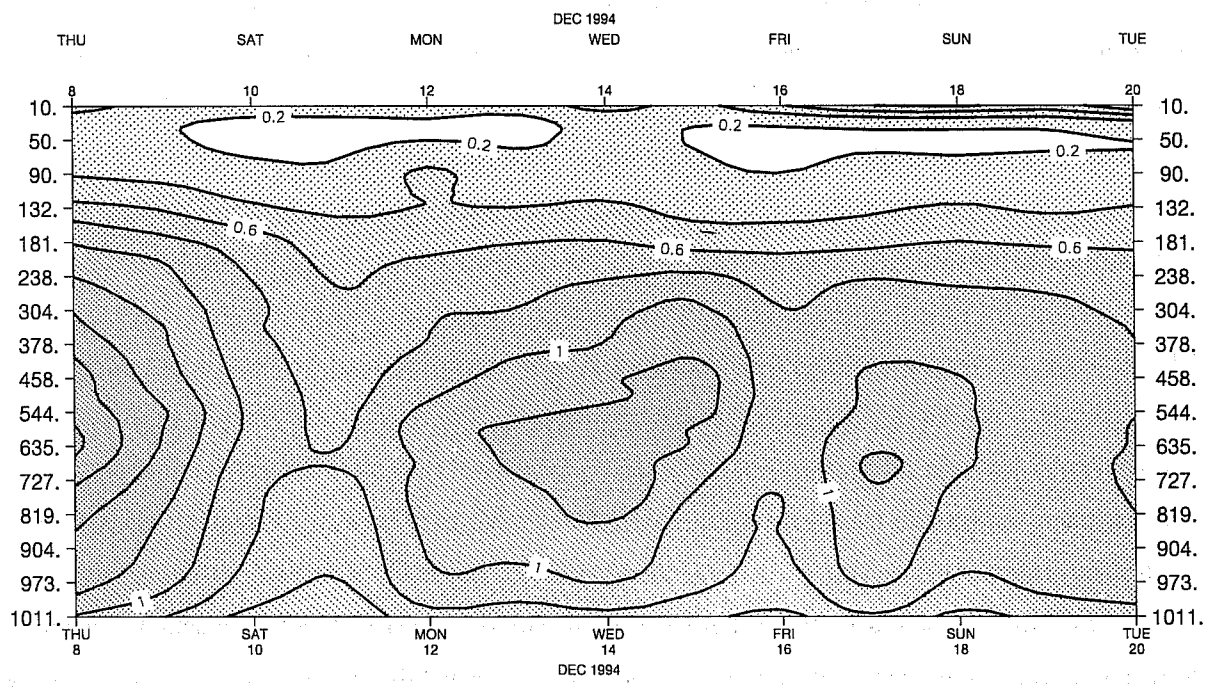


Fig 3 Time-pressure cross-section of root mean square sensitivity gradients of vorticity for the Northern Hemisphere. For the operational model and data assimilation system (a), for a system with modified physics (b).
Units: $10^{-6} s^{-1}$.

mentioned that this sensitivity gradient was dependent on the choice of inner-product, which could affect the results. This will be illustrated in the following.

3.1 Inner-product choice

One can formulate the sensitivity problem in optimisation terms. Let C be the matrix defining a relevant inner-product $\langle x,y\rangle_c = x^T C y$. This inner-product also defines a norm $\|x\|_c^2 = x^T C x$ for any field x.

By definition, perturbations of the form $\delta x_a = \alpha \nabla J_c$ where α is a scalar, and ∇J_c is the gradient of J with respect to the inner-product $\langle .,. \rangle_c$, optimize the first order change in J for a given norm of the perturbation $\|\delta x^a\|_c^2 = \mathbb{N}$ (see *Bouttier*,1993 for another derivation of this property).

Writing the diagnostic function

$$J(x^a) = \tfrac{1}{2}(M(x^a)-x_T^{ref})^T A(M(x^a)-x_T^{ref})$$

where x^a is the analysis at time 0, M represents the integration of the operational model for up to time T=2 days, x_T^{ref} is our reference atmospheric state at time T taken as the operational analysis and A is a matrix defining an inner-product. The approximate first-order change with respect to $\|\delta x^a\|$ in J is $\delta J = (M'_s.\delta x^a)^T A(M(x^a)-x_T^{ref})$ where M'_s is the tangent-linear model of a simplified model M_s. This simplified model is a lower resolution model (T63L31), run with reduced physics (*Buizza*, 1993). δx^a is an increment to the analysis. The problem of maximising δJ under the constraint $\|\delta x^a\|_c^2 = \mathbb{N}$ can be written in terms of the Lagrangian L:

$$L = (M'_s.\delta x^a)^T A(M(x^a)-x_T^{ref}) - \lambda(\delta x^{aT} C \delta x^a - \mathbb{N})$$

where λ is a scalar lagrangian multiplier. Setting the spatial derivatives of L with respect to δx^a and λ to zero yields

$$\begin{cases} M'^T_s A(M(x^a)-x_T^{ref}) - 2\lambda C \delta x^a = 0 \\ \delta x^{aT} C \delta x^a = \mathbb{N} \end{cases}$$

The second equation is simply the constraint, while the first equation can be rewritten as

$$\delta x^a = \frac{1}{2\lambda} C^{-1} M'^T_s A(M(x^a)-x_T^{ref}) = \frac{1}{2\lambda} \nabla J_c$$

A perturbation proportional to the gradient will then clearly solve the optimization of δJ under the fixed-norm constraint. The choice of inner-product will affect the result of this optimization problem, as it enters the definition of the constraint.

From a statistical point of view, the ideal inner-product should reflect the uncertainties in the analysis. However, this is not easily available and some simpler inner-products have to be considered for the time being. The inner-product usually used at ECMWF (for sensitivity studies and singular vector computations) is what is called the "energy" inner-product, which is proportional to u^2+v^2 for the wind field, and T^2 and $(lnp_s)^2$ for the mass field which is a fair first approximation to the ideal one, as discussed in *Molteni et al* (1996). Of course, this "energy" inner-product cannot reflect the dependency of analysis error on either the particular meteorological situation or the geographical location for instance, but it is thought to be globally relevant for the isotropic part of the errors. By using the inner-product defined by the analysis error covariance matrix, one would give a relevant weight to all components of the perturbations, taking into account their likelihood. This inner-product is known to give very small-scale sensitivity structures (see *Rabier et al,* 1996). The question which might be asked is how these structures would be changed, by using a different inner-product. It was chosen to answer this question by selecting an "enstrophy"-like inner-product , which is proportional to ζ^2+D^2 for the wind field. All coefficients of the matrix A defining the "energy" inner-product were multiplied by $n\times(n+1)$, where n is the total wave-number to get a B matrix defining the "enstrophy" inner-product. Going from the "energy" gradient to the "enstrophy" one then amounts to dividing all components by $n(n+1)$ because of the relation $\nabla J_B = B^{-1} A \nabla J_A$. The gradient obtained in such a way will then be much more large-scale. For diagnostic purposes, one can scale the gradients obtained with both inner-products and add the following perturbation field as a perturbation to the analysis. Figs 4 and 5 show the time evolution of the increments based on these gradients for the 6 April 1994 case discussed in *Rabier et al* (1996). The scaling factor was chosen to give a good improvement in 2-day forecasts. The initial increments (panel a) are much broader for the "enstrophy" inner-product calculation (Fig 5) than for the "energy" inner-product calculation (Fig 4). One can also notice that the amplitudes are much larger with an rms value of 8.37 m compared with an rms value of 0.41 m for the "energy" increment.

However, these perturbations of different scale evolve to a similar structure after an integration of 48-hours (panel c, in Figs 4 and 5). The large-scale "enstrophy" perturbations also improve the quality of the forecast to the same extent as the "energy" perturbations, as can be seen in Fig 6. Note the larger rms differences from the operational OI, for the "enstrophy" increments in panel b (Fig 6).

Although both initial perturbations are "optimal" for the sensitivity problem under investigation (with respect to different norms) and they both considerably improve the forecast, one might prefer to investigate the first one more for diagnostic studies. As a matter of fact, if this type of sensitivity study is also aimed at hinting at

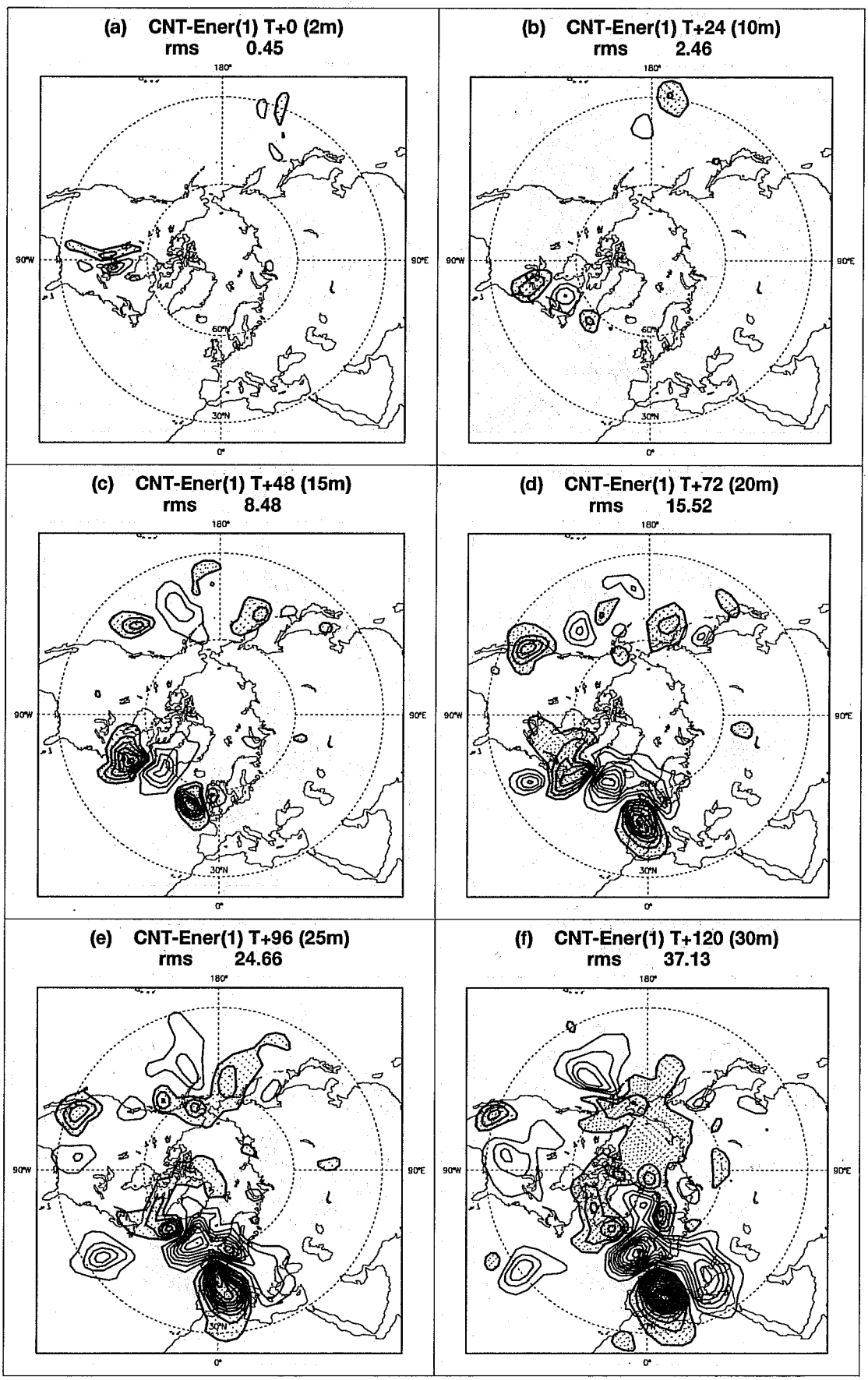


Fig 4 Difference between the sensitivity integration started from the analysis perturbed by a scaled gradient with respect to energy inner-product and the control forecast for forecast ranges day 0, day 1, ..., day 5 for 6 April 1994. Units: gpm. Note the different contour intervals used in the panels (shown in brackets): day 0, 2 m; day 1, 10 m; day 2, 15 m; day 3, 20 m; day 4, 25 m; day 5, 30 m.

