TECHNICAL MEMORANDUM

205

Linearity of optimal perturbation time evolution and sensitivity of ensemble forecasts to perturbation amplitude

R. Buizza

Research Department

September 1994

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

SUMMARY

We discuss the perturbations which grow most rapidly over a finite time interval in a primitive equation atmospheric model. They are the singular vectors of a linear approximation of the European Centre for Medium-Range Weather Forecasts primitive equation model. They are computed using the adjoint technique at horizontal spectral truncation T21 with 19 vertical levels.

Linear combinations of singular vectors, named optimal perturbations, can be used in ensemble prediction to generate the initial conditions of perturbed integrations. Having specified the initial amplitude to be comparable to the amplitude of analysis error estimates, we study the non-linear time evolution of optimal perturbations when added to the control initial conditions. In particular we estimate the time limit τ_{LIN} after which non-linear processes cannot be neglected. Considering optimal perturbations generated using singular vectors with maximum growth over a 36 hour time interval, and characterized by maximum amplitudes comparable to analysis error estimates, two different methods give an estimated value $\tau_{LIN} \sim 2\text{–}2.5$ *days*. In the second part of this paper we analyze the sensitivity of ensemble predictions to the optimal perturbation amplitude. This sensitivity study gives useful indications for future developments.

An estimate of the possible impact of the reduction of the amplitude of analysis errors on the skill of numerical weather prediction is deduced from the comparison of ensemble experiments run with initial perturbations characterized by different amplitudes. Results indicate that a reduction of the root mean square amplitude of the analysis error by a factor $\sqrt{2}$ may lead to an improvement of medium-range predictability up to 1 day, and that a reduction by a factor $2\sqrt{2}$ may reduce the errors of the 7-day forecast to values shown, at present, at forecast day 5.

1. INTRODUCTION

Ensemble prediction is a probabilistic approach to numerical weather prediction (*Epstein*, 1969; *Leith*, 1974). Instead of providing one (deterministic) predicted state only, it can give an estimate of the probability distribution of forecast states. Since December 1992 both the US National Meteorological Center (NMC) and the European Centre for Medium-Range Weather Forecasts (ECMWF) have been producing and disseminating ensemble prediction products (*Tracton and Kalnay*, 1993; *Palmer et al*, 1993).

The ECMWF Ensemble Prediction System (EPS) is based on 33 integrations of the ECMWF primitive equation model, at horizontal spectral triangular truncation T63 with 19 vertical levels (T63/L19). *Molteni et al* (1994) describe the ECMWF EPS, which uses singular vectors, i.e. structures with maximum growth over a finite time interval (*Farrell*, 1988) called the optimisation time interval, to define the initial conditions of the perturbed members. If the initial error distribution is isotropic in phase space and then evolves to an ellipsoid, then the major axes of the ellipsoid are determined by singular vectors of a linear primitive equation model. Singular vectors can therefore be used to generate initial perturbations for ensemble forecasts (*Mureau et al*, 1993; *Palmer et al*, 1993; *Buizza et al*, 1993). Linear combination of singular vectors will be named optimal perturbations.

Buizza and Palmer (1994) provide a description of the basic formalism and a comparison between singular vectors and other unstable vectors (e.g. normal modes, Lyapunov vectors), together with singular vector statistics of routine calculations made between December 1992 and August 1993.

In contrast with this approach, the NMC EPS is based on perturbations computed using the method denoted "breeding of growing modes" (*Toth and Kalnay*, 1993). This method consists of running an additional, perturbed short-range forecast, in addition to the regular forecast in an analysis cycle. The difference between the control and perturbed six-hour (first-guess) forecast is scaled back to the size of the initial perturbation, and then reintroduced onto the new atmospheric analysis. Thus, the perturbation evolves along with the analysis, ensuring that after a few days of cycling the perturbation consists of the superposition of fast-growing modes corresponding to the most recent past evolution of atmosphere. *Toth and Kalnay* (1993) and *Buizza and Palmer* (1994) compare the ECMWF and the NMC methods.

In the ECMWF EPS optimal perturbations are added to the control initial conditions to provide an efficient sampling of the components of the probability distribution function of analysis error which will lead to rapid growth of forecast errors. Their amplitude is set to be close to the typical amplitude of Optimal Interpolation estimates of analysis errors. Both the strength and the weakness of the singular vector approach to dynamical instability is the restriction to linear dynamics that permits the use of linear model equations.

Considering a more simplified model, *Lacarra and Talagrand* (1988) experimentally showed that the barotropic time evolution of small perturbations (with amplitude comparable to analysis errors) can be described by its linear approximation if the time interval is not longer than 2-3 days.

Rabier and Courtier (1992) addressed the question of how important the influence of non-linear terms is in the evolution of the atmospheric flow in cases of strong baroclinic development. They considered the zonal part of a flow as its basic state, the eddy part of the flow as a perturbation, and they quantified the relative importance of the linear and the non-linear terms during the perturbation time evolution. Their initial perturbation was characterized by an amplitude that was far from negligible, with surface pressure components up to 20 hPa and lowest level temperature components up to 10 K. Their results showed that the tangent-linear model describes to a good degree of approximation the evolution of the eddy part of the perturbation for about 2 days, while the total (eddy plus zonal) evolution of the perturbation is less linear than the evolution of the eddy part only.

Vukicevic (1991) investigated linearity of initial error evolution using a primitive equation limited area model, and demonstrated that the major portion of the initial forecast errors (with magnitude comparable

with analysis errors) can be described by the tangent model solutions for periods of 1-1.5 days duration. In a later paper, *Errico et al* (1993) with the same limited area model studied the 72h time evolution of small perturbations with amplitude comparable to analysis errors, during a summer and a winter case. They compared the difference between the non-linear and the linear time integration (error) with the perturbation field itself, and they concluded that in terms of maximum absolute values of fields, at the end of the integration period, the ratio of perturbation to error values is approximately five for the winter case and two for the summer case.

Buizza (1994a), comparing subjectively the time evolution of integrations started adding and subtracting the same structure to the control initial conditions with amplitude comparable to (Optimal Interpolation) analysis error estimates, concluded that non-linear effects are small up to forecast day 2 but they can be quite large after forecast day 4.

In the first part of this paper we verify that the time evolution of optimal perturbations with initial amplitude comparable to analysis error estimates can be linearly approximated up to optimisation time, and we estimate the time limit τ_{LIN} after which the contribution of non-linearity to the perturbation time evolution cannot be neglected.

In the second part of this paper we focus on the EPS skill. *Toth and Kalnay* (1993) stated that a requirement for successful ensemble forecasting is that "the time evolution of the atmosphere should be a plausible member of the ensemble". The analysis of the ECMWF EPS results so far have shown that, for some cases, the trajectory followed by the atmosphere appeared to be outside the range of forecast states estimated by the EPS (*Molteni et al*, 1994). This could be due to a model deficiency in predicting that particular situation, or it could indicate that the ensemble members should be characterized by a larger spread to actually cover all the probable forecast states. We investigate whether multiplying the root mean square (hereafter rms) perturbation amplitude by a factor $\sqrt{2}$ can enlarge the spread among the perturbed forecasts and increase the EPS skill.

Ensemble experiments have been run also dividing the rms perturbation amplitude by a factor $\sqrt{2}$. The analysis of the variation of the spread when multiplying or dividing the rms perturbation amplitude can be used to evaluate the possible impact of the reduction of analysis errors on the forecast skill of a deterministic prediction.

Our results are based on ensemble experiments run for three cases, with starting dates 92.12.28, 93.02.14 and 93.09.06, in three different configurations named normal-amplitude, small-amplitude and large-

amplitude. Considering the amplitude of the normal-amplitude perturbations as reference (amplitude 1), the large-amplitude and the small-amplitude configurations are characterized by the same perturbations but with rms amplitude multiplied, respectively, by a factor $\sqrt{2}$ and $1/\sqrt{2}$. Moreover, for one case only (92.12.28), an ensemble has been run with rms amplitude multiplied by a factor of 0.5.

Section 2 briefly describes the EPS and the three atmospheric situations we studied. In section 3 we investigate the linearity of the optimal perturbation time evolutions, while the impact of the perturbation amplitude on the EPS performance is analyzed in section 4. The results from sections 3 and 4 are further discussed in section 5, where some conclusions are drawn.

2. ENSEMBLE PREDICTION AT ECMWF

In the first part of this section we briefly review the ECMWF EPS characteristics, while we refer to *Molteni et al* (1994) for the complete description of the system. In the second part of this section we review the atmospheric circulation associated with the three analyzed case studies.

2.1 **The Ensemble Prediction System**

The ECMWF EPS is based on an ensemble of 32 perturbed and one unperturbed (control) time integration of the T63/L19 version of the ECMWF model. The initial conditions of the 32 perturbed members are created by adding optimal perturbations to the control initial conditions. The optimal perturbations are defined using the singular vectors of a linear primitive equation model. The singular vectors are computed using the adjoint technique. The four main steps of the EPS are listed below.

i) Singular vectors computation

If $\underline{x}$ is a perturbation, and

$$\underline{x}(t) = L\,\underline{x}_0 \tag{1}$$

is the solution of the linear model equations

$$\frac{\partial \underline{x}}{\partial t} = A_l\,\underline{x}, \tag{2}$$

where A_l is an approximation of the tangent linear version of the ECMWF primitive equation non-linear model A, then the total energy norm of the perturbation at time t can be computed as

$$\|\underline{x}(t)\|^2 = (T\,L\,\underline{x}_0; T\,L\,\underline{x}_0) = (x_0; L^*\,T^*\,T\,L\,x_0), \tag{3}$$

where L^* is the adjoint of the linear propagator L (see eq (1)) with respect to the total energy scalar product (...;...) (*Buizza et al*, 1993, eq (5.1)], and T^* is the adjoint of the local projection operator T

(*Buizza and Palmer*, 1994). The application of the local projection operator permits the identification of singular vectors characterized by maximum growth over the region of the northern hemisphere (NH) with latitude $\lambda \geq 30°N$ (*Buizza*, 1994a). At present, the trajectory is computed applying the ECMWF version of the vertical diffusion and surface drag scheme, while the linear model and its adjoint use a simplified linear scheme (*Buizza*, 1994b).

The singular vectors identify the directions of the phase space of the system that maximize the energy growth defined by eq (3). They depend on the optimization time interval t and on the choice of norm. For our EPS experiments they are computed with the T21/L19 version of the Integrated Forecasting System (*Courtier et al*, 1991).

ii) Generation of 16 perturbations

From the singular vectors computed in step i) 16 singular vectors are selected, rejecting singular vectors located in the same region. Since the 16 selected singular vectors can be very localized in space, a phase-space rotation is applied to generate 16 less localized fields. Moreover, the 16 singular vectors are re-scaled in order to have local maxima comparable to local analysis errors estimated by the Optimal Interpolation procedure. The fields generated following this procedure are named optimal perturbations. As a consequence of the phase-space rotation they are characterized by similar amplification rates.

iii) Generation of 32 initial conditions

The 32 initial conditions of the perturbed forecasts are generated adding to the control initial conditions the optimal perturbations generated in step ii), either with a positive or a negative sign.

iv) Model integrations

A control 10-day integration and 32 10-day integrations are performed with a version of the ECMWF spectral model, with horizontal triangular spectral truncation T63 and 19 vertical levels, with full physics.

2.2 Synoptic situations

We now describe the synoptic situations during the three studied 10-day periods. They can be classified as cases of good (93.02.14), medium (92.12.28) and poor (93.09.06) forecast skill by considering the skill of the control at forecast day 7 and onwards, measured by the anomaly correlation coefficient (ACC) between the forecast and the analysis of the 500 hPa geopotential height (see Appendix A). Spread and skill scores reported in this work are computed using 500 hPa geopotential height fields for the NH and the European region. The first is defined as the region of the NH with latitude $\lambda \geq 30°N$, while the European region is defined as the area with $30 \leq lat \leq 75$, and $-20 \leq lon \leq 45$. As a reference value for spread and skill scores we consider that a value of 0.6 ACC indicates a limit of useful predictions. Due to

limitations of space, only the 92.12.28 case characterized by medium skill scores will be investigated in detail, especially over the European region.

a) 92.12.28 (medium skill)

Fig 1 shows the 500 hPa geopotential height at 92.12.28 12GMT (panel a) and 7 days later (panel b)). At initial time, both the Pacific and the eastern Atlantic and western Europe sectors are characterized by a blocked flow. The strong ridge over the eastern Pacific weakens during the following days, although an anti-cyclonic circulation persists for the whole 10 day period. In the eastern Atlantic and western European sector, the initial anti-cyclonic circulation over the British Isles changes to a more zonal circulation during the following 5 days. An area of positive height anomaly continues to influence central Europe, while a cut-off low system develops over the Mediterranean sea. This leads, after forecast day 7, to a split circulation over Europe (see Fig 1b). This blocked situation lasts for only 2 days, later evolving into a more zonal flow.