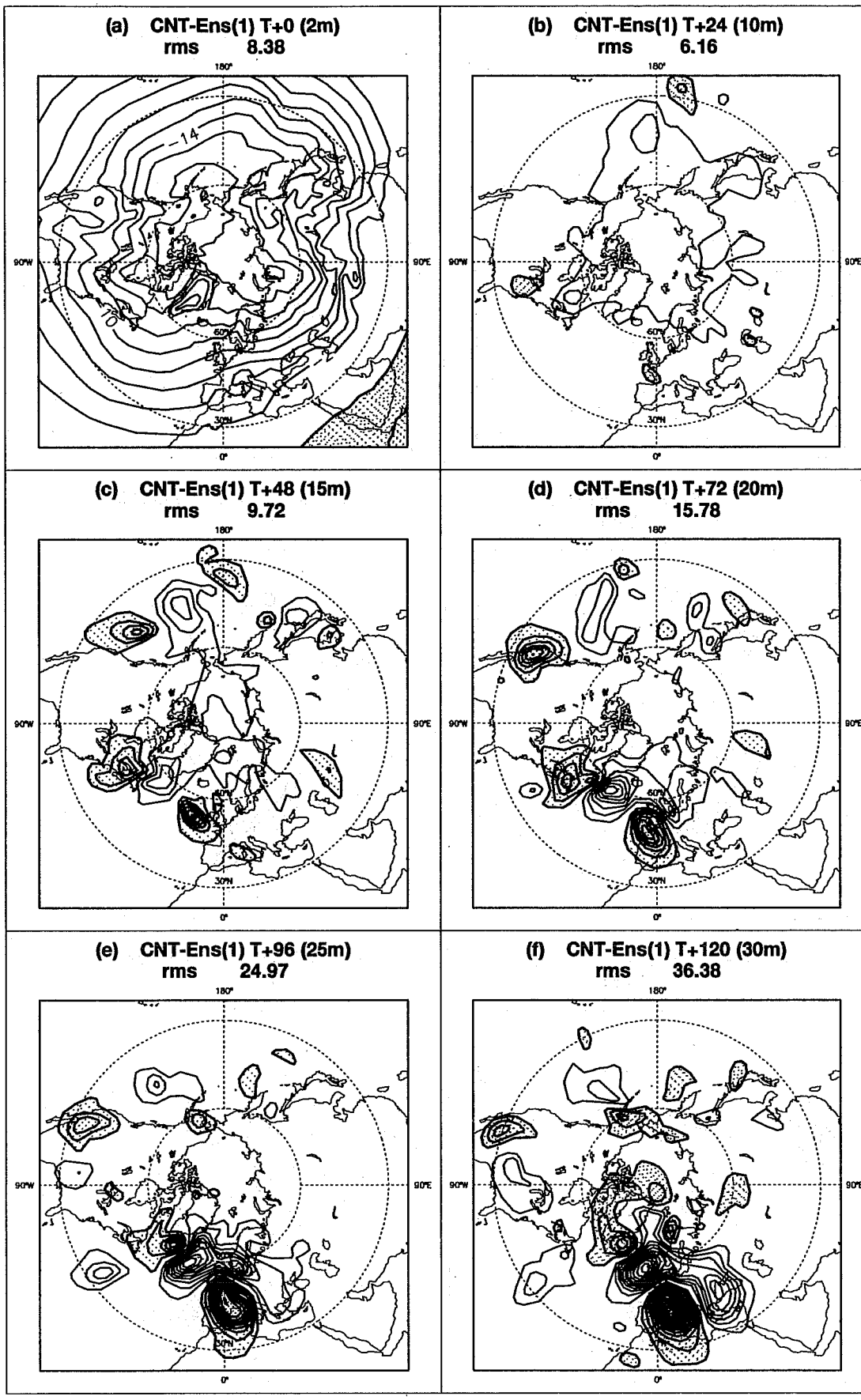


Fig 5 Difference between the sensitivity integration started from the analysis perturbed by a scaled gradient with respect to enstrophy inner-product and the control forecast for forecast ranges day 0, day 1, ..., day 5 for 6 April 1994. Units: gpm. Note the different contour intervals used in the panels (shown in brackets): day 0, 2 m; day 1, 10 m; day 2, 15 m; day 3, 20 m; day 4, 25 m; day 5, 30 m.

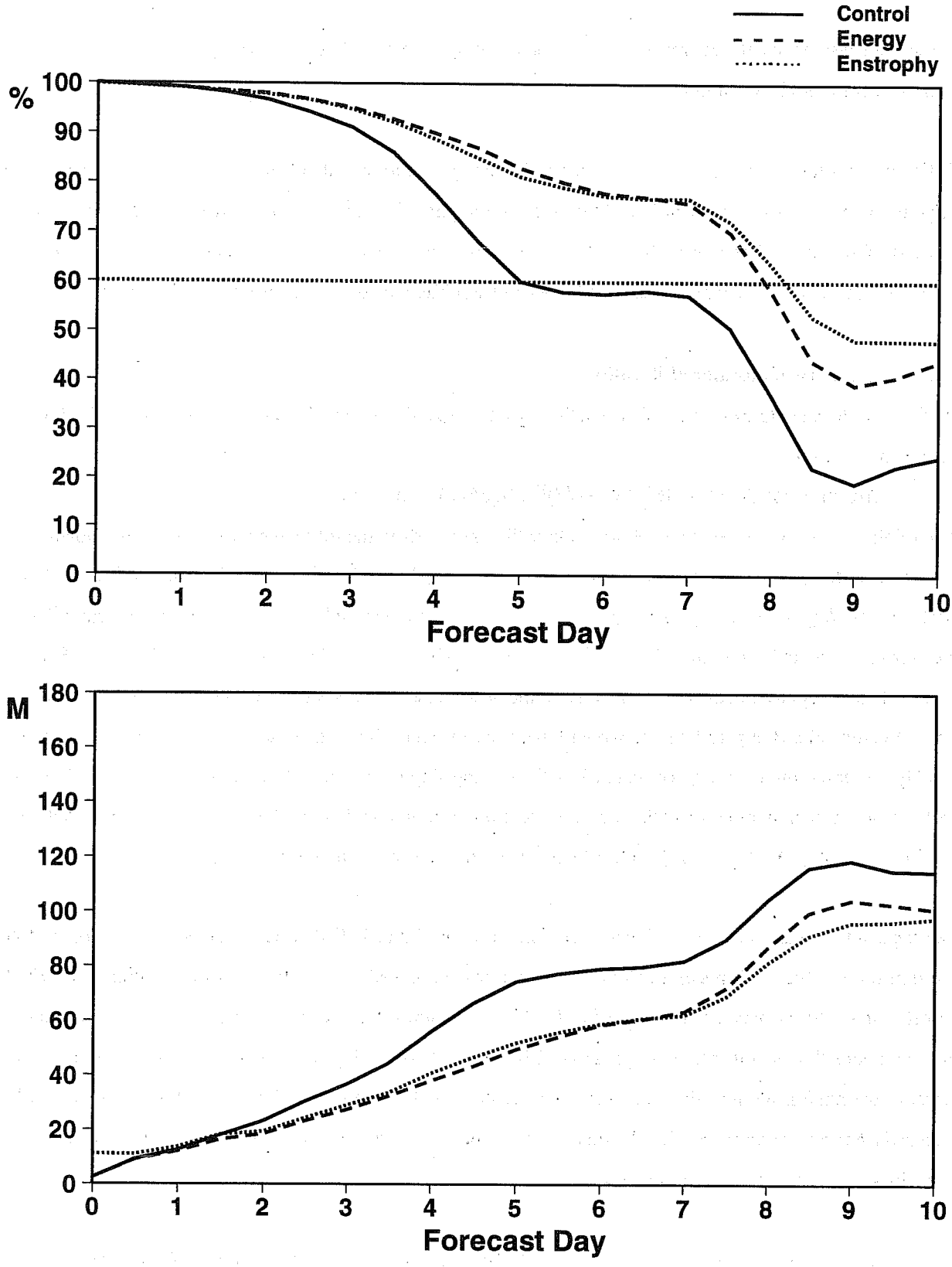


Fig 6 Forecast scores for 6 April 1994 (anomaly correlation and root-mean-square error for the Northern Hemisphere). Scores are shown for the control forecast (solid line), the energy-based gradient sensitivity integration (dashed line) and the enstrophy-based gradient sensitivity integration (dotted line).

possible analysis errors, the energy-based increment is much more likely to represent part of this analysis error both in terms of scale and amplitude.

Following this line of thought, the sensitivity problem will be further extended: let us now investigate a range of perturbations of the analysis which might both represent part of the analysis error fairly well and improve the forecast substantially. These perturbations might no longer be "the most sensitive" perturbations, but will still be called "sensitive" if they grow reasonably fast and significantly affect the forecast performance.

3.2 **Influence of number of iterations**

A slightly different optimization problem will then be addressed: find a perturbation δx^a minimizing the forecast error squared norm

$$J(\delta x^a) = \tfrac{1}{2}(M(x^a) + M'_s.\delta x^a - x_T^{ref})^T A(M(x^a) + M'_s.\delta x^a - x_T^{ref}).$$

The minimization, using the Quasi-Newton algorithm, proceeds in the following way: the initial gradient of J with respect to the inner-product $<.,.>_A$ is $(\nabla J)_o = A^{-1} M'^T_s A(M(x^a) - x_T^{ref})$. The first step will be to find a perturbation $\delta x_o^a = -\alpha_o(\nabla J)_o$ bringing a decrease in J. The new value of J and of its gradient $(\nabla J)_1$ is recomputed at the updated point $x^a + \delta x_o^a$. The next step is then to find a perturbation $\delta x_1^a = -\alpha_1 \, H_1(\nabla J)_1$ where H_1 is an approximation of the inverse of the hessian (second derivative) of the cost-function J, and so on ... In Appendix B, a parallel is drawn with the Lanczos algorithm, which illustrates how the minimization quickly converges towards the space spanned by the leading singular vectors of the tangent linear model. If one performs a sufficient number of such iterations, the minimum of J will be reached. This minimum is obtained for $\delta x^a = -(M'^T_s \, A \, M'_s)^{-1} \, M'^T_s \, A(M(x^a) - x_T^{ref})$ (in the case of a non-singular M'_s).

Ten steps of the minimization algorithm were performed in the 6 April 1994 case, which is not sufficient to find the minimum. Figure 7 (panel a), shows the increments produced after 3 steps of minimization while Fig 8 (panel a) shows the increments produced after 10 steps of minimization. Comparing panel a in Figs 4, 7 and 8, one can notice that new increments appear at different locations during the course of the minimization. The effect of the minimization is also to modify the amplitude of the increments. However, it does not drastically change the horizontal scales associated with these increments. After a few days of integration the three sets of perturbations evolve into a similar pattern.