Fig 2 shows the spread (panel a) and the skill (panel b) of the normal-amplitude experiment, computed over the NH. The spread among the ensemble members is quite large, with ACC differences from control ranging from 0.24 to 0.84 at forecast day 7. It is a case of medium skill, with the control ACC skill score of 0.6 at forecast day 6, and with a value of 0.52 at forecast day 7. The control fails to predict the persistent positive height anomaly over the eastern Pacific. Almost 2/3 of the perturbed forecasts score better than the control, with the best having ACC skill score of 0.74 at forecast day 7. Table 1 summarizes the spread and skill scores at forecast day 7.

b) 93.02.14 (high skill)

Fig 1 shows the situation at 93.02.14 12GMT (panel c) and 7 days later (panel d). On 14 February a region of high pressure dominates the European area, with a south-westerly flow north of the British Isles and a trough over Russia. A zonal flow characterizes the circulation over the Asian continent and the western Pacific, with a deep ridge north-west of the Rockies and a region of low pressure over the eastern Canada. After 2 days, the ridge north-west of the Rockies starts evolving into a blocking structure which lasts until forecast day 7. During the same time, an anti-cyclonic circulation starts characterizing the eastern Atlantic, and on the 21st a north-westerly flow dominates the British Isles (see Fig 1d). The trough over the central Europe deepens during the following days, and a cut-off low develops on the 23rd over the Adriatic sea.

Fig 3 shows the NH spread and the skill scores. This is a case of particularly low spread, with the ACC spread scores ranging from 0.67 to 0.90 at forecast day 7 (see Table 1). It corresponds to a very good control forecast, which is characterized by an ACC skill score higher than 0.60 during the whole 10 day period (0.72 at forecast day 7).

c) 93.09.06 (poor skill)

Fig 1 shows the situation at 93.09.06 12GMT (panel e) and 7 days later (panel f). A dipole circulation, with a ridge over the north-western America and a cut-off low over the eastern Pacific, characterizes the circulation of the eastern Pacific-American sector during the first 3 days, becoming more zonal after forecast day 4. In contrast, a trough affects the western Pacific flow during the whole 10 day period. During the first 5 days, a cut-off low persists in the British Isles region, while a zonal circulation characterizes the southern Europe. After forecast day 6, the cyclonic circulation over the south of England becomes stronger, deepening the day after (see the situation on the 13th in Fig 1f) due to the interaction of the middle tropospheric flow with a very strong low-level cyclone coming from the Atlantic ocean. The flow becomes gradually more zonal after the 14th.

The NH spread and the skill scores are shown in Fig 4. Among the three analyzed situations, this is the case with the largest spread, with ACC at forecast day 7 ranging between 0.20 to 0.90 (see Table 1). It also corresponds to the poorest control prediction, with the control ACC skill score curve crossing the 0.60 threshold at forecast day 6, but dropping to 0.36 ACC skill score at forecast day 7. After forecast day 4 the control forecast starts to predict a too smooth cyclonic circulation over south England, leading to a completely wrong ridge over the Scandinavian countries at forecast day 7.

3. LINEARITY OF OPTIMAL PERTURBATION TIME EVOLUTION

Following *Lacarra and Talagrand* (1988), *Vukicevic* (1991) and *Rabier and Courtier* (1992) but considering optimal perturbations, we estimate τ_{LIN}, the time limit after which non-linear processes dominate the optimal perturbation time evolution when they have initial amplitudes comparable to Optimal Interpolation analysis error estimates.

Considering the time evolution of singular vectors computed using primitive equation models, *Buizza* (1994a) compared the T21/L19 linear evolution of optimal perturbations with their non-linear T63/L19 evolution. The two time evolutions were not compared objectively, since the linear and the non-linear evolutions were evaluated using different models with different resolution. To avoid comparing time integrations computed using different models, the approach followed in this work is to use non-linear time evolutions to estimate the importance of the non-linear terms with respect to the linear one. This allows us to take into account all the dynamical and physical processes that influence the atmospheric circulation. It is worth stressing that at present, in fact, only the adiabatic part of the model equations and diffusion processes due to vertical diffusion and surface drag have been linearized. Sub-section 3.1 describes the methodology used for this study, while the results are reported in sub-section 3.2.

3.1 Methodology

Let us denote by $f(z_0,t)$ the solution of the non-linear model equations

$$\frac{\partial z}{\partial t} = A(z,t), \tag{4}$$

with initial conditions z_0. Let us denote by

$$\underline{x} = \epsilon\, \underline{\eta} \tag{5}$$

a perturbation that maximizes the energy norm (3), where $\underline{\eta}$ is a normalized vector. Let us suppose that the perturbation is small compared to the initial state

$$\epsilon \ll \|z_0\|_g, \tag{6}$$

where $\|z_0\|_g$ is the rms amplitude of z_0 computed in grid point space (physical space). Applying a Taylor expansion we can then approximate $f(z_0,t)$ as

$$f(z_0 \pm \epsilon\underline{\eta},t) = f(z_0,t)$$
$$\pm\, \epsilon\underline{\alpha}_1 + \epsilon^2\underline{\alpha}_2 \pm \epsilon^3\underline{\alpha}_3 + \epsilon^4\underline{\alpha}_4. \tag{7}$$

For each perturbation we can run two non-linear integrations adding and subtracting the perturbation to the control initial conditions, and two other integrations adding and subtracting the perturbation multiplied by a factor $\sqrt{2}$ to the control initial conditions. So the functions $f(z_0,t)$ (control integration), $f(z_0 \pm \epsilon\underline{\eta},t)$ and $f(z_0 \pm \epsilon\sqrt{2}\underline{\eta},t)$ are known. They define left hand side terms of eq (7). Then, applying for each perturbed integration the Taylor expansion (7) truncated at the fourth term, we can write a linear system in the unknown variables $\underline{\alpha}_i$ i=1,..,4.

Once the unknown variables $\underline{\alpha}_i$ i=1,..,4 have been evaluated, we can compare their amplitudes to estimate the importance of non-linear terms with respect to the linear one. Let us define

$$\underline{\alpha}_0 = f(z_0,t), \tag{8}$$

and let us define the normalized vectors

$$\underline{u}_i = \frac{\underline{\alpha}_i}{\|\underline{\alpha}_i\|_g} \quad i=0,..,4. \tag{9}$$

Then eq (7) can be written as

$$f(z_0 \pm \epsilon\underline{\eta},t) = \underline{\alpha}_0$$

$$\left(\underline{u}_0 \pm \epsilon \frac{\|\underline{\alpha}_1\|_g}{\|\underline{\alpha}\|_{g0}} \underline{u}_1 + \epsilon^2 \frac{\|\underline{\alpha}_2\|_g}{\|\underline{\alpha}_0\|_g} \underline{u}_2 \pm \epsilon^3 \frac{\|\underline{\alpha}_3\|_g}{\|\underline{\alpha}_0\|_g} \underline{u}_3 + \epsilon^4 \frac{\|\underline{\alpha}_4\|_g}{\|\underline{\alpha}_0\|_g} \underline{u}_4 \right). \tag{10}$$

Let us define the coefficients

$$c_n(t) = \epsilon^n \frac{\|\underline{\alpha}_n\|_g}{\|\underline{\alpha}_0\|_g} \quad n=1,..,4. \tag{11}$$

The analysis of their time evolution can be used to compare the relative amplitude of the first four terms of the Taylor expansion (7).

A second estimate of the non-linearity of the perturbation dynamics can be deduced from the comparison of two perturbed forecasts started by adding and subtracting the same perturbation to the control initial conditions. Let us denote by $ens_+(t)$ and by $ens\_(t)$ the two perturbed runs. The time limit after which non-linear terms dominate the time evolution can be estimated from the analysis of the function $l(t)$ defined as

$$l(t) = - \frac{<(ens_+(t)-con(t));(ens\_(t)-con(t))>_g}{\|(ens_+(t)-con(t))\|_g \ \|(ens\_(t)-con(t))\|_g}. \tag{12}$$

The function $l(t)$ is equal to 1 at initial time $t=0$, when the two difference fields are perfectly correlated, and $l(t) \to 0$ after a time limit τ_{LIN} when non-linearity becomes dominant.

3.2 Linearity of optimal perturbation dynamics

We first deduce an estimate of the time after which non-linear terms become non-negligible from the analysis of the time evolution of the Taylor expansion coefficients $c_n(t)$ $n=1,..,4$, defined in eq (11). As described in section 2.1, each ensemble experiment consists of 16 perturbed integrations with initial conditions generated adding, and 16 subtracting, the optimal perturbations. So, for each case study, using experiments of the normal-amplitude and large-amplitude type we can get 16 evaluations of the four coefficients $c_n(t)$. Then, for every forecast time, averaged values can be computed among them to define the mean coefficients $\bar{c}_n(t)$.

Our results depend strongly on the perturbation amplitudes that are specified to be comparable to analysis error estimates (see sub-section 2.1, step iii)). Figs 5 and 6 show, respectively, the rms amplitude of the normal-amplitude perturbations and of the optimal interpolation estimate of the analysis error for the 92.12.28 case. Wind and temperature components are shown at model level 11 (around 500 hPa). The action of the local projection operator confines the optimal perturbation to the NH. In this region, the

maxima of the perturbation temperature component are similar to the estimated analysis error maxima, while the maxima of the perturbation wind components are almost 1/5 of the estimated analysis error maxima. Fig 5 shows that at model level 11 the perturbations are characterized by local maxima of the wind and temperature components around 1 m/s and 1.8 K (in terms of 500 hPa geopotential height, the perturbations have amplitude smaller than 12.5 m, not shown). Table 2 summarizes the rms amplitude of the estimated analysis error and of the normal-amplitude perturbations averaged over the NH region at model level 11. In rms terms, the perturbation wind components are one order of magnitude smaller than the estimated analysis error while the perturbation temperature component has a value comparable to the estimated analysis error.

Fig 7 shows the results of the non-linearity studies relative to the 92.12.28 case. For each panel, solid lines identify the 16 estimated coefficients $c_n(t)$, while the thick solid lines with large dots refer to the mean curves $\bar{c}_n(t)$. Fig 7 shows that, for each coefficient, the 16 curves do not differ too much from their mean values, with the second order term showing the largest variations. Note that they all remain much smaller than one, confirming that hypothesis (6) remains valid during the whole 10-day period.

To compare the four Taylor expansion terms, the four mean coefficients $\bar{c}_n(t)$ relative to this case have been plotted in Fig 8a. The third and the fourth order terms (dotted and chain-dashed lines) have similar values and remain smaller than the first two order terms (solid and dashed lines) for the whole period. The second order curve (dashed) crosses the first order curve (solid) around forecast day 5, and is the leading term from this time onwards. Note that already around forecast day 2 its magnitude is half the value of the first order term. We can consider the latter as an estimate of the time when non-linear terms are no longer negligible compared to the linear term.

Figs 8b and 8c show the mean coefficients $\bar{c}_n(t)$ computed for, respectively, the 93.02.14 and the 93.09.06 cases. These two cases also show that the first and second order terms are the most important for the whole 10 day period. The first order curves for the 92.12.28 and the 93.02.14 cases have similar values, while the linear term has smaller values for the 93.09.06 case. This reflects the fact that the NH flow is more unstable during cold than warm seasons (*Buizza and Palmer*, 1994), as confirmed by the amplification factor of the most unstable perturbations for the three cases (Fig 9).

If, in Fig 8, we consider the time when the second order curve crosses the first order curve, we can see that the 93.02.14 and the 93.09.06 estimates are slightly larger than the 92.12.28 estimate. Considering these two cases, we can estimate the time when the second order curve has half the value of the first order curve to be between forecast day 2 and 2.5, in agreement with the 92.12.28 results.