In terms of forecast skill, performing a few steps of the minimization algorithm provides slightly better forecasts (see Fig 9). It then seems beneficial to perform a few steps of the minimization algorithm in order to describe more accurately certain components of the analysis error. A similar approach was used in *Zupanski* (1995). After 10 iterations his resulting estimate of the analysis improved the forecast substantially but was clearly too large to be realistic. The question is then which number of iterations would be desirable. As the purpose is to find some "sensitive", fast-growing components of the analysis error, three iterations would be a relevant choice

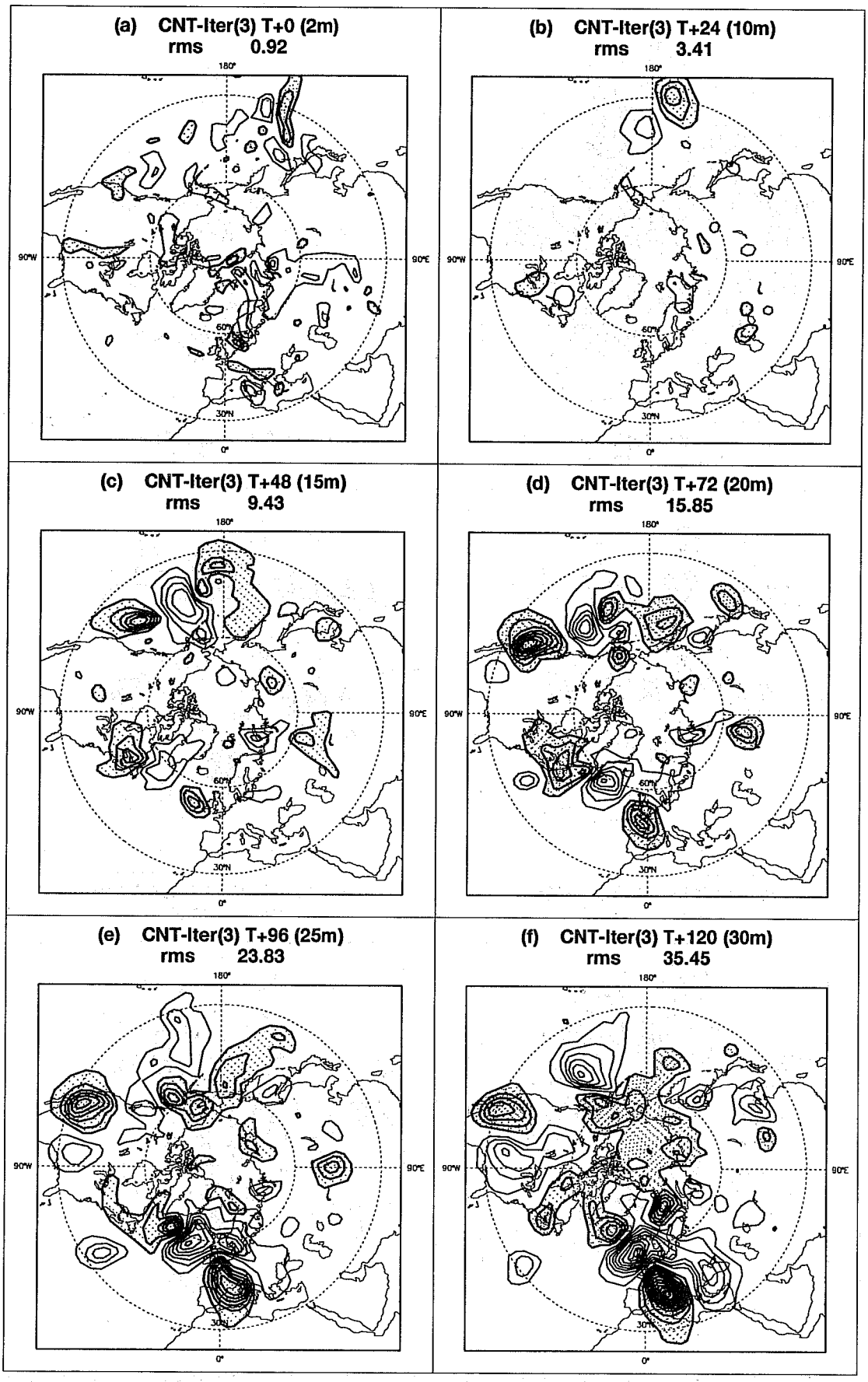


Fig 7 Difference between the sensitivity integration started from the analysis perturbed by the increment based on 3 iterations and the control forecast for forecast ranges day 0, day 1, ..., day 5 for 6 April 1994. Units: gpm. Note the different contour intervals used in the panels (shown in brackets): day 0, 2 m; day 1, 10 m; day 2, 15 m; day 3, 20 m; day 4, 25 m; day 5, 30 m.

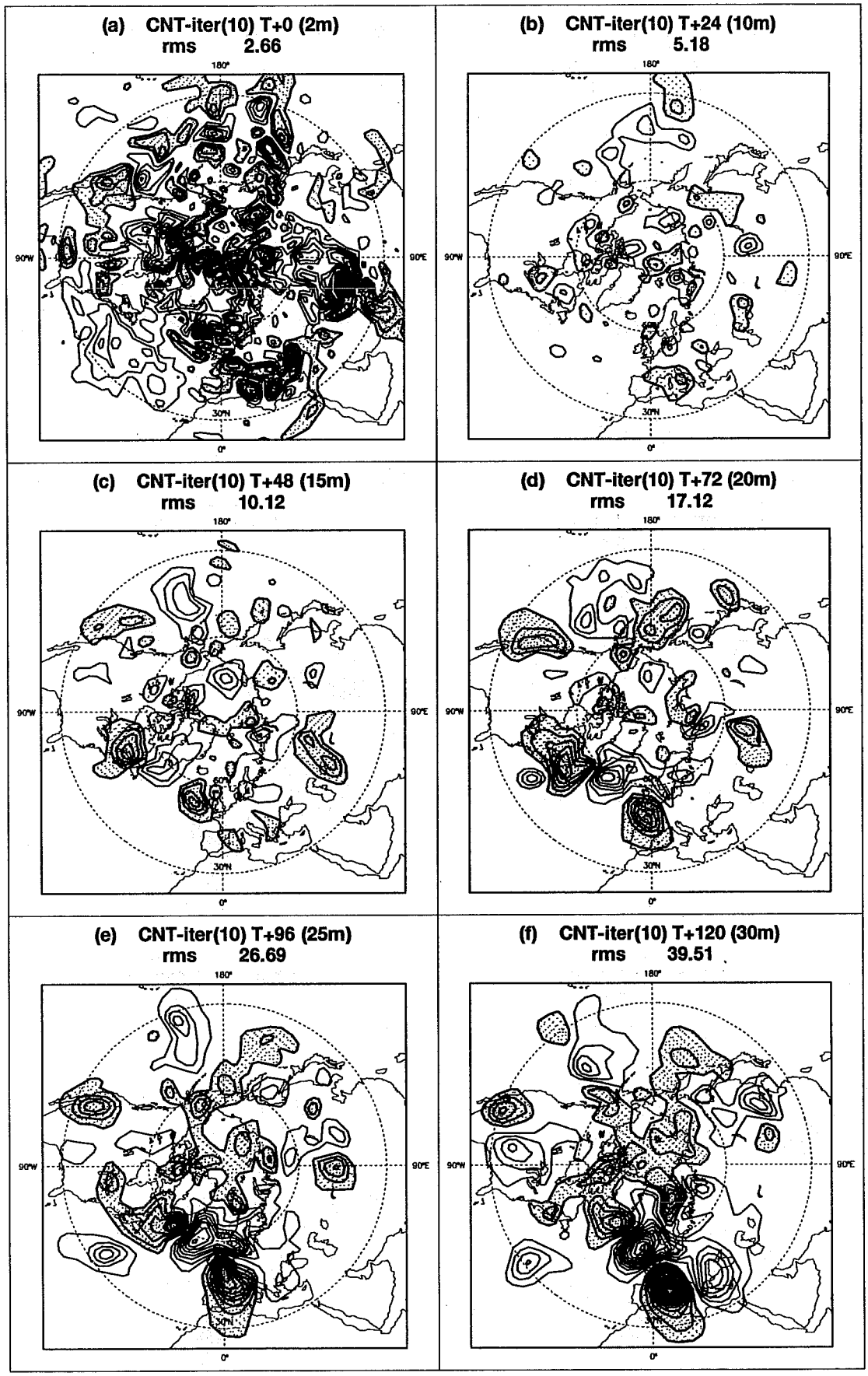


Fig 8 Difference between the sensitivity integration started from the analysis perturbed by the increment based on 10 iterations and the control forecast for forecast ranges day 0, day 1, ..., day 5 for 6 April 1994. Units: gpm. Note the different contour intervals used in the panels (shown in brackets): day 0, 2 m; day 1, 10 m; day 2, 15 m; day 3, 20 m; day 4, 25 m; day 5, 30 m.

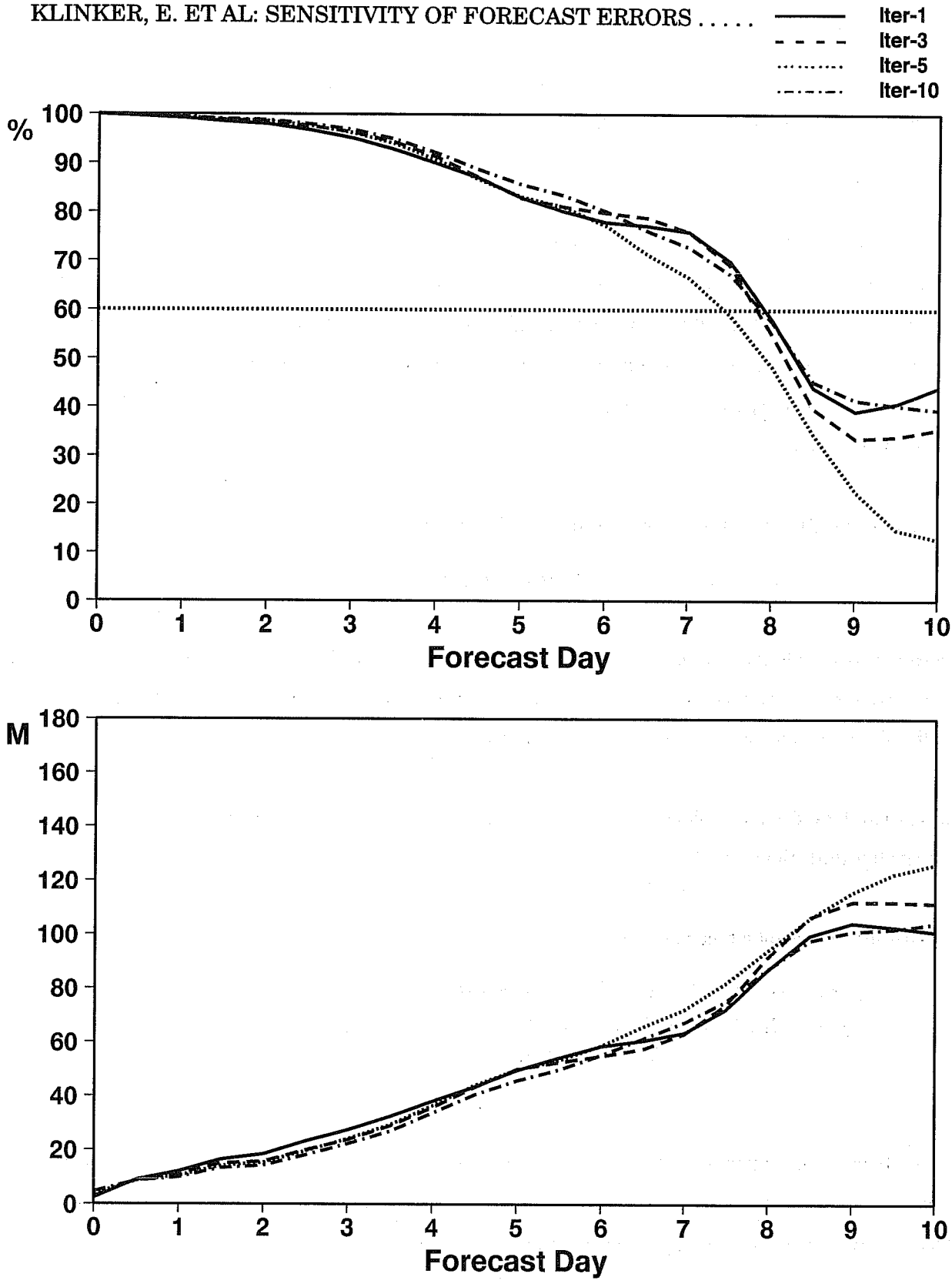


Fig 9 Forecast scores for 6 April 1994 (anomaly correlation and root-mean square error for Northern Hemisphere). Scores are shown for the sensitivity integrations based on the first iteration of the minimization (equivalent to energy-based scaled gradient, solid line) the third iteration (dashed line), the fifth iteration (dotted line) and the tenth iteration (dash-dotted line).

in our case, with a ten-fold increase in rms for the perturbation during the first 48 hours of the forecast. In comparison, the first iteration and fifth iteration perturbations were growing respectively by a factor of 20 and 5 during the same period. Some justification of this choice will be discussed in the following sections.

4. COMPARISON WITH OBSERVATIONS

The fit to the observational data is one criterion which can be used to discriminate between different perturbations to be added to the analysis.For any given perturbation δx^a the goal is to compare the likelihood of two atmospheric states: the (initialised) analysis x^a and the modified analysis $x^a+\delta x^a$. The fit to the available data will be evaluated for atmospheric states of the form $x^a+\alpha\delta x^a$, which is for points in the direction of the perturbation δx^a.

For any α, the fit to the observations y is measured by the quantity

$$J_o(\alpha) = (H(x^a+\alpha\delta x^a) - y)^T R^{-1}(H(x^a+\alpha\delta x^a)-y)$$

where H is the observation operator projecting a model state onto the observation vector y and R is the observation error covariance matrix. For more details on the computation of this cost-function currently used in the 3D-Var system at ECMWF see *Andersson et al* (1995).