We now estimate non-linearity of optimal perturbations time evolution applying the second method based on eq (12). Moreover, we consider not only the normal-amplitude and large-amplitude ensembles, but also results of ensembles run with the smallest perturbation amplitude. Since the $l(t)$ curves of the three cases are similar, only mean curves computed averaging over the three cases are shown. For each forecast time and each experiment type we computed the average among the three cases of the maximum and of the minimum values of $l(t)$, and the average of the mean values of the $l(t)$ functions. In mathematical terms, for each of the three configurations, we computed

$$l_{MINSP}(t) = \frac{1}{3}\left(\max[\, l_{921228}(t) \,] + \max[\, l_{930214}(t) \,] + \max[\, l_{930906}(t) \,] \right), \tag{13a}$$

$$\bar{l}(t) = \frac{1}{3}\left(\bar{l}_{921228}(t) + \bar{l}_{930214}(t) + \bar{l}_{930906}(t) \right), \tag{13b}$$

$$l_{MAXSP}(t) = \frac{1}{3}\left(\min[\, l_{921228}(t) \,] + \min[\, l_{930214}(t) \,] + \min[\, l_{930906}(t) \,] \right), \tag{13c}$$

where the **min[...]**, the **max[...]** and the mean $\overline{...}$ in eqs. (13) have been evaluated among the 16 respective $l(t)$ functions. Thus $l_{MINSP}(t)$ is the curve of minimum spread, and $l_{MAXSP}(t)$ is the curve of maximum spread.

For clarity, we plotted the mean and the extreme curves on different panels of Fig 10: panel a) shows the averaged mean curves, while panel b) shows the averaged extreme curves.

A reference line corresponding to 0.7 correlation coefficient has been drawn in Fig 10. Considering the two fields difference between the perturbed and the control forecasts, a 0.7 correlation coefficient means that half of the squared norm of one field can be explained by taking its projection onto the other field. We can estimate the time limit τ_{LIN} either by considering the time at which the function $\bar{l}(t)$ crosses the reference value of 0.6 correlation coefficient (see sub-section 2.2), or when it crosses the reference line corresponding to 0.7 correlation coefficient.

First, let us consider the normal-amplitude results. We can see that the $\bar{l}(t)$ curve crosses the threshold line 0.7 around forecast day 2.5 (Fig 10a), while the two extreme curves cross it slightly before forecast day 1.5 and at forecast day 3 (Fig 10b). Let us consider as time limit for the normal-amplitude experiments the value computed using the mean curve, i.e. $\tau_{LIN} \approx 2.5$ ***days***. A slightly longer time limit can be estimated by considering the reference value of 0.6 correlation coefficient.

Let us now consider the results of the ensembles run in the other two configurations. The times for which the large-amplitude curves cross the 0.7 line are smaller than the normal-amplitude values. In particular,

the mean spread curve for the large-amplitude ensembles has very similar values to the minimum spread curve relative to the normal-amplitude ensemble until forecast day 3. This indicates that multiplying the rms perturbation amplitudes by a factor $\sqrt{2}$ reduces τ_{LIN} by around 12h. If we consider the mean spread curve computed for the ensembles run in the small-amplitude configuration, we can see that it crosses the 0.7 line just before forecast day 3.

The results of the normal-amplitude ensembles can be used to verify the validity of the methodology applied to compute the 36h optimal perturbations. Fig 10 shows that $\bar{I}(t=36h) \approx 0.9$. This indicates that the T63/L19 time evolution of the 36h optimal perturbations, when the amplitude is set as in the normal-amplitude configuration, is dominated by linear dynamics up to optimisation time. Optimal perturbations are computed in the linear approximation at a lower resolution (T21/L19 instead of T63/L19), using only a simplified vertical diffusion and surface drag scheme. So, we cannot conclude from these results that the T21/L19 linear evolution of the optimal perturbations is a good approximation of their T63/L19 non-linear time evolution. We can only conclude that up to forecast day 2-2.5 the T63/L19 dynamics of optimal perturbations generated from T21/L19 singular vectors is dominated by linear processes, given the initial amplitude of our perturbations.

Concluding this section, the analysis of the correlation between time evolutions of perturbed forecasts generated adding and subtracting the same perturbation to the control initial conditions, confirm the estimate $\tau_{LIN} \approx 2\text{-}2.5$ *days* obtained by Taylor expansion for the normal-amplitude perturbations. Moreover, they show that non-linearity dominates the perturbations time evolution after forecast day 6 (see Fig 10a). The multiplication of the rms perturbation amplitude by a factor $\sqrt{2}$ decreases this limit to less than 2 days, while a decrease of the rms amplitude by the same factor increases this limit to almost 3 days.

It is worth stressing that these estimates are based on singular vectors optimized over a 36h time interval, and on a small sample of 3 ensembles. Experimentation has shown that longer values for τ_{LIN} are obtained when considering singular vectors optimized over longer time periods. In fact, as shown by *Buizza* (1994b), singular vectors optimized over different time intervals are different. Consider two sets of singular vectors, computed with optimisation time intervals t and $t+\Delta$. The singular vectors computed for an optimisation time interval $t+\Delta$ are sub-optimal for shorter time intervals. By definition, the singular vectors computed with an optimisation time interval t grow faster than the others up to time t, and so enter the non-linear regime earlier, provided the time t is not so far from the time non-linearity becomes important.

4. EPS SENSITIVITY TO PERTURBATION AMPLITUDE

In this section we study the impact of the perturbation amplitude on the ensemble spread and skill scores.

4.1 Medium skill case (92.12.28): sensitivity of individual perturbations to initial amplitude

First, we focus on the impact of increasing or reducing the perturbation amplitude on the EPS prediction for the 92.12.28 case, in particular over the European area (this case has been chosen among the three situations because it is characterized by moderate forecast skill).

Figs 11, 12 and 13 show, respectively, the 7-day 500 hPa geopotential height forecasts generated by the EPS run in the normal-amplitude, large-amplitude and small-amplitude configurations. Table 3 summarizes the results of the three ensembles. The first two columns of the first line show the control skill score computed for the NH and the European region. Then, for each configuration, the third and the fourth columns indicate how many members of the ensemble are characterized by skill scores higher than the control skill score over the two areas. The next three columns list, respectively, the number of the ensemble member with the best performance in terms of ACC for the NH, its ACC skill score for the NH and its ACC skill score for the European area. The last three columns show, respectively, the number of the ensemble member with the best skill score over Europe, and its NH and European skill scores.

Considering the normal-amplitude ensemble, Table 3 shows that 23 perturbed forecasts performed better than the control if the skill score is computed over the NH, while 15 performed better if we restrict attention to the European region. The forecast with the best NH performance is member 21, with an ACC skill score of 0.74. This is not the best if we restrict the region of interest to Europe, since member 8 performs better with an ACC skill score of 0.82 compared to the 0.68 value scored by member 21. In fact, looking at Fig 11, we can see that the flow predicted over the European region by member 8 is more similar to the analysis than the flow predicted by member 21 (note that the area shown in the maps is smaller than the European area used to compute the ACC). This is particularly true over the British Isles, Spain and central Europe, although the closed low over the Mediterranean Sea is still missing in member 8. The prediction by member 17 is very similar to the forecast by member 8, and indeed their skill scores are very similar, member 17 having a skill score of 0.81 in term of ACC. The worst forecast is member 10, characterized by a completely wrong zonal flow over central Europe (ACC skill score -0.15). Table 3 summarizes the impact of perturbation amplitude on the EPS results, while Figs 11, 12 and 13 show the differences in the flow fields predicted for the 500 hPa geopotential height. Although the impact of the increase of the perturbation amplitude by a factor $\sqrt{2}$ is not very large, we can see that the large-amplitude ensemble is characterized by the worst results, whilst the normal-amplitude and the small-amplitude ensembles have very similar skill scores.

The comparison among the ensemble forecasts of Figs 11, 12 and 13 characterized by the same sequence number shows the impact of perturbing the control initial conditions along the same direction but varying the magnitude of the perturbation. In some cases this has almost no impact: see for example forecasts with sequence number 3, which have a skill score over Europe 0.70, 0.69 and 0.71, respectively, in the small-amplitude, normal-amplitude (reference) and large-amplitude ensembles. In other cases, such as forecasts number 22, an increase of the perturbation amplitude can reduce the skill score from 0.66 (small-amplitude ensemble), to 0.51 (normal-amplitude ensemble) and further down to 0.40 (large-amplitude ensemble) corresponding to a worsening of the forecast over central Europe. For other directions, on the contrary, an increase of the perturbation magnitude can improve the forecast. For example forecasts number 6, whose skill score increase from 0.63 (small-amplitude ensemble), to 0.64 (normal-amplitude ensemble) and up to 0.73 (large-amplitude ensemble) over the European region.

The comparison of Figs 11, 12 and 13 confirms the general impression that, on average, the impact of the perturbation amplitude on the ensemble skill score is small, in particular over Europe, at forecast day 7. Fig 14 shows the mean spread and skill scores of the normal-amplitude (solid lines), the large-amplitude (dashed lines) and the small-amplitude (dotted lines) ensembles. The comparison of the three mean skill scores over Europe (panel d) confirms the comments above. The impact of the perturbation amplitude on the mean score is slightly more evident if we consider the NH (see panel c), with the small-amplitude ensemble characterized by slightly higher skill scores than the other two ensembles.

4.2 **Medium skill case (92.12.28): sensitivity of ensemble spread and skill to initial amplitude**

If we now consider the spread of the ensembles for the same case, both the curves relative to the NH (Fig 14a) and the European area (Fig 14b) indicate that the mean spread increases if the perturbation amplitude increases. Fig 15 shows, for each ensemble, the two extreme curves with the largest and the smallest values of the spread and skill scores.

From this figure we can see that for the 92.12.28 case increasing the perturbation amplitude increases the NH mean spread of the ensemble, but it has a small impact on the ensemble skill, with the large-amplitude worst-skill-score curve showing the smallest values. Considering the European region, the impact on the mean spread is similar: the fc time at which the mean spread curve crosses the 0.6 ACC line decreases by 12-18 hours when multiplying the perturbation amplitude by a factor $\sqrt{2}$. The comparison among Figs 11, 12 and 13 indicates that each of the three ensembles offer a range of probable forecast states which can include the analysis as a plausible member of the ensemble.

4.3 Medium skill case (92.12.28): effect of initial amplitude on individual members

Before considering the results obtained for the other two cases it is useful to analyze the impact of the perturbation amplitude on three particular ensemble members. These three members have been chosen because perturbations of the control initial conditions along these three directions of the phase space of the system give the best forecasts over Europe among the respective ensemble members (see table 3). These are members number 8 (best among the normal-amplitude ensemble members), number 17 (best among the small-amplitude ensemble members) and number 32 (best among the large-amplitude ensemble members). Table 4 shows the NH and European skill scores of ensemble members 8, 17 and 32, when run in the three configurations. Perturbing the control initial conditions along the direction identified by member 8 gives the best European skill score when the amplitude is fixed to the reference value (normal-amplitude ensemble). Adding this perturbation with a smaller amplitude produces a slightly worse forecast (0.80 ACC instead of 0.82), while multiplying it by a factor $\sqrt{2}$ has a more negative impact (0.66 ACC skill score). The initial perturbation of member 8 (Fig 16) with normal-amplitude has values smaller than 12.5 m in terms of geopotential height at 500 hPa. Fig 17 shows the difference between the control and the perturbed forecast (left panels) and the difference between the analysis and the perturbed forecast (right panels), in terms of 500 hPa geopotential height, at forecast day 7, of ensemble member 8 run in the small-amplitude configuration (panels a) and b)), in the normal-amplitude configuration (panels c) and d)) and in the large-amplitude configuration (panels e) and f)). (We remind the reader that figs. 11, 12 and 13 show the full fields.) In other words, left panels of Fig 17 show the field to be added to the perturbed member to obtain the control, and right panel of Fig 17 the field to be added to the perturbed member to obtain the analysis.

All the forecasts are characterized by too high values over south-western Europe and by too low values over northern and eastern Europe (see right panels of Fig 17). The 7-day difference between the analysis and the perturbed forecast has maximum values up to 220 m for the small-amplitude forecast, 180 m for the normal-amplitude forecast and 300 m for the large-amplitude forecast. Looking at the difference between the control and the perturbed forecast we can see that increasing the perturbation amplitude produces largest differences between the control and the perturbed forecast. Multiplying the initial perturbation by a factor 2 produces in this case differences up to a factor 1.9 after 7 days of non-linear integration.

Considering one particular configuration, e.g. the normal-amplitude, it is possible to trace back the areas of origin of the difference appearing over Europe at forecast day 7 using the field difference between the control and the perturbed forecast for different forecast times (not shown). Results show that it is due to a combined action of the perturbation added over north America and eastern Pacific, but that it is dominated by the propagation of the positive-negative pattern localized across the Rockies (see Fig 16).

Similar considerations can be drawn comparing ensemble members 17 and 32. Table 4 confirms the conclusions drawn from Fig 15d that the skill scores of the best among the ensemble members in the small-amplitude and in the normal-amplitude configurations are higher than the skill score of the best among the large-amplitude ensemble members.

4.4 Three case studies: average results

We now consider the results from the ensembles run with the other two starting dates (93.02.14 and 93.09.06). For reason of space we will not analyze in detail the normal-amplitude, large-amplitude and small-amplitude results. We have computed averages over the three cases of the mean skill scores, the best-skill-scores and the worst-skill-scores. Moreover, we have computed averages among the three cases of the mean, the smallest and the largest spread curves.

Fig 18 shows the average mean spread and mean skill score curves computed for the NH and the European region. Generally speaking, Fig 18 is consistent with the results of the 92.12.28 case. The spread increases with the perturbation amplitude in both regions. The impact on the skill score is smaller and it indicates that reducing the perturbation amplitudes slightly increases the EPS mean skill score.

Fig 19 shows the average of the extreme spread and skill curves. Figs 19a and 19b show that the perturbation amplitude has a little impact on the minimum spread curve (high ACC), but has a measurable effect on the maximum spread curve (low ACC). Similar considerations can be drawn considering the extreme skill score curves: the perturbation amplitude has a little effect on the most skilful forecast, but a large effect on the skill of the least skilful forecast.

4.5 Three case studies: simulation of a reduction of analysis errors in a perfect model

Let us suppose that the forecast model were perfect. We can consider the control trajectory as the "truth" and the perturbations added to the control initial conditions as analysis errors. Then the comparison of the mean spread score curves can give an indication of the impact of errors in the initial conditions, represented by the optimal perturbations, on the forecast skill. (As we have seen in sub-section 3.2, the amplitude of the perturbations in the normal-amplitude configuration is comparable to the amplitude of the estimated analysis error.)

The comparison among the mean spread curves of Fig 18 can be used to estimate the impact on the forecast skill of a reduction of the amplitude of the analysis error by a factor $\sqrt{2}$. If we consider the forecast day at which the two mean normal-amplitude and small-amplitude spread curves crosses the reference line of 0.6, we can see that it increases by almost 1.5 day if we consider the NH, and by almost 1 day if we refer to the European region. In particular the estimated gain in NH medium-range predictability is around

1.5 days for the less skilful case (93.09.06), it is around 1 day for the 92.12.28 case (see Fig 14a) and it is more than 1.5 days for the most skilful among the three cases (93.02.14).

Moreover, if we consider Figs 19a and 19b we can see that a reduction of the analysis errors decreases the spread of the ensemble. In particular, it significantly reduces the largest spread curve. This suggests that, on average, smaller analysis error should increase the skill scores of the worst forecasts. In contrast, it only slightly reduces the smallest spread curve, since for the three case studies the simulations of the best forecasts are already nearly perfect in all three configurations.

Another way of looking at these results is to compute the factor by which the analysis error should be reduced to diminish the mean spread of forecast day 7 to the value of day 5. To address this question we ran another ensemble with starting day the 92.12.28, fixing the rms perturbation amplitude to half the normal-amplitude value. Table 5 shows, for the NH, the averaged rms spread and error, for forecast day 5 and 7, computed for the 500 hPa geopotential height field. The first column of Table 5 lists, for the four experiment sets, the ratio of the rms amplitude of the perturbations with respect to the square of the reference amplitude. Consider the experiment sets run with perturbations characterized by rms amplitudes $\sqrt{2}$ and 0.5 times the reference amplitudes. As can be seen by comparing the day 5 and the day 7 spread, a reduction of the rms perturbation amplitude by a factor of $2\sqrt{2}$ is necessary to reduce the spread of forecast day 7 to the value of forecast day 5.

Unfortunately, model deficiencies reduce the possible gain in medium-range predictability, as can be seen comparing the difference between the normal-amplitude and small-amplitude mean spread curves with the difference between the normal-amplitude and small-amplitude mean skill score curves (see Fig 18). The comparison among the mean normal-amplitude and small-amplitude spread curves indicates that the possible gain in predictability is less than 12 hours for the NH (smaller for Europe) when reducing the rms amplitude by a factor $\sqrt{2}$. Fig 18 shows that the distance between the large-amplitude and normal-amplitude mean spread curves is similar to the distance between the normal-amplitude and small-amplitude mean spread curves, while the distance between the normal-amplitude and small-amplitude mean skill score curves is almost half the distance between the large-amplitude and normal-amplitude mean skill score curves. This confirms that model errors become more important when the errors in the initial conditions decrease and reduce the possible gain in predictability that can be achieved by an analysis error reduction. Note that this result is also confirmed by the decrease of the errors with the perturbation amplitude (see Table 5).

error amplitude. Considering the control trajectory as the "truth" (perfect model hypothesis) and the initial perturbations as analysis errors, the spread among the ensemble members can be used as a measure of the impact of analysis errors on the forecast performances. Comparing the mean spread curves of ensembles run in the normal-amplitude and the small-amplitude configurations we have shown that a reduction of the rms amplitude by a factor $\sqrt{2}$ can lead to a gain of medium-range predictability of around 1.5 days if measured over the Northern Hemisphere, and of around 1 day if measured over Europe.

We have also addressed the question of how large the improvement of the analysis error should be to reduce the ensemble spread at forecast day 7 to the value shown at present at forecast day 5. To answer to this question, we have run, for one case only, another experiment set with perturbation amplitudes smaller than the reference value by a factor 0.5. The comparison between the mean spread curves computed for the ensembles run with the largest and the smallest perturbation amplitude, suggests that a reduction of the analysis error by a factor $2\sqrt{2}$ can lead to the required reduction of the forecast day 7 spread.

Note that this may overestimate the sensitivity to error in the initial conditions if there are error components that do not project well onto the fastest growing singular vectors.

These conclusions are based on many (more then 300) T63/L19 10-day non-linear integrations run for three cases characterized by medium, high and poor skill score. Although they should be considered as preliminary, we believe they give useful indications.

ACKNOWLEDGEMENT

Singular vector computation using the Integrated Forecasting System developed in collaboration between ECMWF and Météo-France, and the implementation of the ECMWF Ensemble Prediction System are the results of the work of many staff and consultants during the past years. The author appreciation goes to all of them, and in particular to R Mureau and J Tribbia who contributed to the implementation of the ECMWF Ensemble Prediction System. The author would also like to thank F Molteni and T Palmer for very useful discussions we had during the development of this work, and P Courtier, A Hollingsworth and A Simmons for helpful comments on an earlier version of this paper.

	NH spread max - min	NH skill-scores	
		control	worst - best
92.12.28	0.24 - 0.84	0.52	0.40 - 0.74
93.02.14	0.67 - 0.90	0.72	0.45 - 0.80
93.09.06	0.20 - 0.90	0.36	0.30 - 0.61

Table 1 Spread and skill scores, computed for the NH, of the three analyzed cases.

	An-error	Perturbations
U-comp	4.3 m/s	0.41 m/s
V-comp	4.3 m/s	0.37 m/s
T-comp	1.0 K	0.68 K

Table 2 Root mean square amplitude, averaged over the NH region at model level 11, of the analysis error estimate and of the normal-amplitude perturbations, relative to the 92.12.28 case.

Ampl	Control		Ens better Con		Best over NH			Best over Eur		
	NH	Eur	NH	Eur	n	NH	Eur	n	NH	Eur
norm	0.52	0.60	23	15	21	0.74	0.68	8	0.62	0.82
small			23	17	19	0.72	0.75	17	0.50	0.82
large			20	14	21	0.74	0.59	32	0.53	0.76

Table 3 Skill scores of the normal-amplitude, large-amplitude and small-amplitude ensembles run with starting date 92.12.28. First two columns: control skill scores over NH and Europe. Then, for each configuration: third and fourth columns: number of ensemble members performing better than the control over NH and over Europe. Next three columns: ensemble member performing better than the control over NH, its skill score over NH and over Europe. Last three columns: ensemble member performing better than the control over Europe, its skill score over NH and over Europe.

forecast number	small-ampl		normal-ampl		large-ampl	
	NH	Eur	NH	Eur	NH	Eur
8	0.63	0.80	0.62	0.82(*)	0.53	0.66
17	0.50	0.82(*)	0.49	0.80	0.49	0.72
32	0.66	0.78	0.59	0.72	0.53	0.76(*)

Table 4 Skill scores of ensemble members number 8 (first line), 17 (second line) and 32 (third line) run with starting date 92.12.28. First two columns: skill scores of the small-amplitude ensemble members. Third and fourth columns: skill scores of the normal-amplitude (reference) members. Last two columns: skill scores of the large-amplitude members. For each ensemble member, a star close to the European skill score identifies the configuration in which the member is characterized by the best European skill score.

rms ampl	forecast day 5		forecast day 7	
	spread	error	spread	error
$\sqrt{2}$	67.2	85.9	100.9	104.4
1	58.8	82.2	93.3	100.7
$1/\sqrt{2}$	49.6	80.3	83.0	97.1
1/2	39.8	79.5	69.4	98.2

Table 5 Mean rms spread and rms error (m), computed using 500 hPa geopotential height over NH, for experiments run with starting date 92.12.28. First column: rms perturbation amplitude with respect to the reference value (normal-amplitude configuration). Next two columns: spread and error at forecast day 5. Last two columns: spread and error at forecast day 7.

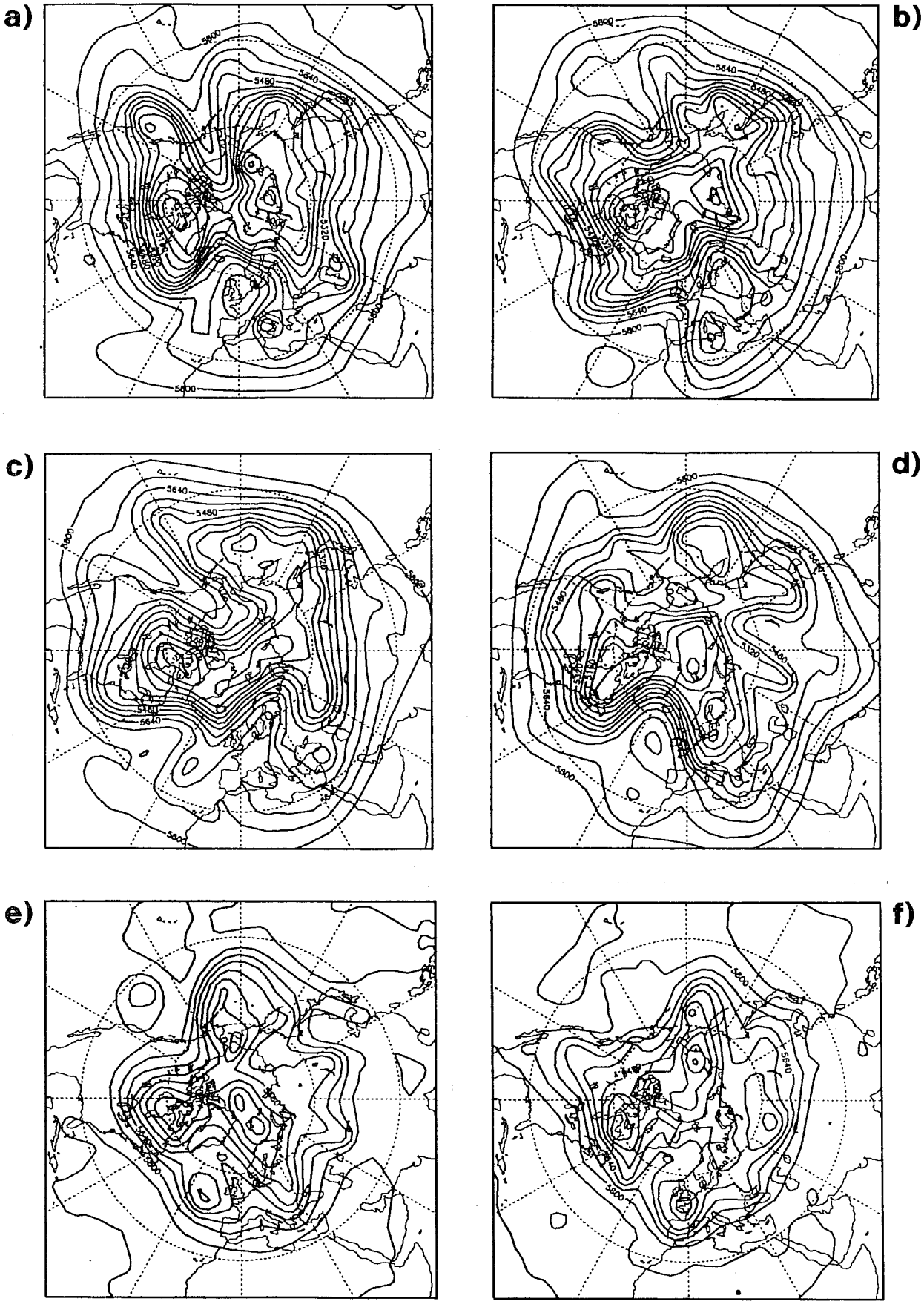

Fig 1 500 hPa geopotential height of the three analyzed cases (all referring to 12GMT). Panel a): 92.12.28; panel b): 93.01.04. Panel c): 93.02.14; panel d): 93.02.21. Panel e): 93.09.06; panel f): 93.09.13. Contour intervals 80 m.