It is assumed that the tangent-linear hypothesis is valid, which means that $H(x^a+\alpha\delta x^a) \simeq H(x^a) + \alpha H'.\delta x^a$ for small perturbations $\alpha\delta x^a$.

$J_o(\alpha)$ is then a quadratic function of α:

$$J_o(\alpha) = (H(x^a)+\alpha H'.\delta x^a-y)^T R^{-1}(H(x^a)+\alpha H'.\delta x^a-y)$$
$$J_o(\alpha) = (H(x^a)-y)^T R^{-1}(H(x^a)-y)+2\alpha(H'.\delta x^a)^T R^{-1}(H(x^a)-y) + \alpha^2(H'.\delta x^a)^T R^{-1}H'.\delta x^a$$

Let us denote α_0 the minimum of this parabola . The atmospheric states projecting on the line with origin x^a and direction δx^a with a projection of the form $x^a + \alpha_0\ \delta x^a$ are optimizing the fit to the observations along this line. α_0 can be found by expressing a condition of nil derivative $J'_o(\alpha_0) = 0$

$$\delta x^{aT} H'^T R^{-1}(H(x^a)-y+\alpha_0\ H'.\delta x^a) = 0$$
$$(\delta x^{aT} H'^T R^{-1}\ H'\delta x^a)\alpha_0 = -\delta x^{aT} H'^T R^{-1}(H(x^a)-y)$$

As α_0 is expressed as a function of the random variable y, its variance can be computed. If one denotes α_t the coefficient describing the projection of the true atmospheric state x^t onto the line with origin x^a and direction

δx^a, one obtains for the variance of α_0:

$$\sigma_0^2 = \overline{(\alpha_0 - \alpha_t)^2} = (\delta x^{aT} H'^T R^{-1} H' \delta x^a)^{-1}$$

This means that the larger the curvature of the parabola, the more accurate the location of the state best fitting the observation in that direction. The parabola $J_o(\alpha)$, through its minimum and its curvature, contains useful information for comparing the two atmospheric states we are interested in: x^a and $x^a + \delta x^a$. The way this parabola can be built is to compute J_o for a few α's (three values are enough) and then fit a parabola between these values.

The fit to the observations for the analysis x^a is given by $J_o(\alpha=O) = 41512.1$, taking into account all the observations over the globe on the 6-hour window centred on 940406, 12Z.

Table 1 indicates, for each relevant direction δx^a, the fit to the observations for $x^a + \delta x^a$ (i.e. $\alpha=1$), the coefficient α_0 such that $x^a + \alpha_0 \delta x^a$ minimizes the fit to the observations in the direction δx^a, and σ_0 the standard-deviation of error associated with α_0.

δx^a	enstrophy <..> scaled gradient	energy <..> scaled gradient	energy <..> 3 iterations	energy <..> 5 iterations	energy <..> 10 iterations
$J_o(\alpha=0)$	41512.1	41512.1	41512.1	41512.1	41512.1
$J_o(\alpha=1)$	44087.8	41514.3	41399.5	41409.1	41733.7
α_0 optimal	-0.06	0.48	1.02	0.61	0.39
σ_0	0.02	0.14	0.10	0.05	0.03
$\alpha=0$ (Analysis) "fits the data"	yes	yes	no	no	no
$\alpha=1$ (Increment) "fits the data"	no	no	yes	no	no

Table 1: Fit of different atmospheric states to observations.

Comparing only the different J_o's obtained in this table, one notices two major points. The first one is that the scaled gradient obtained with the "energy" inner-product fits the observations better than the one obtained with the "enstrophy" inner-product. The second point is that, among the increments presented in this table, only two exhibit a better fit to the data than the control $J_o(\alpha=O)$: the increments obtained after 3 and 5 iterations, the best results being obtained after 3 iterations.

One can consider the (arbitrary) criterion: x "fits" the observations if it is less than 4 standard-deviations away from the state best fitting the observations. The last two lines in the table show when the analysis and modified analyses meet this criterion. Then, along the 5 directions studied, the analysis x^a represented by $\alpha=0$ fits the observations only twice. This is not unreasonable, knowing that the analysis was not an accurate description of the atmospheric state, as can be judged by the quality of the subsequent forecast. As far as the increments are concerned, it is only for 3 iterations that the increment fits the observations (although the first iteration was on the edge). In this particular direction (3rd column in the table), the increment almost exactly fits the state best representing the data, whereas the analysis is more than 10 standard-deviations away.

The conclusion of these computations is that, among the perturbations computed in this study, the one corresponding to three iterations of the minimization algorithm is the only one moving the analysis significantly closer to the observations. Consequently further experimentation with this choice of sensitive increments has been performed in the following.

5. LIMITATIONS IN THE DERIVATION AND USE OF SENSITIVITY PERTURBATIONS

5.1 **Influence of the simplified low-resolution model on the quality of the increments**

The limits of the method are investigated in a simplified situation in which it is assumed that the analysis error is known. It will then be possible to judge the quality of the increments produced by the minimization method.

The operational analysis is supposed to be the "true" atmospheric state x^t for 20 September 1994. Another analysis x^a is obtained by performing the optimal interpolation on the same date, but removing one observation (an AIREP in the middle of the Pacific Ocean). Furthermore the operational model M is assumed to be perfect, so that the divergence of the forecasts started from x^a and x^t actually measures the forecast error as an evolution of the analysis error.

Figure 10 shows the "analysis error" x^a-x^t for the stream function field at model level 13 (roughly 300 hPa). One clearly sees the impact of removing the observation located at 180W, 40N. After 2 days, the forecast error $M(x^a)-M(x^t)$ is shown in Fig 11 for the same field. The error pattern has moved eastwards and northwards towards the North-Eastern Pacific and Western Canada.

As already explained in the previous sections, the minimization of the forecast error is done with a simplified physics low-resolution model denoted M_s. The cost function in this case writes $J_M(\delta x^a) = \langle M(x^a)+M'_S\,.\delta x^a-M(x^t),\ M(x^a)+M'_s\,.\delta x^a-M(x^t)\rangle$ where M'_s denotes the tangent-linear of M_s. Figure 12 shows the increments obtained after three iterations. These should be compared to the analysis error in Fig 10. The pattern of the analysis error, with a dipole located at 40N, 180W is reproduced fairly well by the increments. However, the horizontal scale of the increments is too large and there is a spurious signal over

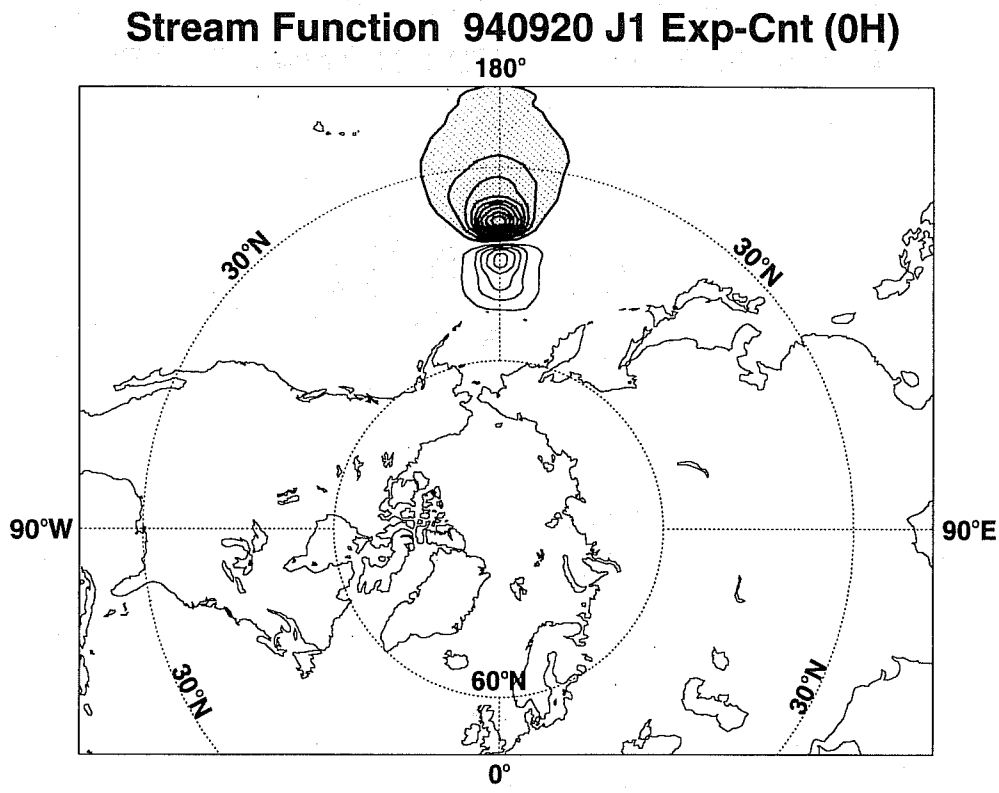


Fig 10 Difference between the analysis $\mathbf{x}^a$ and the "true" atmospheric state $\mathbf{x}^t$, for the streamfunction field at model level 13 (roughly 300 hPa). The contour interval is 5×10^5 $m^2 s^{-1}$.

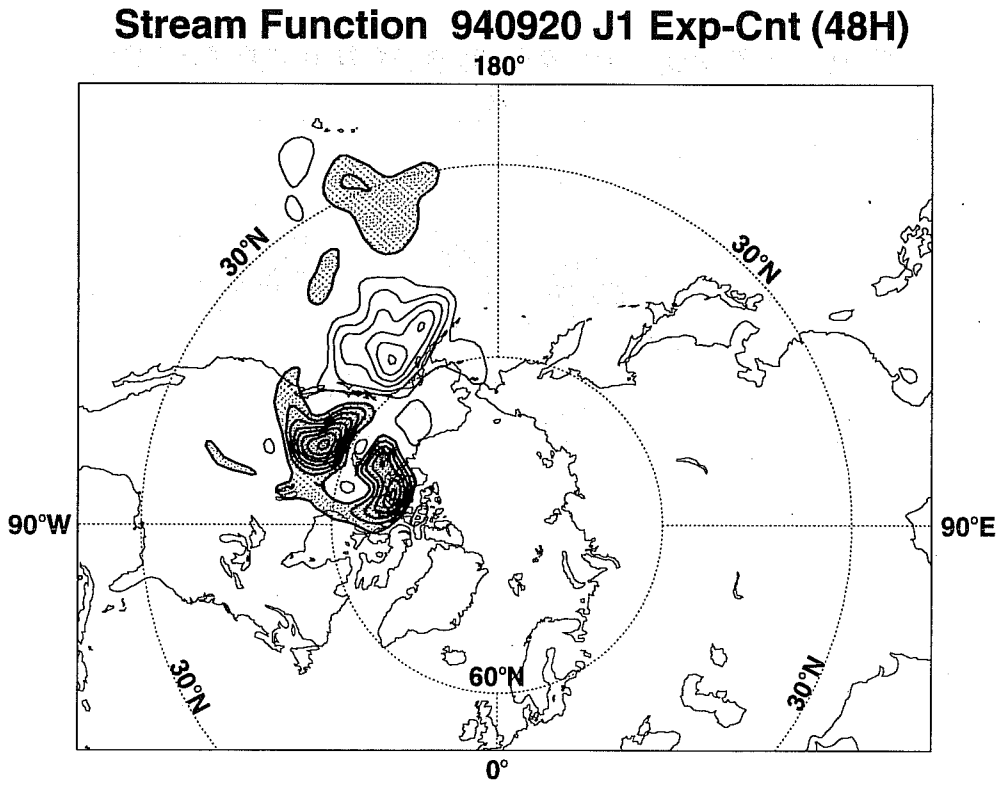


Fig 11 Difference between the forecast started from the analysis $\mathbf{x}^a$ and the forecast started from the "true" atmospheric state $\mathbf{x}^t$, after 48 hour integration of the operational T213L31 model, for the streamfunction field at model level 13 (roughly 300 hPa). The contour interval is 5×10^5 $m^2 s^{-1}$.