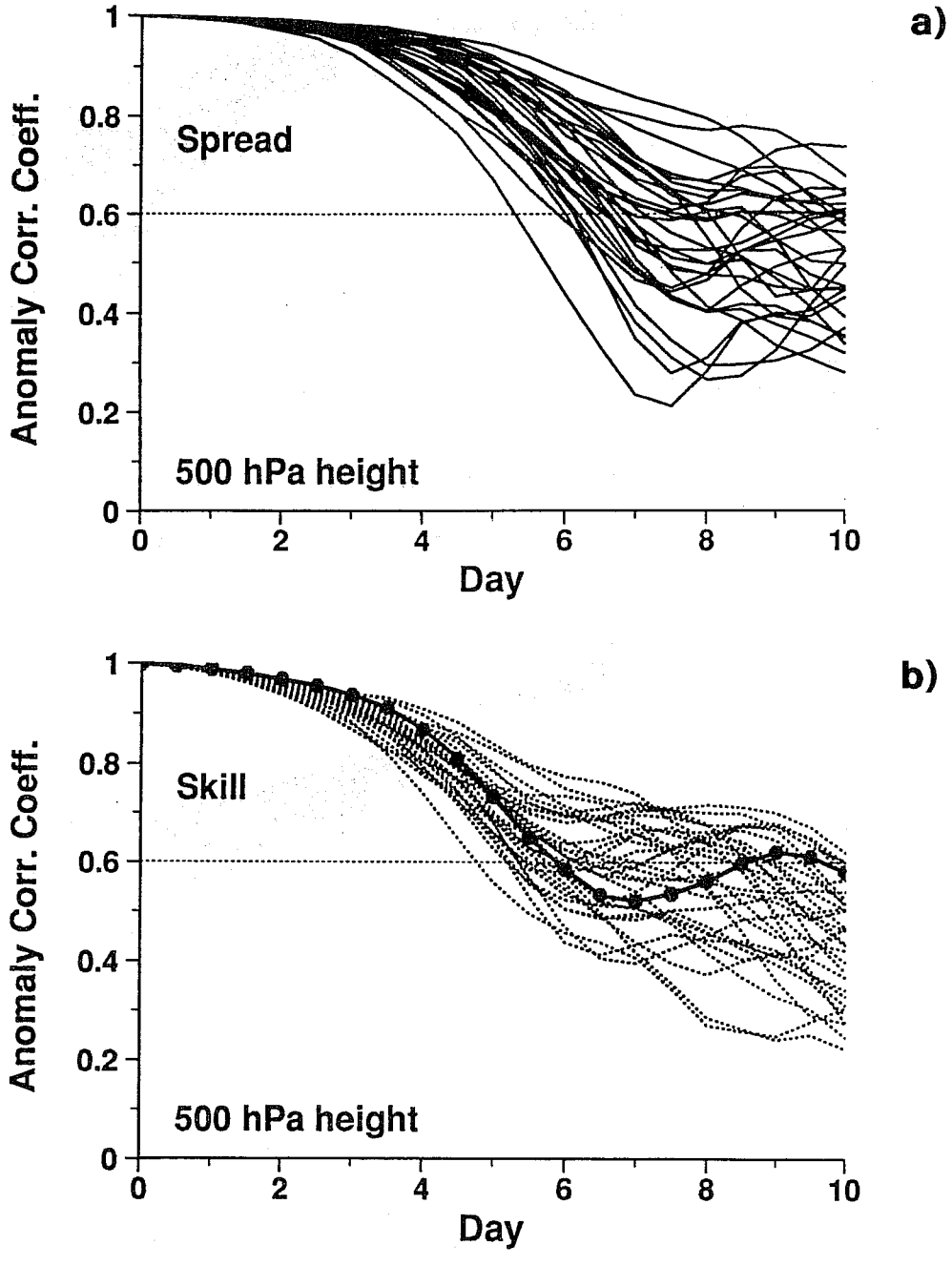

Fig 2 Spread (panel a)) and skill (panel b)) scores of the normal-amplitude ensemble prediction started at 12GMT of 92.12.28, computed using ACC. Panel a): each solid line refers to a perturbed forecast. Panel b): each dotted line refers to a perturbed forecast, while the thick solid line refers to the control.

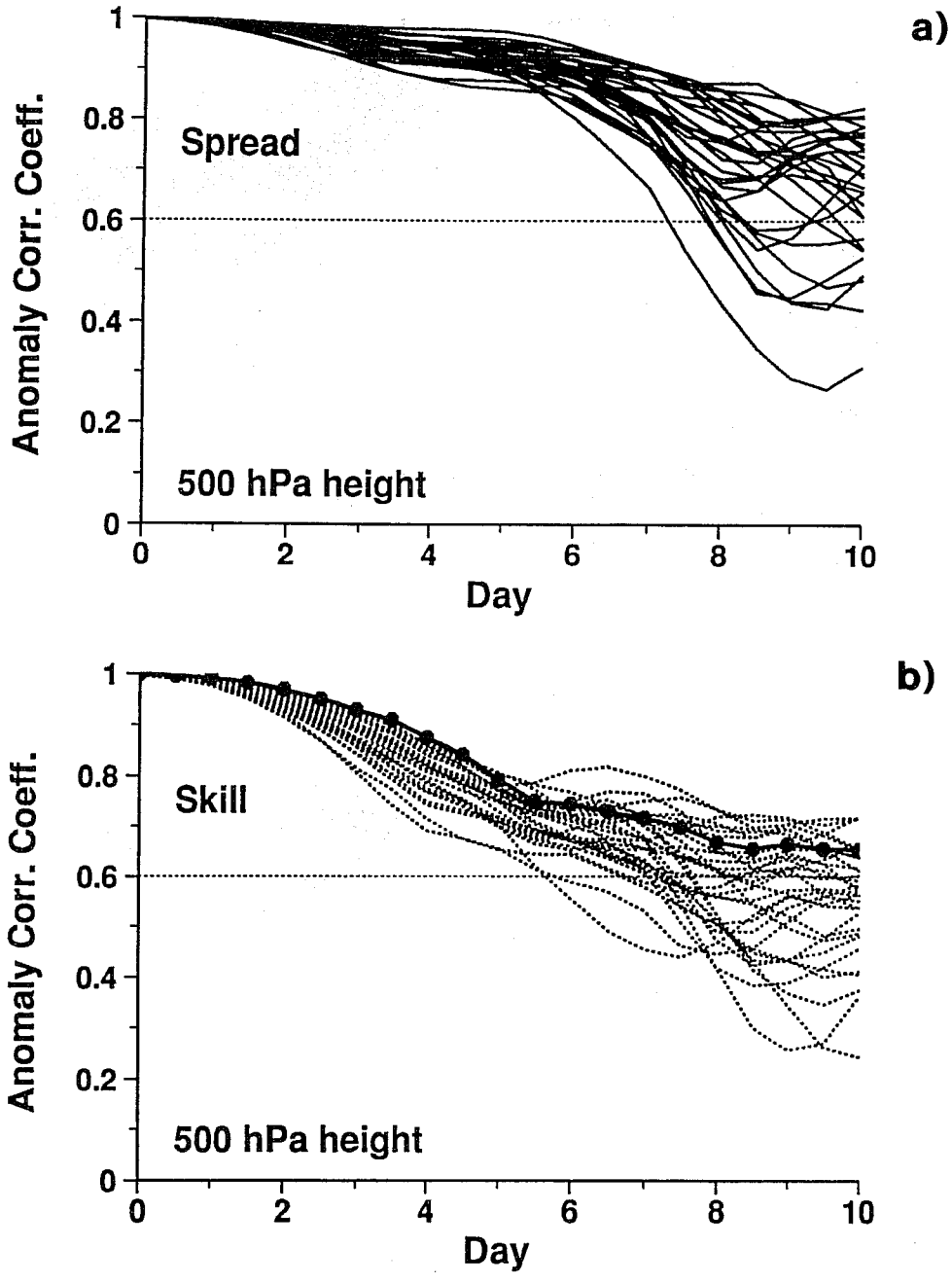

Fig 3 As in Fig 2, but for the normal-amplitude ensemble started at 12GMT of 93.02.14.

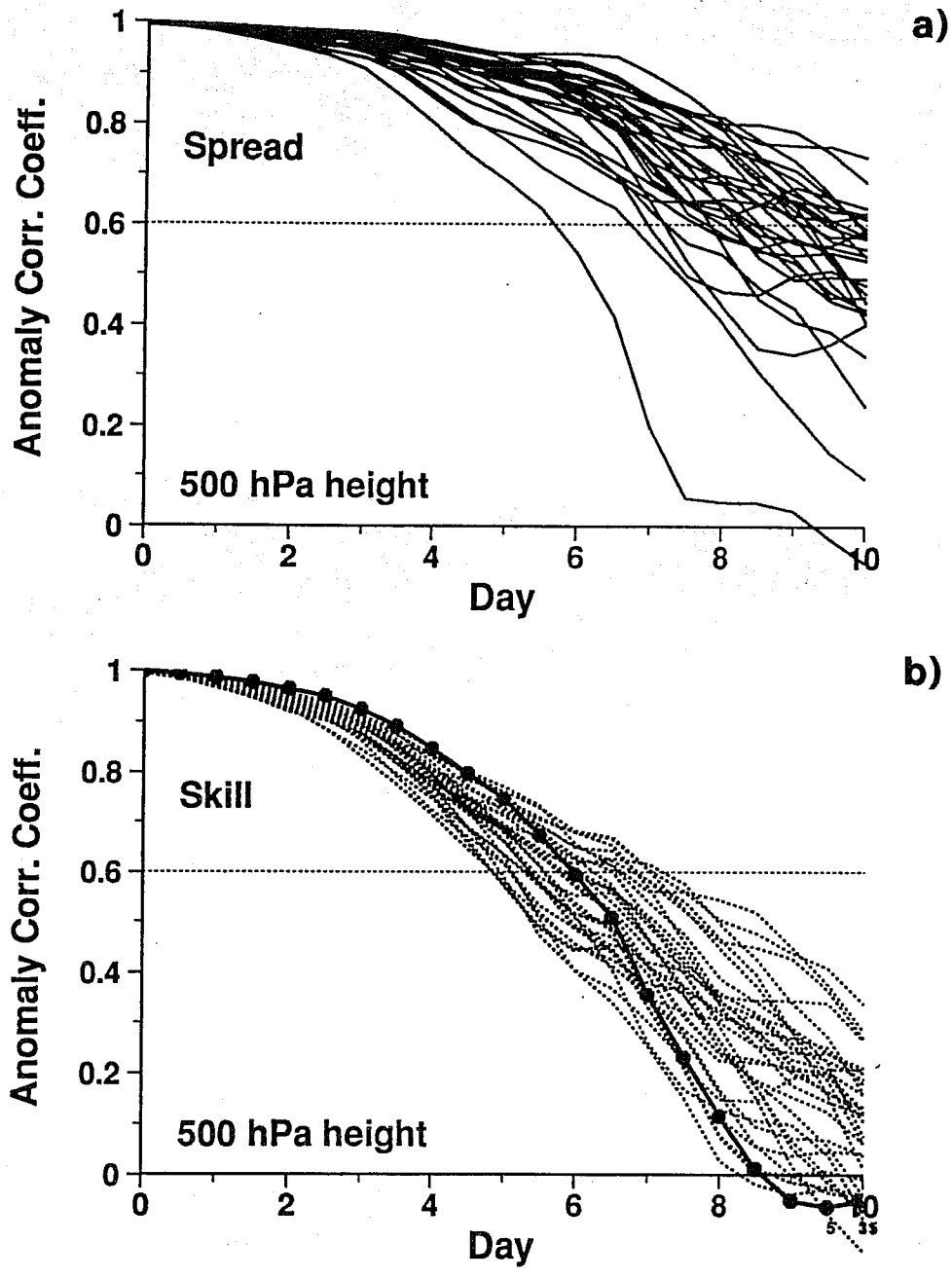

Fig 4 As in Fig 2, but for the normal-amplitude ensemble started at 12GMT of 93.09.06.

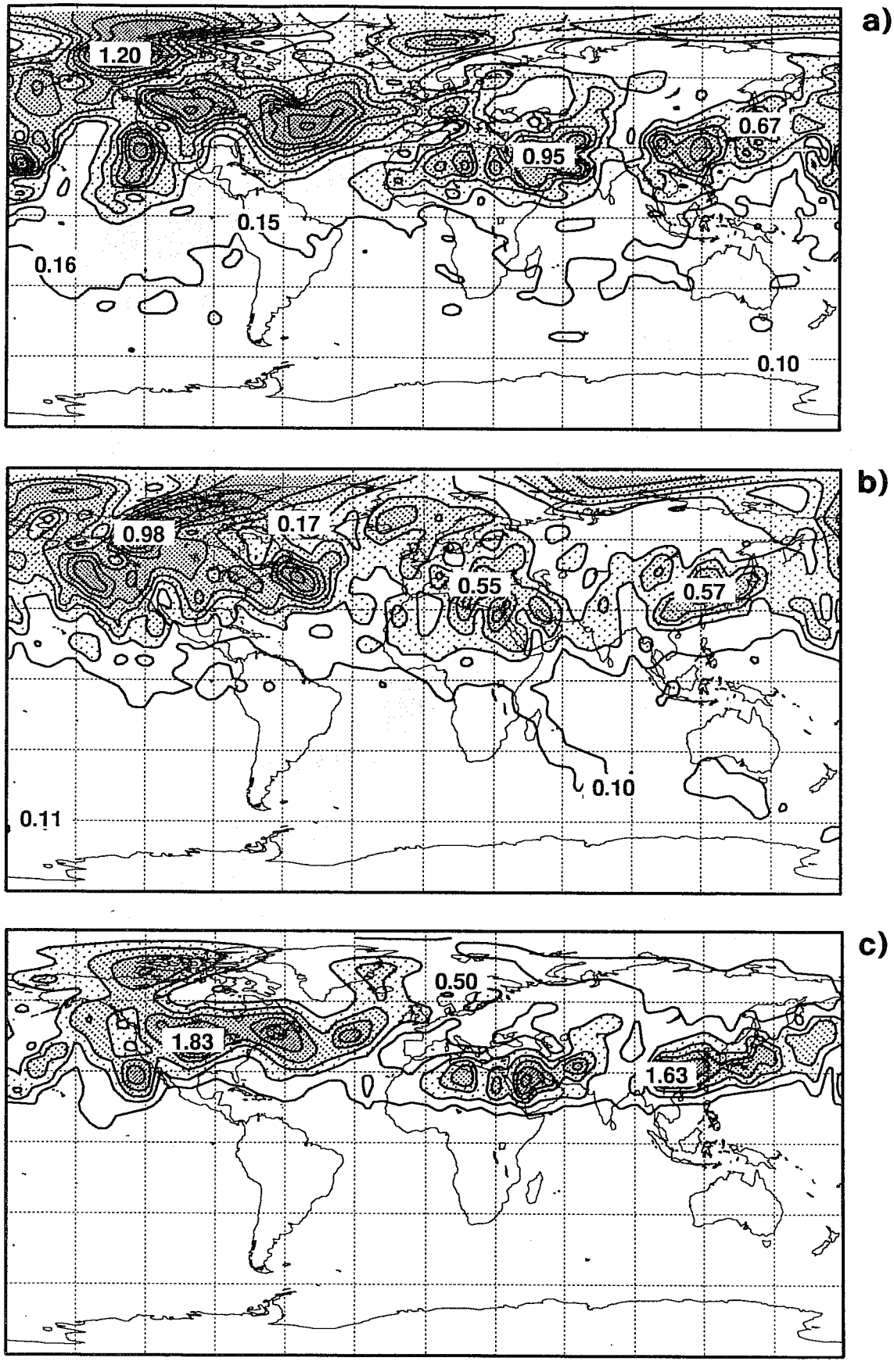

Fig 5 Root mean square amplitude of the normal-amplitude perturbations relative to the 92.12.28 case. U (panel a)) and V (panel b)) components, and temperature (panel c)) are shown at model level 11 (around 500 hPa). Contour isolines 0.1 m/s, starting from 0.05 m/s, in panels a) and b), and 0.25 K, starting from 0.125 K, in panel c).

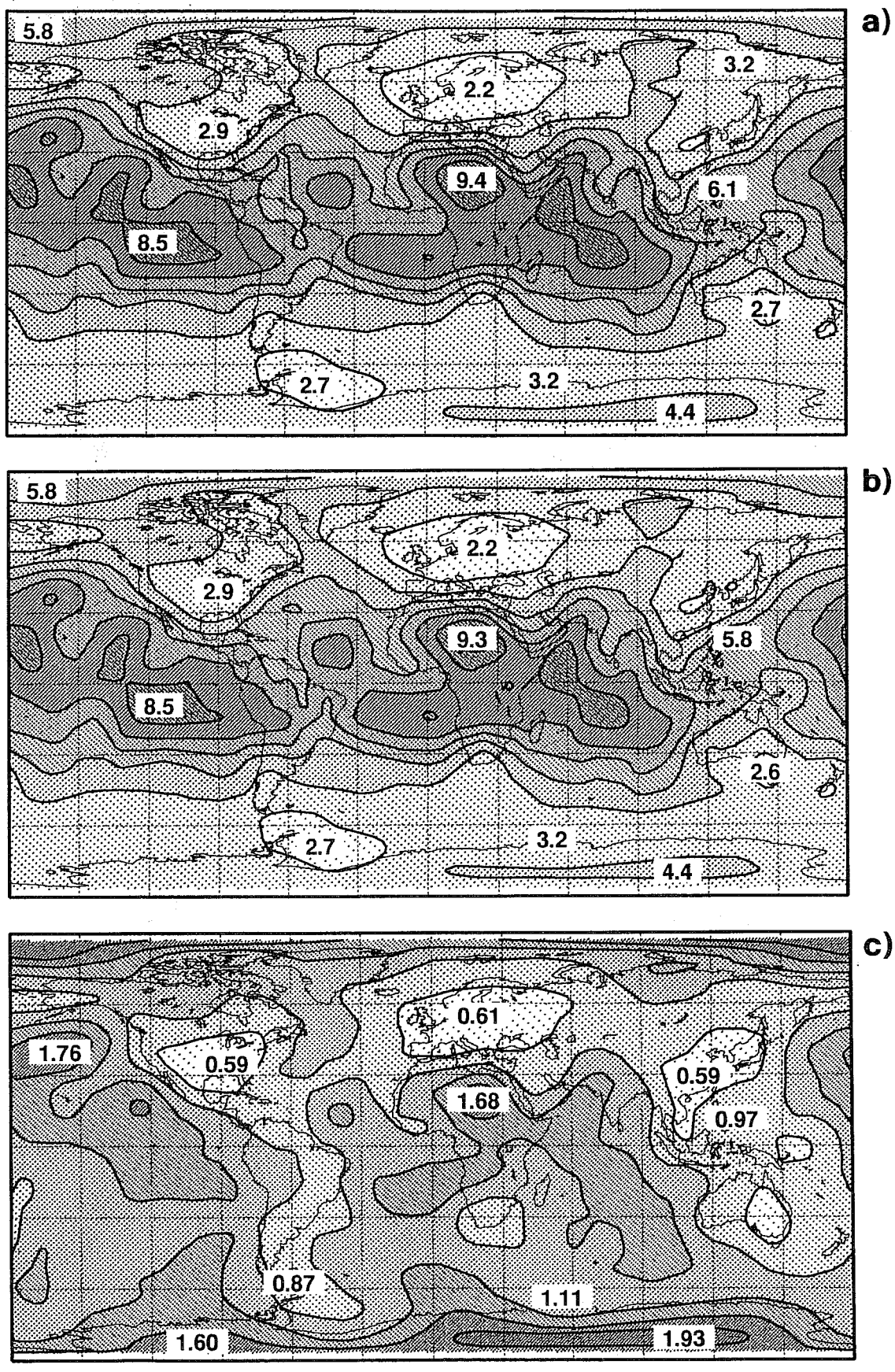

Fig 6 As Fig 5, but for the optimal interpolation estimate of the 92.12.28 analysis error field. Contour isolines 1 m/s, starting from 0.5 m/s, in panels a) and b), and 0.25 K, starting from 0.125 K, in panel c).

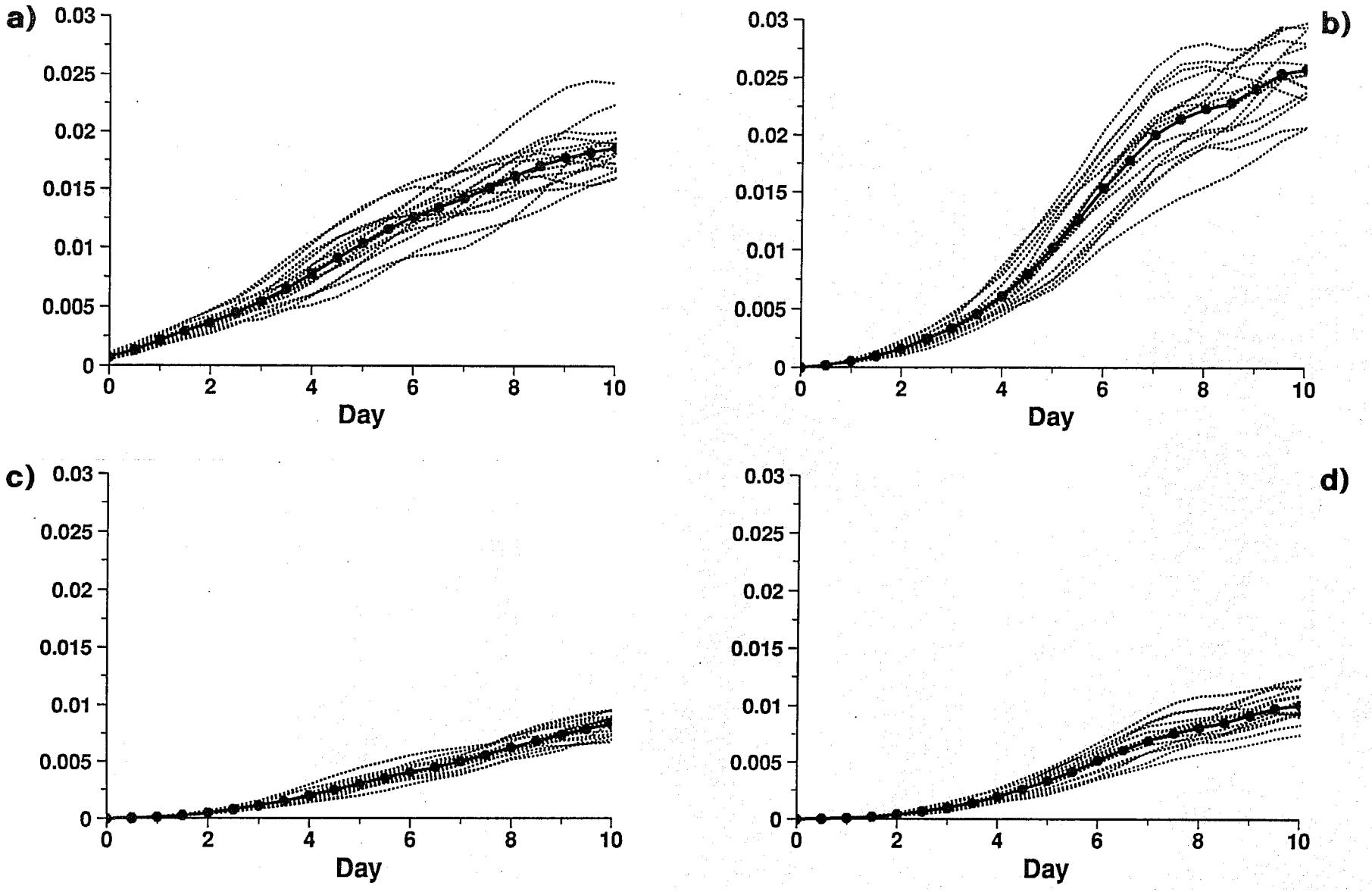

Fig 7 Taylor expansion coefficients $c_n(t)$ $n=1,..,4$ (see eq (14)), computed from the normal-amplitude and large-amplitude ensembles run with starting date the 92.12.28. Dotted lines refer to the 16 coefficient estimates, while the thick solid line refers to the mean coefficients $\bar{c}_n(t)$. Panel a): first order coefficients; panel b): second order; panel c): third order; panel d): fourth order.