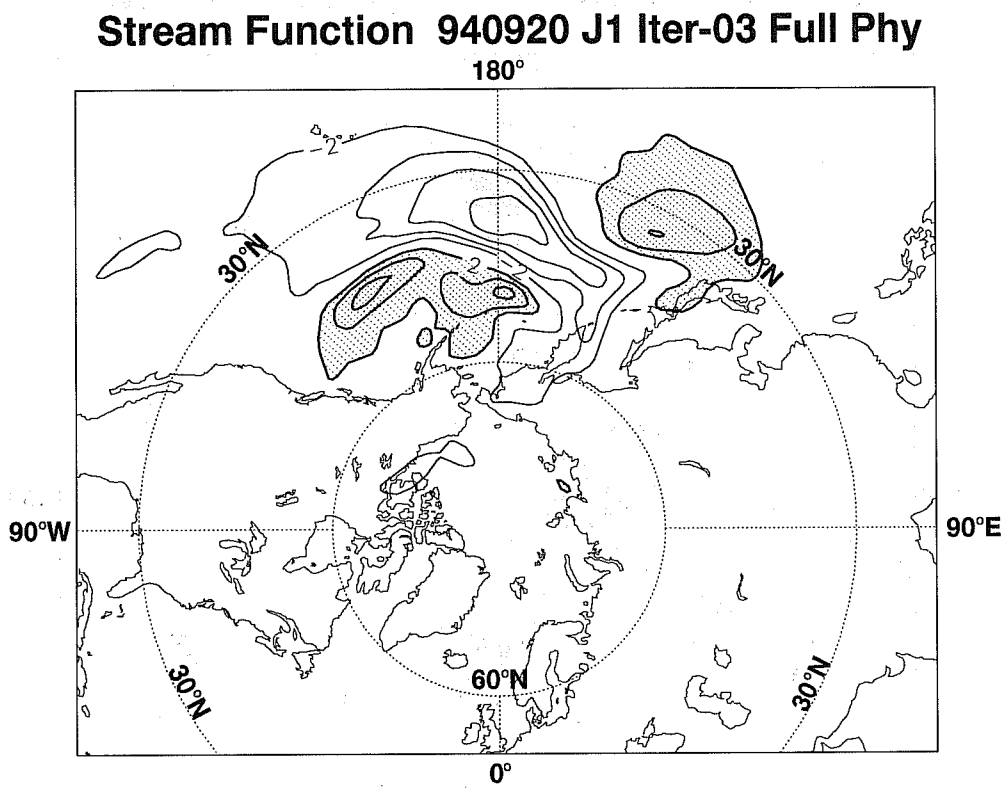


Fig 12 Increments obtained after three iterations of the minimization of the two-day forecast error obtained with the full T213L31 model started from the analysis $\mathbf{x}^a$, for the streamfunction field at model level 13 (roughly 300 hPa). The contour interval is 2×10^4 $m^2 s^{-1}$.

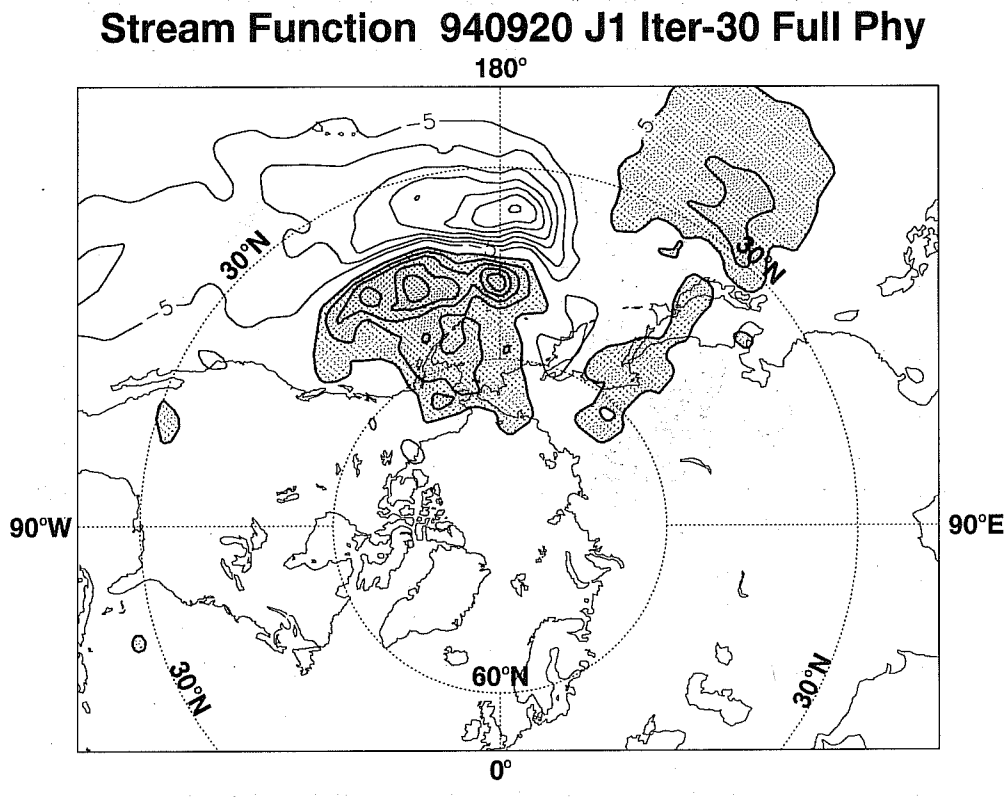


Fig 13 Increments obtained after 30 iterations of the minimization of the two-day forecast error obtained with the full T213L31 model started from the analysis $\mathbf{x}^a$, for the streamfunction field at model level 13 (roughly 300 hPa). The contour interval is 5×10^4 $m^2 s^{-1}$.

the Western Pacific. One should note the size of the increments, which is roughly a tenth of the actual analysis error. This result after three iterations looks like a good estimate of part of the analysis error. Proceeding with the minimization produces increments with a larger amplitude as can be seen from Fig 13, after 30 iterations. However, the noise does not disappear and its amplitude increases. This is clearly detrimental for extracting the relevant signal out of the structure of the increments.

The influence of using M_s instead of M in the minimization on the quality of these increments can be investigated as follows. Let us now assume that the simplified model M_s is perfect. It is then the divergence of the forecasts started from x^a and x^t, and using the model M_s which measures the forecast error. This forecast error is shown in Fig 14. One can compare the actual forecast error produced with the full T213 physics model M in Fig 11, with that obtained with the T63 simplified physics model M_s in Fig 14. The high resolution model clearly gives a larger forecast error than the low resolution one. However, the structure of the forecast error is quite similar, apart from some large positive errors over Western Canada which are only present when the high resolution model is run.

Minimizing $J_{M_s} = \langle M_s(x^a) + M'_s \,.\delta x^a - M_s(x^t),\ M_s(x^a) + M'_s \,.\delta x^a - M_s(x^t)\rangle$ is the next stage, for which no approximation is involved due to the consistent use of model version M_s obtained to define the forecast error. The increments obtained after three iterations (not shown) are quite similar to the ones obtained after three iterations of the minimization of J_M. Performing more steps of the minimization algorithm for J_{M_s} provides a better description of the analysis error. The increments obtained after 30 iterations (not shown) compare better to the actual analysis error than the results obtained after 30 iterations for the minimization of J_M. In particular, there are no spurious maxima, apart from the relevant dipole at 180W.

It then seems beneficial to perform more steps of minimization in the case when the model used in the minimization is the same as the one used to simulate the "forecast error", whereas it is not beneficial when the models are different.

The minimization of J_{M_s} was extended to 60 iterations. The result is comparable to the one obtained after 30 iterations as far as the structure of the increments is concerned. However, the amplitude comes closer to that of the analysis error (not shown). The full convergence of the minimization was not achieved after 60 iterations, which means that the δx^a such that $M'_s \,.\delta x^a = (M_s(x^a) - M_s(x_t))$, i.e. $\delta x^a = -M'^{-1}_s (M_s(x^a) - M_s(x^t))$ was not reached. In any case, inverting the M'_s over 48 hours is ill-conditioned because of the diffusion included in the model. It could not be expected to retrieve the analysis error perfectly. For a discussion of this problem, see *Thépaut and Courtier* (1991).

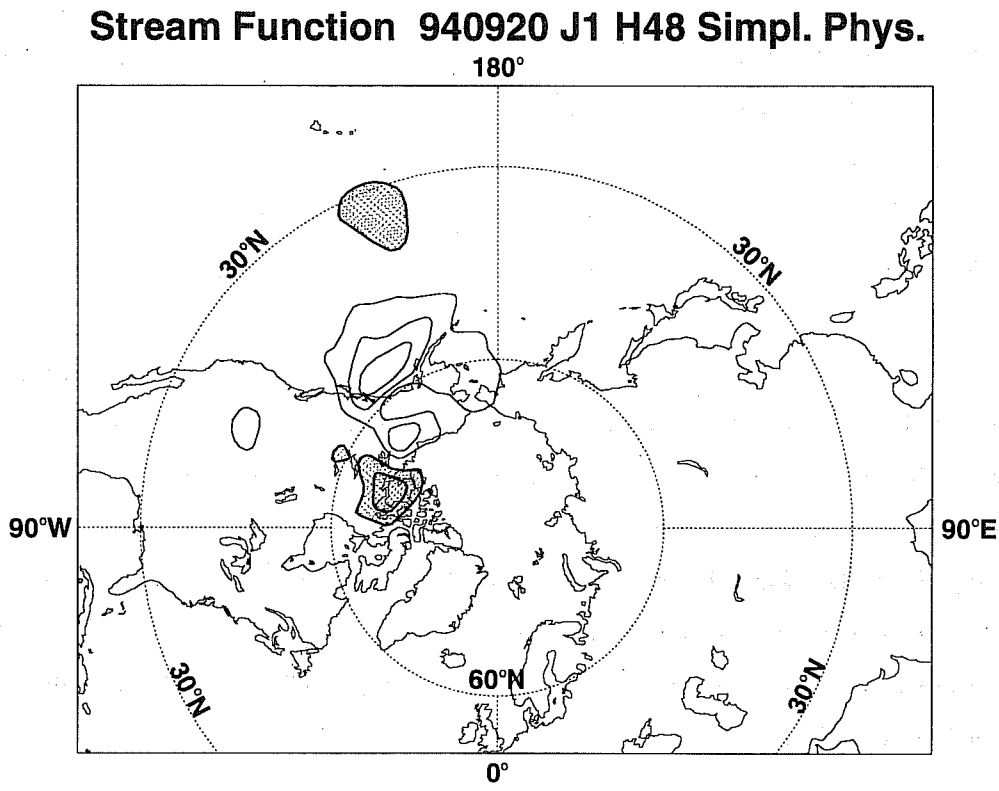


Fig 14 Difference between the forecast started from the analysis $\mathbf{x}^a$ and the forecast started from the "true" atmospheric state $\mathbf{x}^t$, after 48 hour integration of the model with simplified physics, for the streamfunction field at model level 13 (roughly 300 hPa).The contour interval is 5x10 m^2s^{-1}.

The conclusion of this section is that three iterations of the minimization algorithm provide a reasonable structure of the fast-growing part of the analysis error, even if the model used in the minimization is not perfect. However, it cannot be expected to get an accurate description of the analysis error by performing a full minimization, because of the discrepancy between the model M_s and the model M. Even if the same model was used, the full analysis error could not be retrieved because of a loss of conditioning due to diffusion in the model.

Furthermore, in a more realistic context of uncertainty in the definition of the forecast error (which is the case in practice, as it is defined as the discrepancy between the forecast and the verifying analysis, which is not perfect), information can only be extracted along the unstable directions. As a matter of fact, one can write the forecast error as $e_T = Re_o + \varepsilon_T$, where e_o is the error in the initial conditions and ε_T is the noise associated with errors in the verifying analysis. If one wants to invert this expression $\delta x_o = -R^{-1}(Re_o+\varepsilon_T)$ to find the perturbation in the initial conditions cancelling the forecast error, one clearly amplifies the noise in contracting directions of the tangent-linear operator. For an eigenvalue of 1 (neutral direction, neither stable nor unstable), the uncertainty in the estimate of the analysis error δx_o will be the same as the uncertainty ε_T in the forecast error.

5.2 Linear and non-linear error evolution

The sensitivity calculations are performed under the assumption that the first 48 hours of forecast time are dominated by linear growth of relatively small analysis errors. It has been shown that this is a reasonable assumption. This was shown in the case of a barotropic model by *Lacarra and Talagrand* (1986), in the context of large-scale baroclinic instability by *Rabier and Courtier* (1992) and in the framework of a mesoscale model by *Errico et al* (1993b). One way of verifying the linear behaviour is to follow the growth of a pair of small perturbations that are identical apart from the sign. After 48 hours of integration time (top row of panels in Fig 17) the negative perturbations, plotted with the sign reversed (left panel), have evolved into a structure that is almost identical to the structure of the evolved positive perturbations (middle panel). The correlation between the day-2 forecast errors (right panel) and the evolved perturbations is fairly high, indicating that a large part of the forecast errors can be adequately described by the unstable growth of analysis errors.