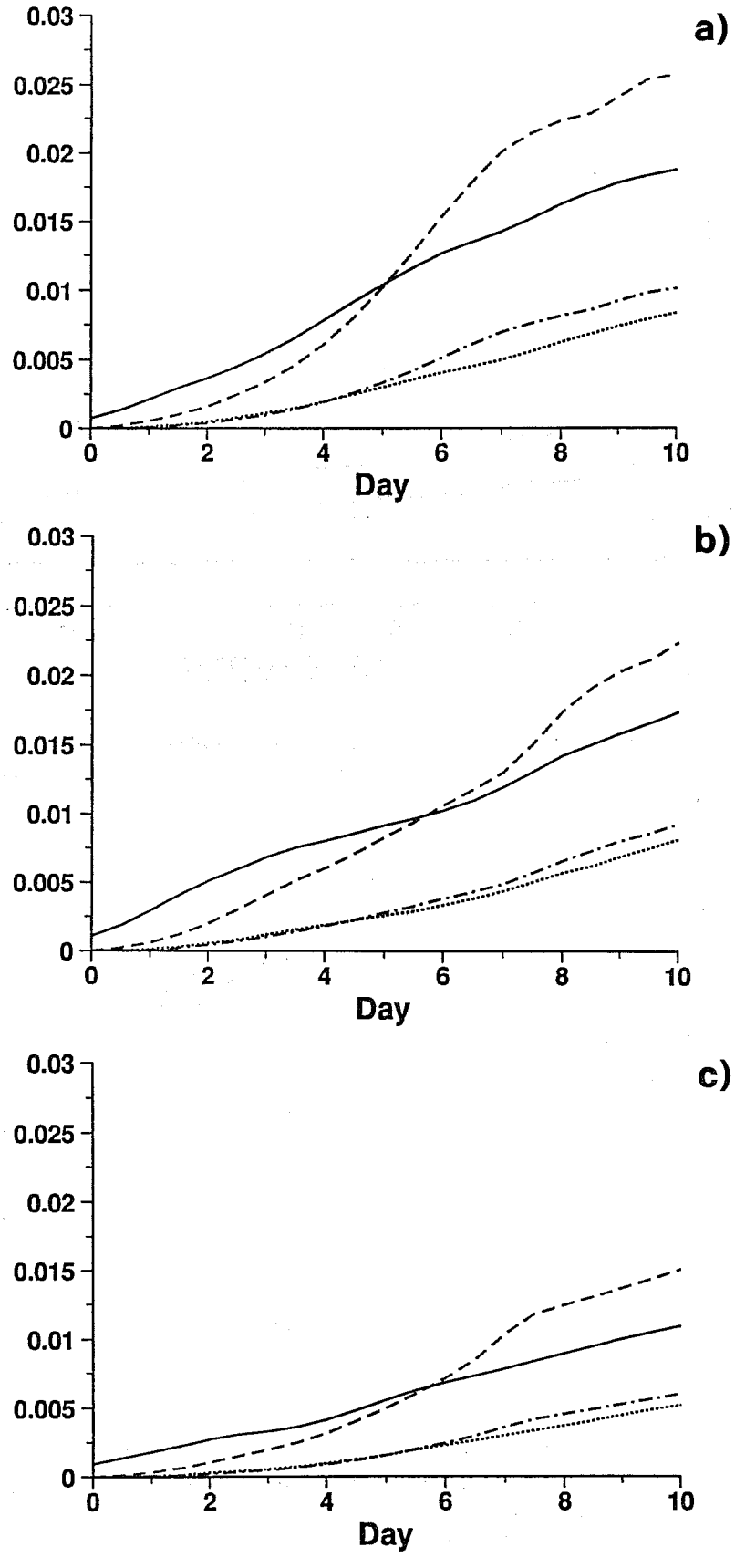

Fig 8 Mean Taylor expansion coefficients relative to the 92.12.28 (panel a)), the 93.02.14 (panel b)) and the 93.09.06 (panel c)) cases. Solid line: first order; dashed line: second order; dotted line: third order; chain-dashed line: fourth order term.

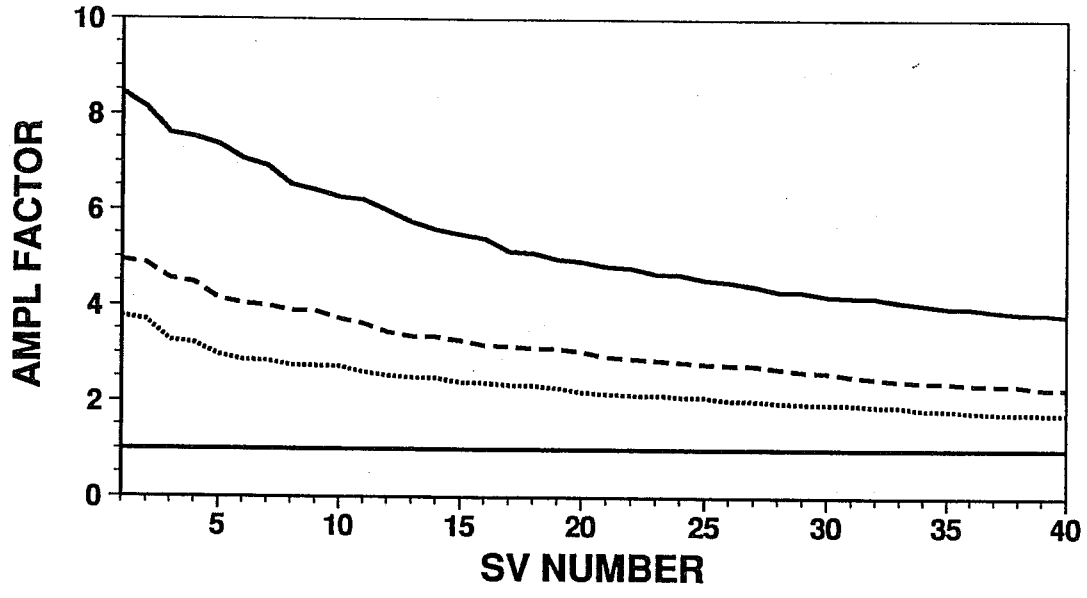

Fig 9 Amplification factors of the most unstable singular vectors, sorted with respect to the amplification factor, for the 92.12.28 (solid line), the 93.02.14 (dashed line) and the 93.09.06 (dotted line) situation.

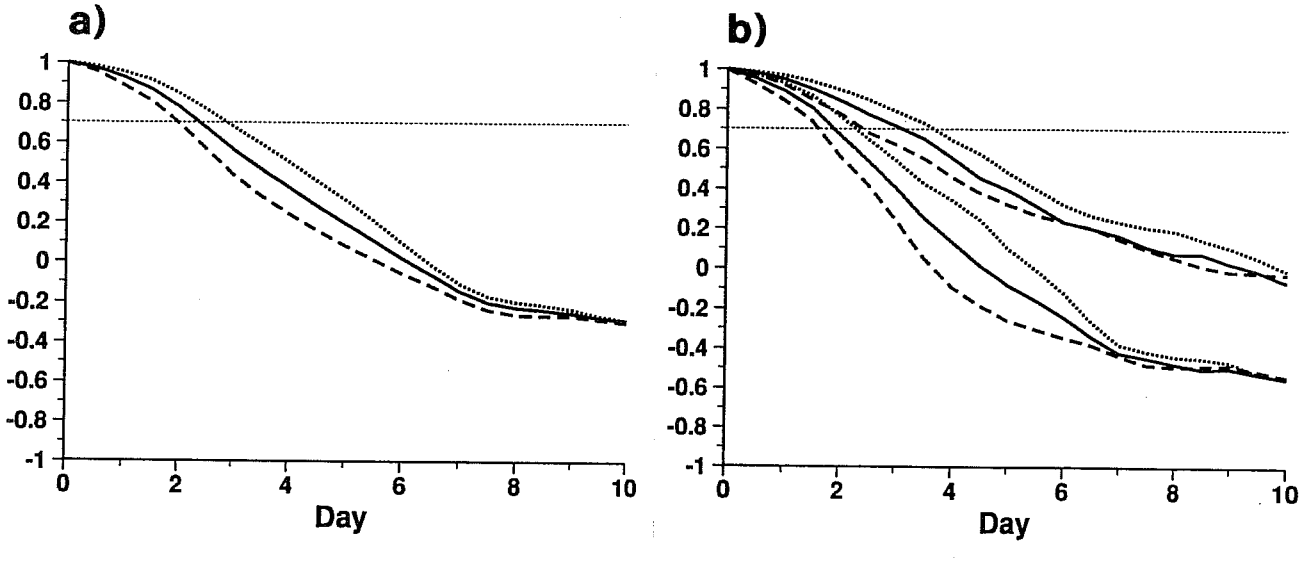

Fig 10 Panel a): functions $\bar{l}(t)$ computed from the normal-amplitude (solid lines), the large-amplitude (dashed lines) and the small-amplitude (dotted lines) ensembles. Panel b): extreme functions $l_{MINSP}(t)$ (upper curves) and $l_{MAXSP}(t)$ (lower curves) computed from the normal-amplitude (solid lines), the large-amplitude (dashed lines) and the small-amplitude (dotted lines) ensembles.

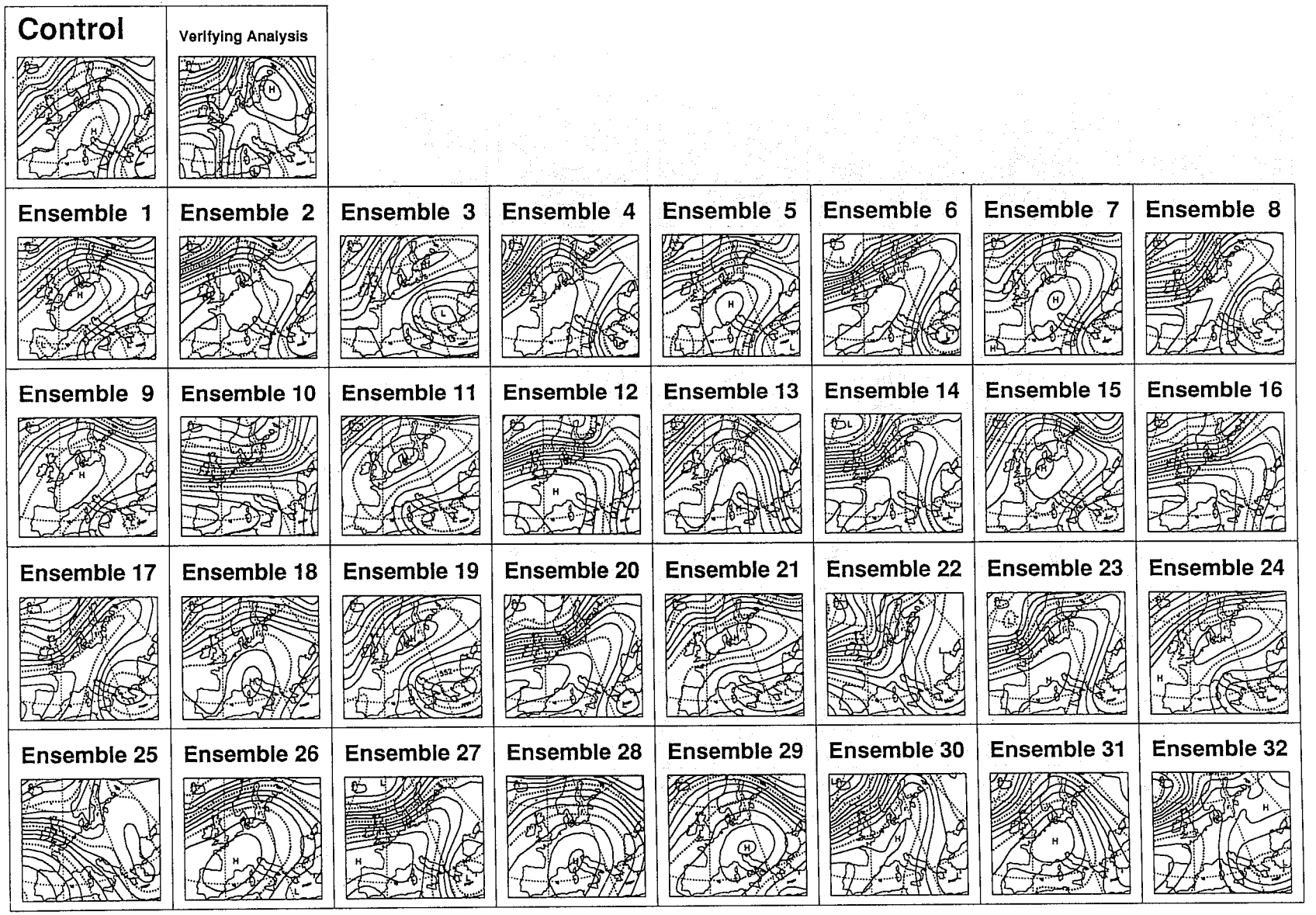

Fig 11 500 hPa geopotential height forecasts predicted over the European region for forecast day 7 by the 92.12.28 normal-amplitude ensemble. First line: control forecast (first panel) and verifying analysis (second panel). All the other panels refer to ensemble members. Contour intervals 60 m.