The linear evolution of the perturbations continues for at least another day of integration time until, by day-4, (middle row of panels in Fig 15) differences between the negative and positive perturbation growth show up. In particular the evolution of the positive perturbation (middle panel) over the East-Atlantic into a strong negative deviation from the control is a much weaker feature in the negative perturbation run (left panel). In the North-East Pacific the linear growth continues for a few more days and is still present by day 6 (bottom row of panels in Fig 15). However, west of the North-American coast the evolved perturbations at day-6 do not contribute to a consistent reduction of forecast errors in that region as perturbations and forecast errors have

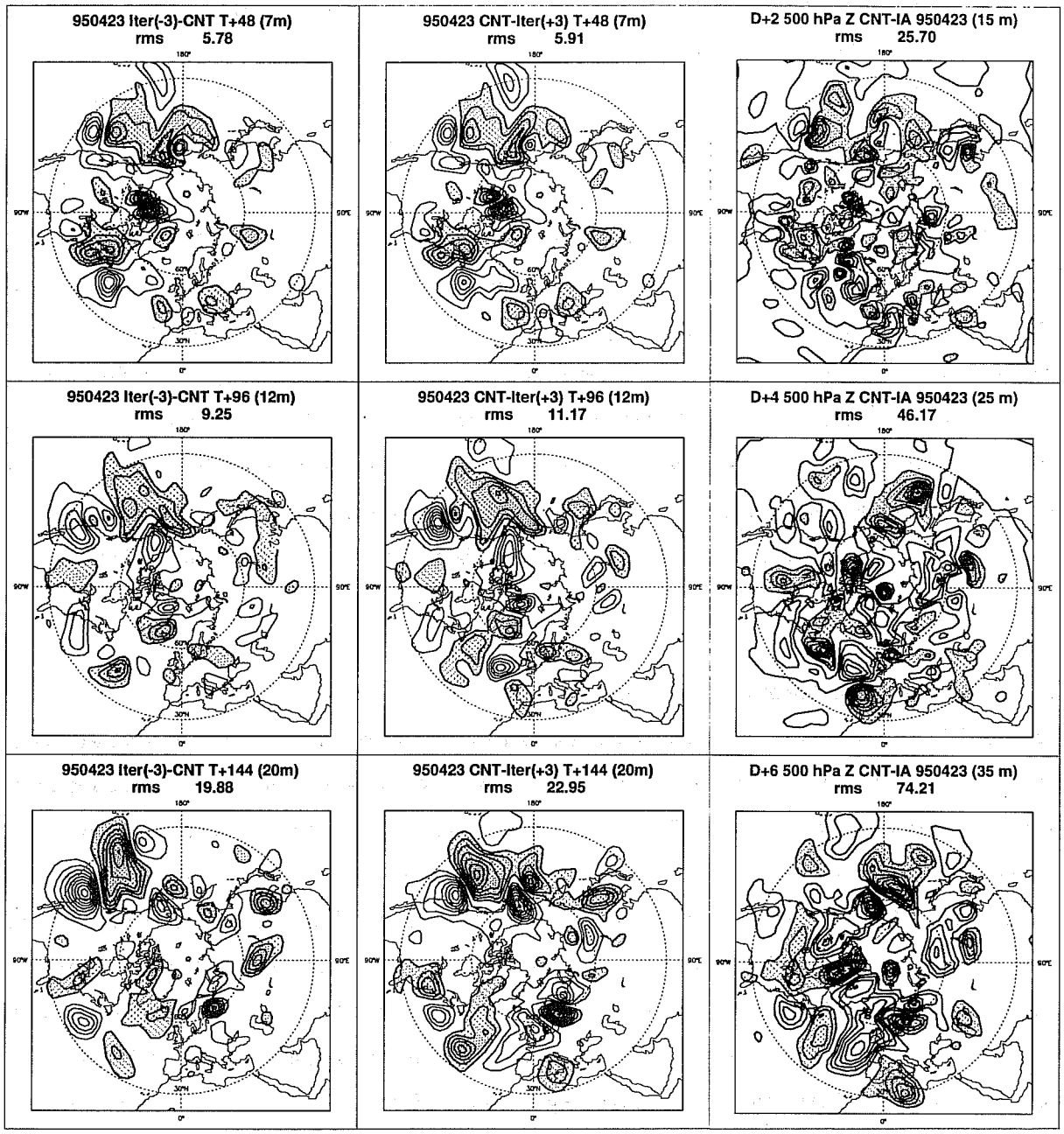
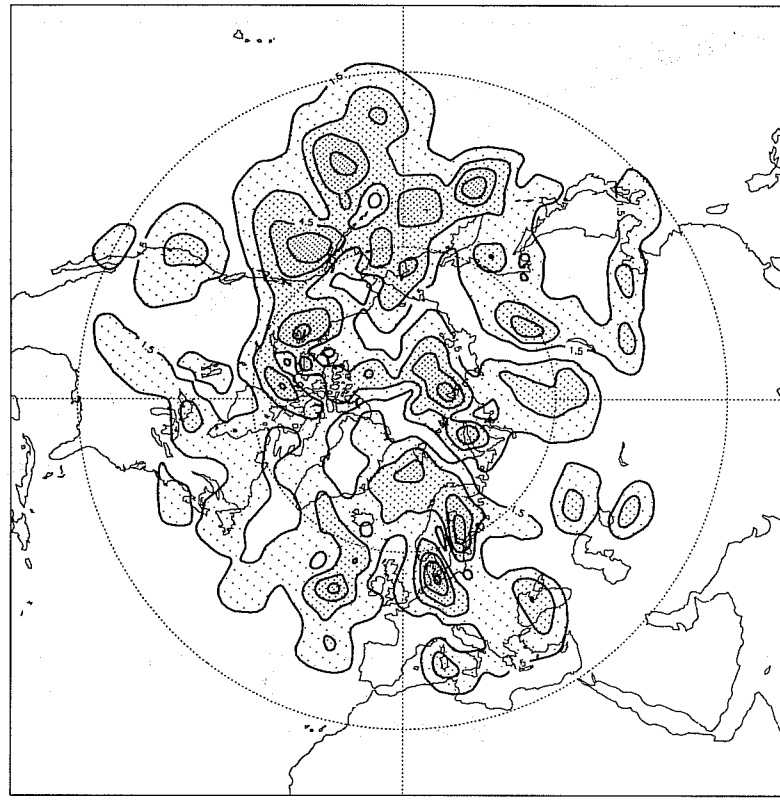


Fig 15 Evolution of negative (left column) and positive (middle column) sensitivity increments and evolution of control forecast errors (right column). Forecast ranges are from top to bottom panel: 2 days, 4 days and 6 days. Units: m.

a) T106 all waves t72 RMS-diff j1-j0 950519/to/950521 500 hPa
rms 1.21



b) T63 all waves t72 RMS-diff j1-j0 950519/to/950521 500 hPa
rms 0.84

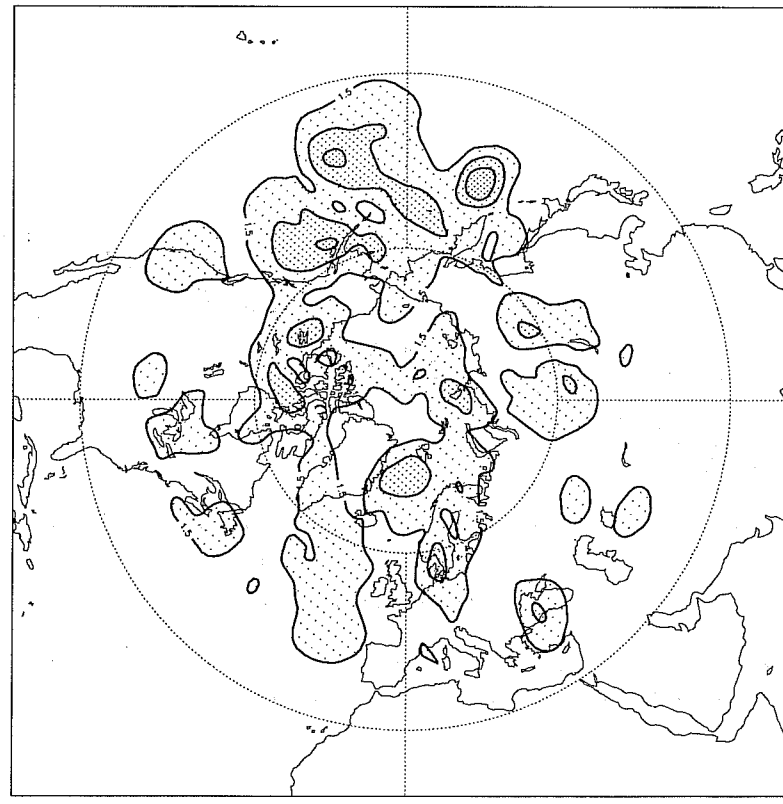


Fig 16 Root mean square values of evolved sensitivity increments after 3 days of integration time. The model resolution used is T106 in (a) and T63 in (b). Units: dam.

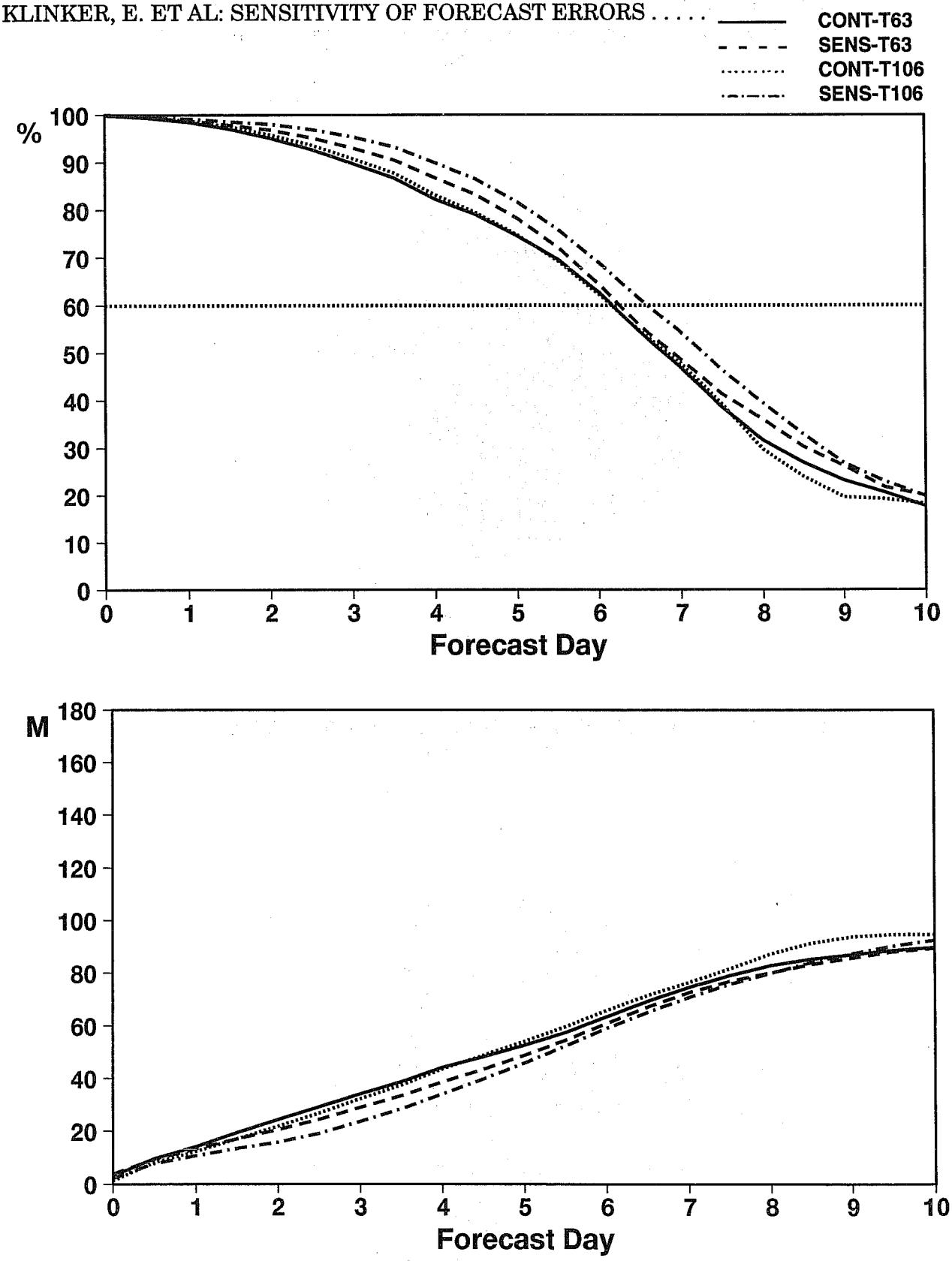


Fig 17 Verification scores for control forecasts (CONT) and sensitivity integrations (SENS) for the Northern Hemisphere. Model resolution T106 and T63. Top panel anomaly correlation, bottom panel root mean square errors.

developed out of phase. It therefore seems that for this area analysis errors further upstream, which project onto the unstable sub-space, are only partly responsible for forecast errors in the medium range.

Over western Europe, where the sensitivity perturbations seemed to affect the non-linear error growth after day-4, a much larger impact could be seen on the medium-range forecast errors. In particular, over the West-Atlantic and Europe the forecast errors and the evolved perturbations are in phase, indicating a substantial reduction of errors in the integration with positive perturbations.

6. MODEL INTEGRATIONS WITH SENSITIVITY PERTURBATIONS

6.1 **Comparison between T106 and T63 integrations**

An important influence on the growth rate of initial perturbations during a model integration could be expected from the horizontal resolution of the model used for the sensitivity integrations. Experience in some cases that were investigated in 1994 based on sensitivity gradients had shown that a change from T106 to T213 increased the perturbation growth even though the spectral resolution of the initial perturbations was T63 in both cases (*Rabier et al*, 1996). For these cases, subjectively scaled sensitivity gradients were used to perturb the analysis. The experimentation presented here is based on using perturbations from three steps in the sensitivity minimization procedure.

Figure 16 shows local rms values of the perturbations for three cases after 72 hours of integration time for the two resolutions of T63 and T106. In almost all areas the perturbations have grown to larger amplitudes in the T106 integration than in the lower resolution run. The T106 version of the model shows more details in many areas as well. The difference in perturbation growth is summarized in the Northern Hemisphere rms value. From an initial value of 0.12 dam the perturbations grow by a factor of 7 to .84 dam in the T63 integration and by a factor of 10 to 1.21 dam in the T106 integrations.

Of course the question remains whether the increased perturbation growth improves the forecast performance as well. Some improvement of the T106 control forecast (J0i0-T106) compared to the T63 (J0i0-T63) control forecast can be seen from the verification scores for the Northern Hemisphere (Fig 17). The scores show that the initial T63 perturbations add more skill to the forecast in the high resolution environment than in the low resolution model.

6.2 **The choice of sensitivity calculations in the operational suite**

During summer 1994 the experimental program related to the sensitivity of forecast errors to initial conditions concentrated on the use of the two day-lagged gradient J48 based on the 48 hour forecast errors, J48. See Appendix for definition of the different gradients. The improvement in the short and medium-range forecast skill based on the modification of the initial conditions using the J48 gradient could not be achieved by the use of a

gradient based on the 24 hour forecast errors, J24. A better performance was noticed from a gradient J48/24 based on the difference between today's day-1 and yesterday's day-2 forecast. Therefore, it was decided to retain the J48 and the J48/24 gradient in the operational suite, both based on the Northern Hemisphere. The additional gradient calculations performed on the European area did not prove to be useful as medium-range forecast errors for Europe frequently originate further upstream over North America or even the Pacific.

The success of correcting errors in the initial conditions led to further research in the direction of a possible application in real forecast mode. A waiting time of 48 hours, as required for modifications of the initial conditions based on J48, is normally not compensated by a gain of forecast skill of at least 48 hours. A relatively small waiting time of 12 hours could be achieved by using the inconsistency between the 0z and 12z forecast. This so called J48/36 gradient follows the singular vector approach of Ron Gelaro to propagate the 12 hour forecast error forward in time for 36 hours and then calculate, with the singular-vector equivalent of the adjoint, the sensitivity at initial time. In a similar way the adjoint J48/36 gradient is based on the forecast inconsistency between the 48-hour forecast from 12 z and the 36-hour forecast from 0z, whereby the adjoint calculations follow the 48-hour trajectory. As for the J48 calculation, 3 iterations of the minimization procedure are carried out to derive optimal perturbations without increasing the computational cost too much.

6.3 **Performance of the T106 integrations**

Based on the experience that the use of T63 as horizontal resolution reduced the growth of perturbations, most research experiments were performed with the T106 resolution. A longer series of 20 days of integrations was run for June 1995. As the inconsistency of the 48-hour and 36-hour forecasts is due to errors in both initial conditions at 12z and 0z, we expect only to see a fraction of the true sensitivity of forecast errors to initial conditions at 12z from J48/36. By multiplying the sensitivity perturbation by a factor of 2 the perturbations had the same amplitude as the J48 perturbations which seemed to be a relevant choice.

The results show behaviour of the sensitivity integrations that has been found for other periods as well. The J48-integrations show a general improvement in forecast skill compared to the control runs (Fig 18 for the NH). The fact that the minimization is done for the Northern Hemisphere area explains why, on smaller sub-areas like Europe (not shown), an occasional deterioration of skill by the use of the J48 perturbations can be seen. The definition of the cost-function for a smaller area sufficiently further upstream would probably improve the performance over Europe.

The anomaly correlation graph for the J48/36-integrations shows the performance depends on the skill of the control forecast. At a number of days when the control forecast changes from a relatively bad to a better score, the J48/36 integration shows the opposite behaviour of improving on the bad control forecast but doing worse for the previous good forecast. This suggests that analysis errors causing a poor 12z forecast for a specific day

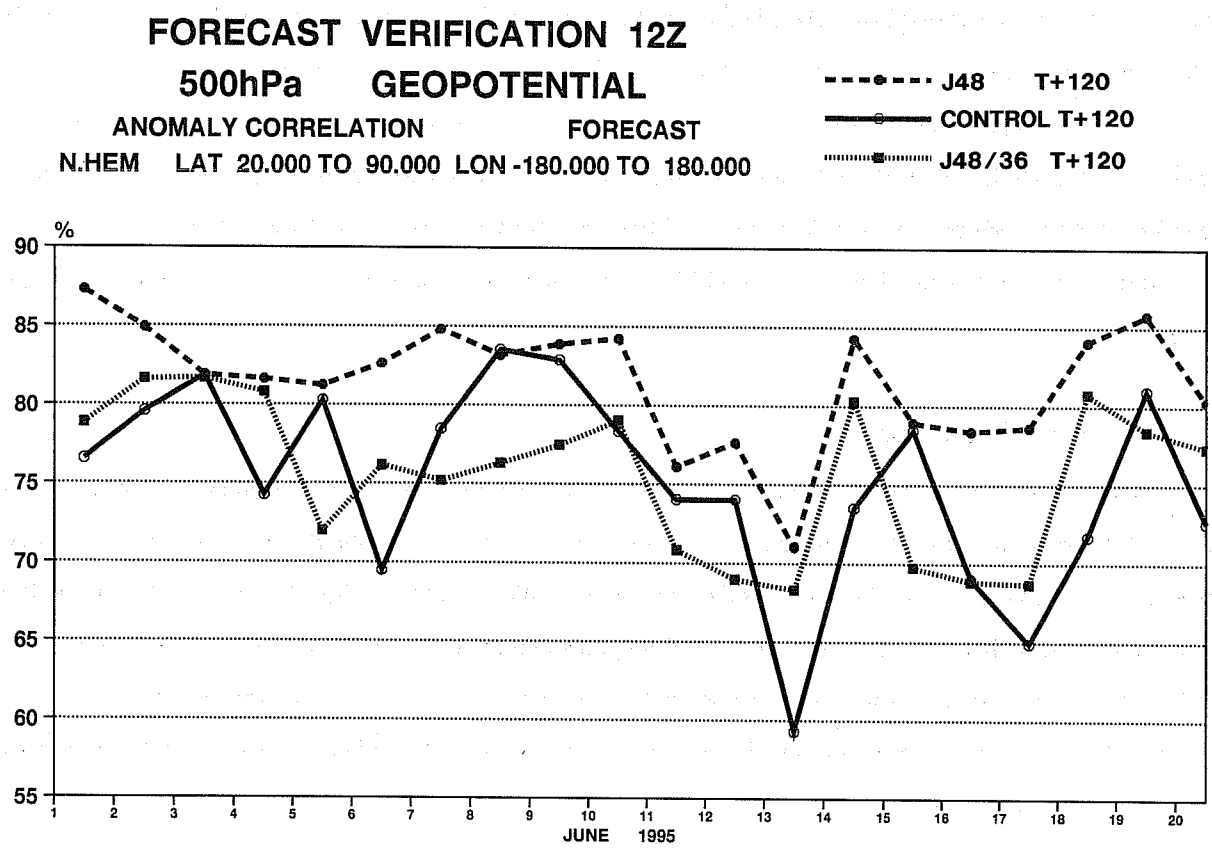


Fig 18 Time series of day 5 anomaly correlations for the Northern Hemisphere. Control forecast (solid line), 2 day-lagged sensitivity integration (dashed line), 12 hour-lagged sensitivity integration (dotted line).

were already present 12 hours earlier in the 0z analysis, causing the J48/36 gradient to correct the previous 12z analysis into the wrong direction by interpreting the inconsistency between the 0z and 12z-forecast as a problem in the 12z-analysis. On average over the 20 cases in June (Fig 19) the J48/36 integrations showed fairly neutral behaviour with slight gains in the short-range forecast and slight losses in the medium range contrary to the J48 integrations which persistently showed an improved skill for all forecast ranges. From these scores that show a gain by J48/36 of less than 12 hours it seems unlikely that an application for real time forecasting would be beneficial if the perturbations are restricted to a positive sign. However, two sets of forecast runs from negative and positive perturbations based on J48/36 would form a valuable extenstion of the existing ensemble prediction system.

7. CONCLUDING REMARKS

This study confirms that the sensitivity of forecast errors to initial conditions, calculated operationally at ECMWF for more than two years, provides useful diagnostic guidance for identifying areas in which analysis errors contribute to large forecast errors. As the sensitivity gradient is closely linked to the stability of the flow it is not surprising that the largest sensitivity is found during the winter when the flow is more baroclinically unstable than in the summer. Apart from the dominating seasonal effect there is also an influence from the model formulation itself. Reduced forecast errors due to improved parametrization schemes have made the short range forecast less sensitive to analysis errors.

The choice of the inner product in the adjoint calculations has an important effect on the horizontal structure of the sensitivity gradient. By using an "enstrophy" like inner product as an alternative to energy the sensitivity gradient obtains an unrealistically large scale. Model experiments with different increments confirm that the large scale part of the enstrophy based sensitivity gradient does not grow as fast as the small scale part that is more similar to the small scale structures identified with the energy norm. The fact that the energy norm provides an efficient way of identifying the relevant scales of analysis errors is confirmed by comparing the modified initial conditions to observational data.

Whereas the sensitivity gradient represents the direction that contributes to the fastest growing analysis error, the application of an iterative procedure minimizing the short-range forecast error leads, after some iterations, to so called sensitivity increments that describe a larger subspace of analysis error directions. A large number of iterations in the minimization algorithm does not seem to be beneficial, in particular, as there is still a discrepancy between the model with simplified physics used for the adjoint integrations and the forward model with the full physics. It seems that three iterations provide a reasonable structure of fast growing analysis errors which substantially improves the fit to observations when subtracted from the analysis.

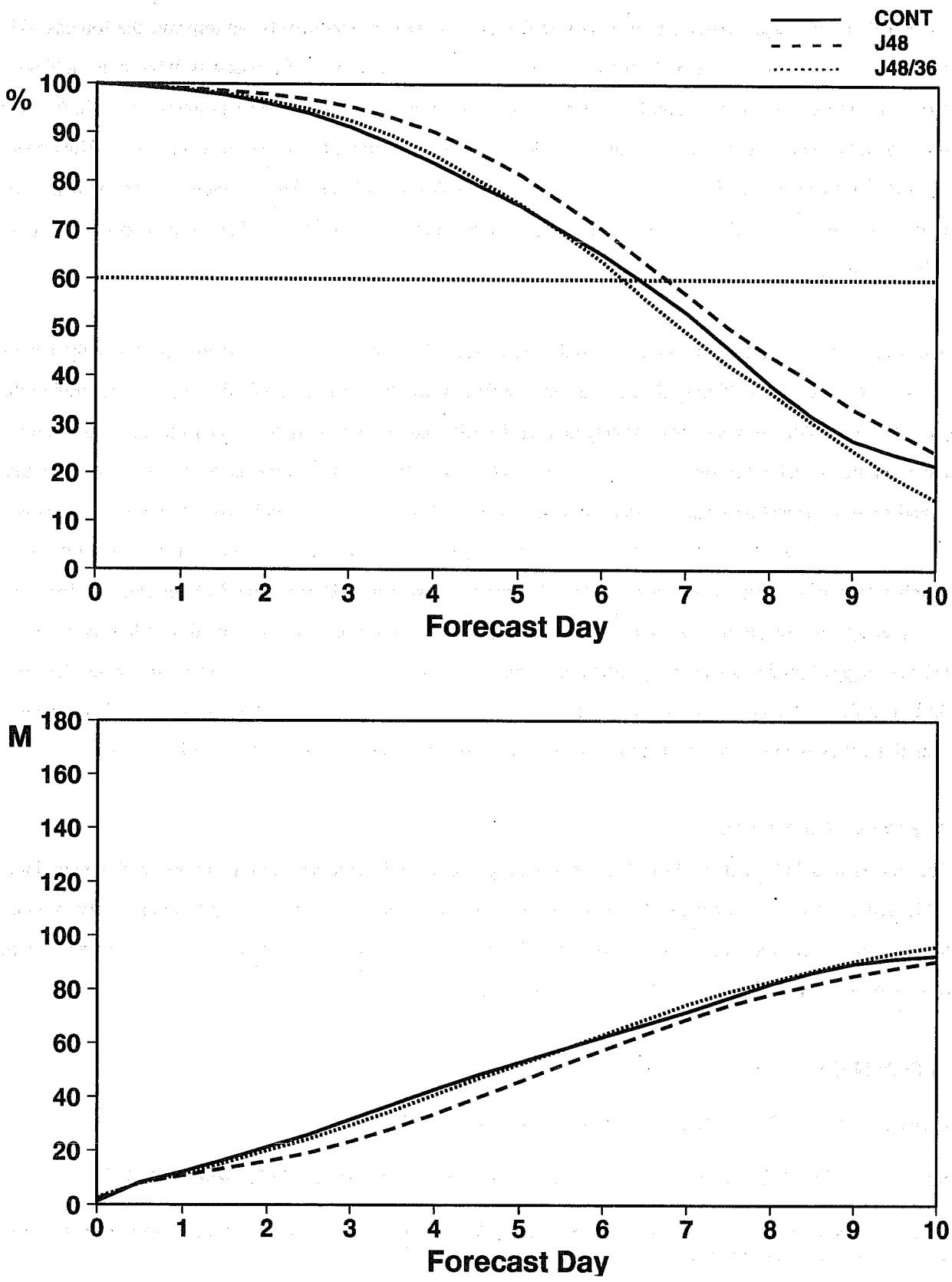


Fig 19 Mean verification scores for control forecasts (solid line), 2 day-lagged sensitivity integration (dashed line), 12 hour-lagged sensitivity integration (dotted line).

Modifications of analyses based on sensitivity gradients or sensitivity increments can improve the forecast skill beyond the optimization time of 48 hours, in some cases spectacular gain of forecast skill has been achieved. However, the success during non-linear error growth is not guaranteed. The fast growing errors in the first 48 hours may be unrepresentative for later error development from initially smaller growing errors. The benefit in model performance is also dependent on the model resolution. With initial increments of identical spectral resolution, higher resolution versions of the model allow a faster growth of perturbations and produce a larger gain in predictive skill

The improvement of forecast skill by adding sensitivity increments to the analysis can normally not be transformed into real predictive skill, as the waiting time of for the verifying analysis is large compared to the gain of forecast skill. A reduction in waiting time can be achieved by shortening the forecast length. As the main data volume available for the analysis arrives only at 12z and 0z, a 12 hour forecast length is the shortest that would allow a verification against an analysis of sufficient quality to identify model errors. Experience, however, has shown that the 12 hour forecast errors are not sufficiently large to obtain a well defined sensitivity field. Much larger differences evolve between the 48 hour forecast from 12z from the 36 hour forecast from 0z. Calculating the sensitivity of this forecast divergence with respect to the initial conditions at 12 z (12 hour-lagged) produces sensitivity structures similar to the 48-hour forecast error sensitivity (2 day-lagged). Although the 12 hour-lagged sensitivity increments do not improve the forecast as much as the 2 day-lagged sensitivity increments, it seems that in particular bad forecast cases a real predictive skill is possible.

ACKNOWLEDGMENTS

The authors would like to thank M Fisher for providing a relevant interpretation of the minimization technique, P Courtier, A Hollingsworth and A Persson for their useful comments which helped to improve the original manuscript. J Rousseau and J-L Perrin are acknowledged for their contribution to the comparison with observations. C Edis-Williams efficiently typed the manuscript.

REFERENCES

Bertsekas, D P, 1995: Nonlinear programming, pub Athena Scientific.

Bouttier, F, 1993: The dynamics of error covariance in a barotropic model. Tellus, **45A**, 408-423.

Buizza, R, R Gelaro, F Molteni and T N Palmer, 1995: Predictability studies with high resolution singular vectors. (Submitted to QJRMS).

Buizza, R, 1994: Sensitivity of optimal unstable structures. QJRMS, **120**, 429-451.

Errico, R, T Vukicevic and K Reader, 1993a: Comparison of initial and lateral boundary condition sensitivity for a limited-area model, **45A**, 539-557.

Errico, R M, T Vukicevic and K Reader, 1993: Examination of the accuracy of a tangent linear model. Tellus, **45A**, 462-477.

Errico, R and T Vukicevic, 1992: Sensitivity analysis using an adjoint of the PSU/NCAR mesoscale model. Mon Wea Rev, 120, 1644-1660.

Farrell, B F,1990: Small error dynamics and the predicatbility of atmospheric flows. J Atmos Sci, 47, 2409-2416.

Gilbert and Lemaréchal, 1989: Some numerical aspects with variable storage Quasi-Newton algorithms. Mathematical programming, **B25**, 407-435.

Hall,M C G, 1986: Application of adjoint sensitivity theory to an atmospheric general circulation model. J Atmos Sci, 43, 2644-2651.

Horanyi, A and A Joly, 1995: Some aspects of the sensitivity of idealized frontal waves. Submitted to Beiträge zur Physik der Atmosphäre.

Hoskins, B J and P J Valdes, 1990: On the existence of storm-tracks. J Atmos Sci, **47**, 1854-1864.

Lacarra, J-F and O Talagrand, 1988: Short-range evolution of small perturbations in a barotropic model. Tellus, **17**, 321-333.

Langland, R H, R L Elsberry and R M Errico, 1995: Evaluation of physical processes in an idealized extratropical cyclone using adjoint sensitivity. QJRMS, **121**, 1349-1386.

LeDimet, F-X and O Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations. Tellus, 38A, 97-110.

Marais, C and L Musson-Genon, 1992: Forecasting the surface weather elements with a local dynamical-adaptation method using a variational technique. Mon Wea Rev, 120, 1035-1049.

Molteni, F, R Buizza, T Palmer and T Petroliagis, 1996: The ECMWF ensemble prediction system: methodology and validation. Q J Roy Meteor Soc, **122**, 73-119.

Nocedal, J, 1980: Mathematics of computation, Vol 35, No. 151, pp 773-782.

Oortwijn, J and J Barkmeijer, 1995: Perturbations that optimally trigger weather regimes. JAS, Vol 52, No. 22, 3932-3944.

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

Rabier, F, P Courtier and O Talagrand, 1992: An application of adjoint models to sensitivity analysis. Beitr Phys Atmosph, 65, 177-192.

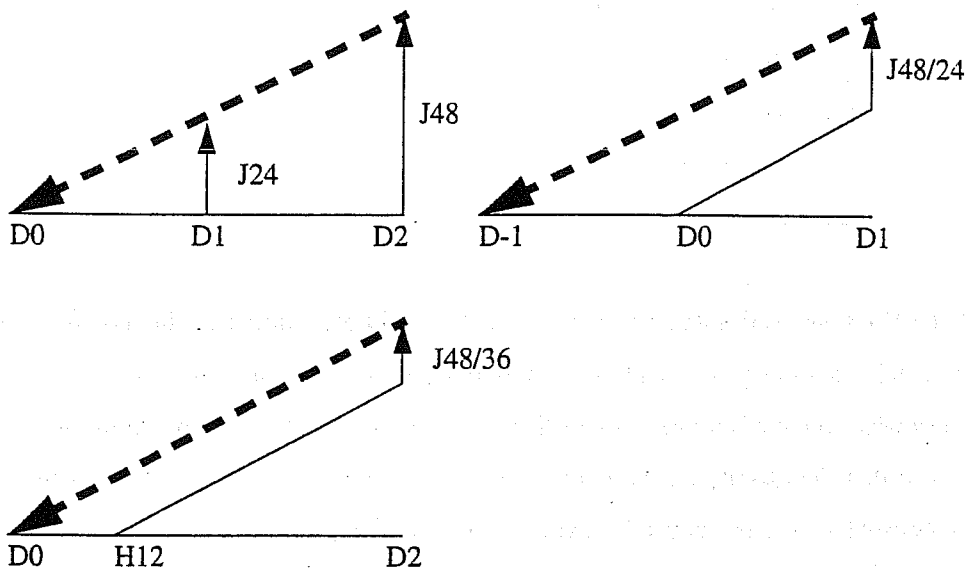
Rabier, F, E Klinker, P Courtier and A Hollingsworth, 1996: Sensitivity of forecast errors to initial conditions. QJRMS, **122**, 121-150.

Rinne, J and H Järvinen, 1993: Estimation of a Cressman term for a barotropic model through optimization with use of the adjoint model. MWR, Vol 121, No. 3, 825-833.

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

Zupanski, M, 1995: An iterative approximation to the sensitivity in calculus of variations. MWR, Vol 123, 3590-3604.

Appendix A



Definition of gradients

APPENDIX B

The minimization uses the quasi-Newton algorithm M1QN3 developed at INRIA (*Gilbert and Lemarechal*, 1989). The algorithm may be written concisely as

$$d_i = -H_i \nabla J(\delta x_i^a)$$
$$\delta x_{i+1}^a = \delta x_i^a + \alpha_i d_i$$
$$H_{i+1} = F(s_0, s_1, \ldots s_i; y_0, y_1, \ldots y_i; H_i^0)$$

where $s_i = \delta x_{i+1}^a - \delta x_i^a$ and $y_i = \nabla J(\delta x_{i+1}^a) - \nabla J(\delta x_i^a)$.

Nocedal (1980) demonstrated that under certain restrictive conditions, successive iterates δx_i^a produced by the algorithm are identical to those produced by conjugate-gradient minimization of J with exact line searches. The conditions under which this equivalence holds are that α_i is chosen at each iteration to minimize J along the line $\delta x_i^a + \alpha d_i$, and that the identity matrix is used for H_i^0. Neither of these conditions is met exactly in our algorithm. Nevertheless, we expect the following result to hold to good accuracy.

It is well known (see for example *Bertsekas*, 1995) that for a quadratic function the n^{Th} iteration of conjugate-gradient minimization with exact line searches minimizes J over all vectors in the Krylov space

$$\{\delta x_0^a,\ \delta x_0^a + (J'')\delta x_0^a,\ \delta x_0^a + (J'')^2\ \delta x_0^a,\ \ldots,\ \delta x_0^a + (J'')^{n-1}\ \delta x_0^a\}$$

where (J'') is the Hessian matrix of J.

As n increases, $(J'')^{n-1}\ \delta x_0^a$ points increasingly in the directions of the leading eigenvectors of (J''), and J is rapidly minimized in these directions. (It is this property of Krylov spaces to rapidly converge towards the space spanned by the leading eigenvectors of a matrix which is exploited by the Lanczos algorithm, which is also closely related to conjugate-gradient minimization.)

The Hessian matrix of J is $M_s'^T A M_s'$. The eigenvectors of this matrix are the singular vectors of M_s' with respect to the inner product $[.,.]_A$. Thus, to the extent to which the quasi-Newton minimization algorithm mimics conjugate-gradients, J is rapidly minimized in the directions spanned by the leading singular vectors of the tangent linear model.