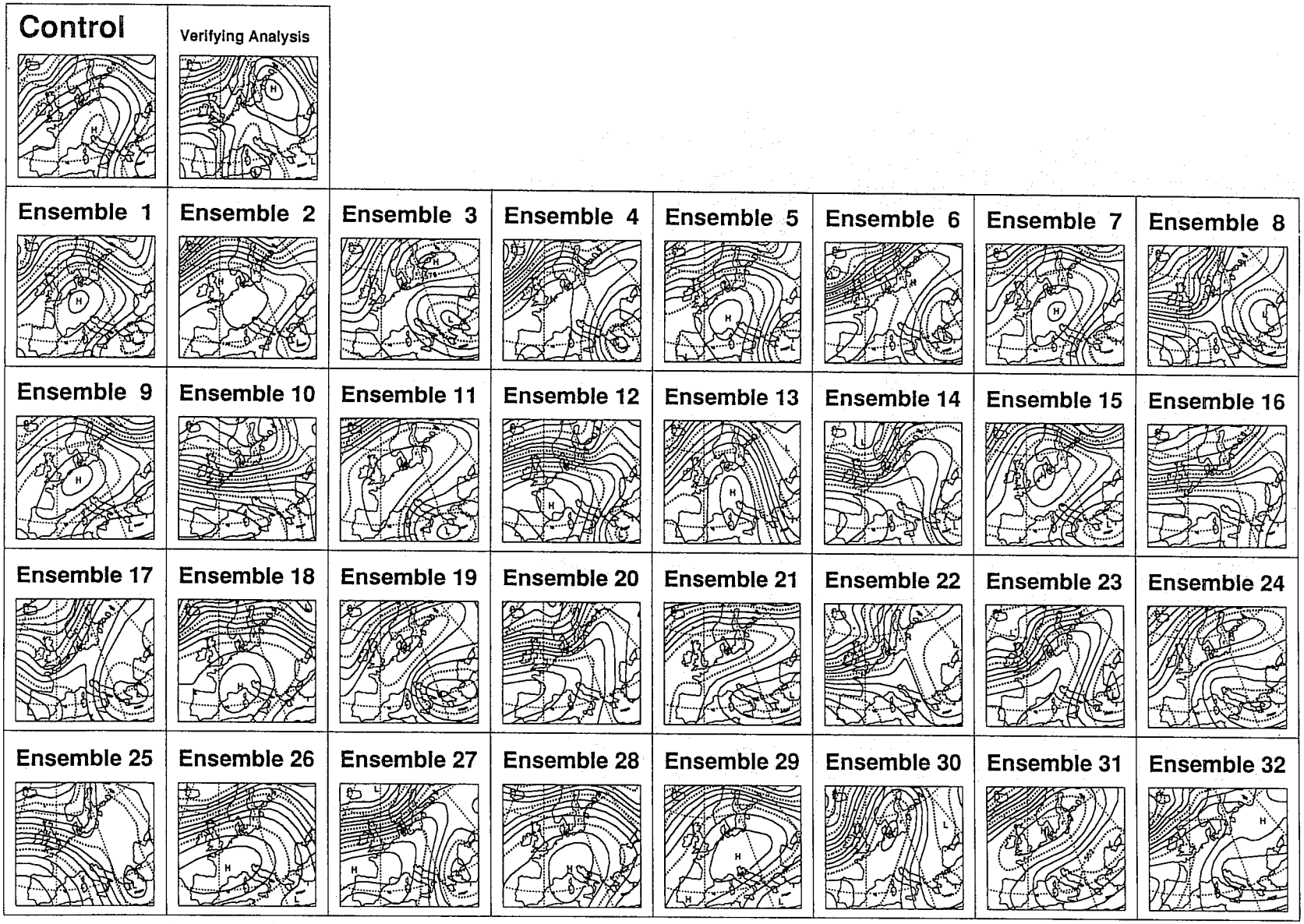

Fig 12 As Fig 11, but for the flow predicted by the large-amplitude ensemble.

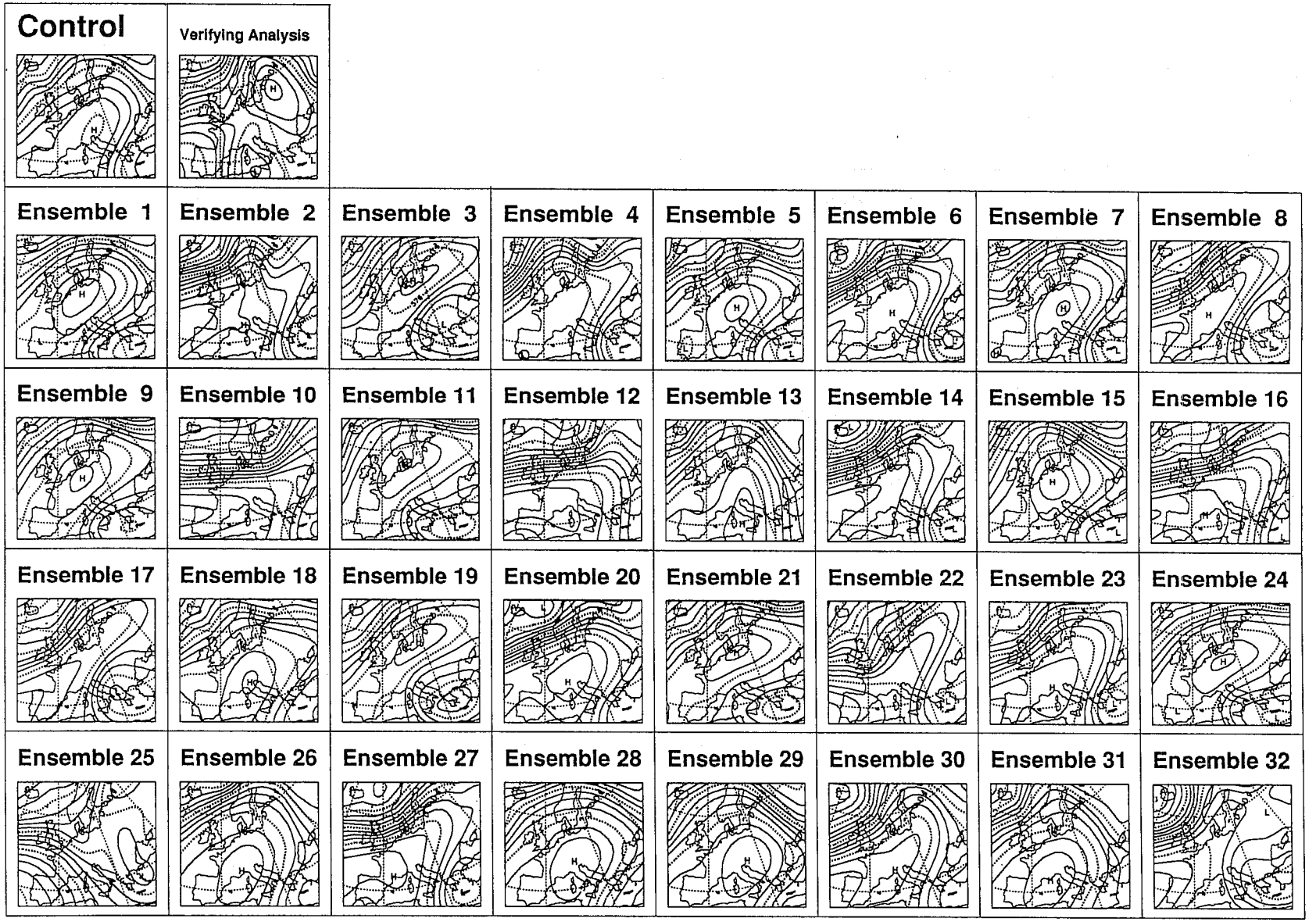

Fig 13 As Fig 11, but for the flow predicted by the small-amplitude ensemble.

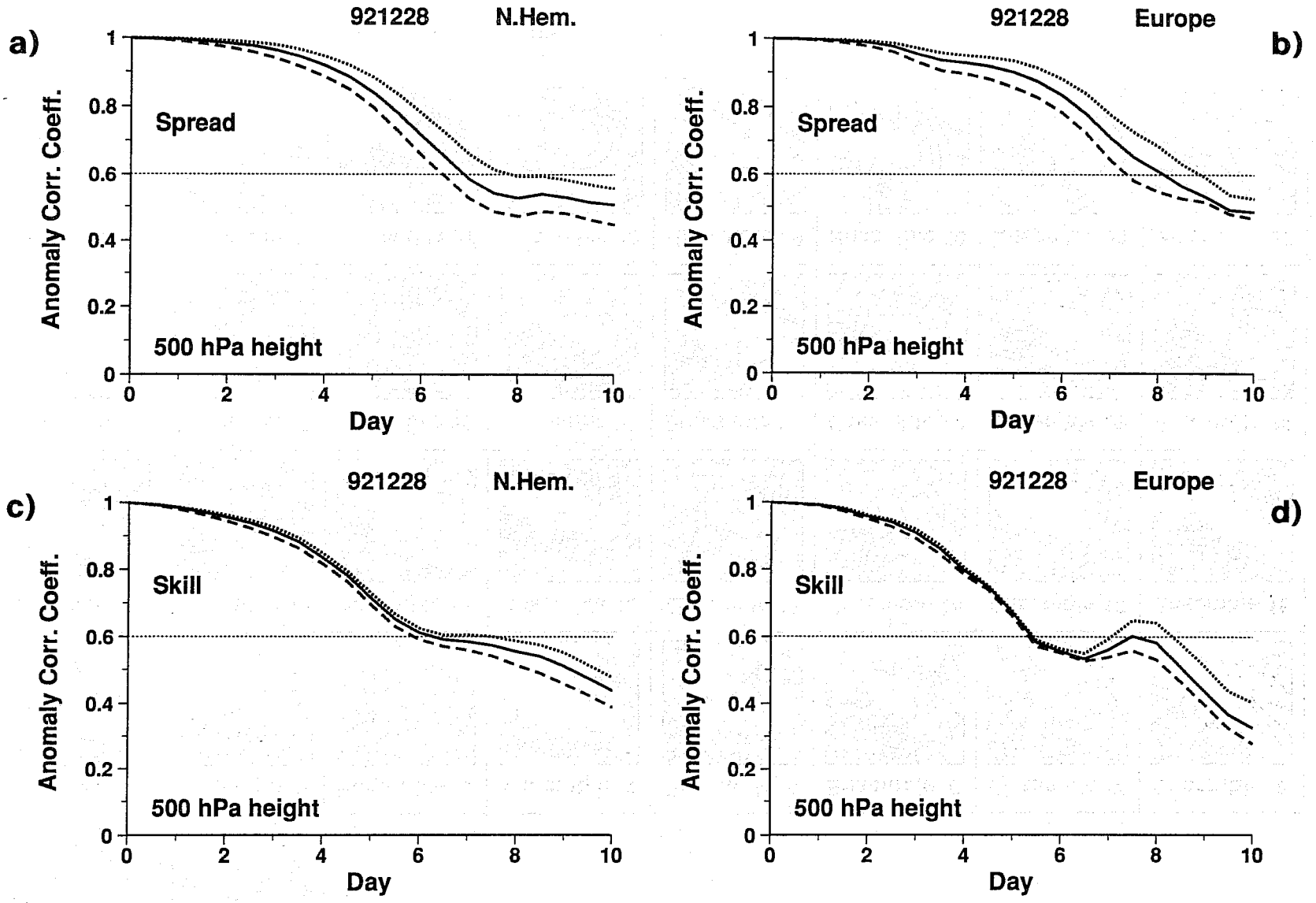

Fig 14 Mean spread and skill score curves averaged among the normal-amplitude (solid line), large-amplitude (dashed line) and small-amplitude (dotted line) ensembles started on 92.12.28. Panel a): NH spread; panel b): spread computed over Europe; panel c): NH skill score; panel d): skill score computed for Europe.

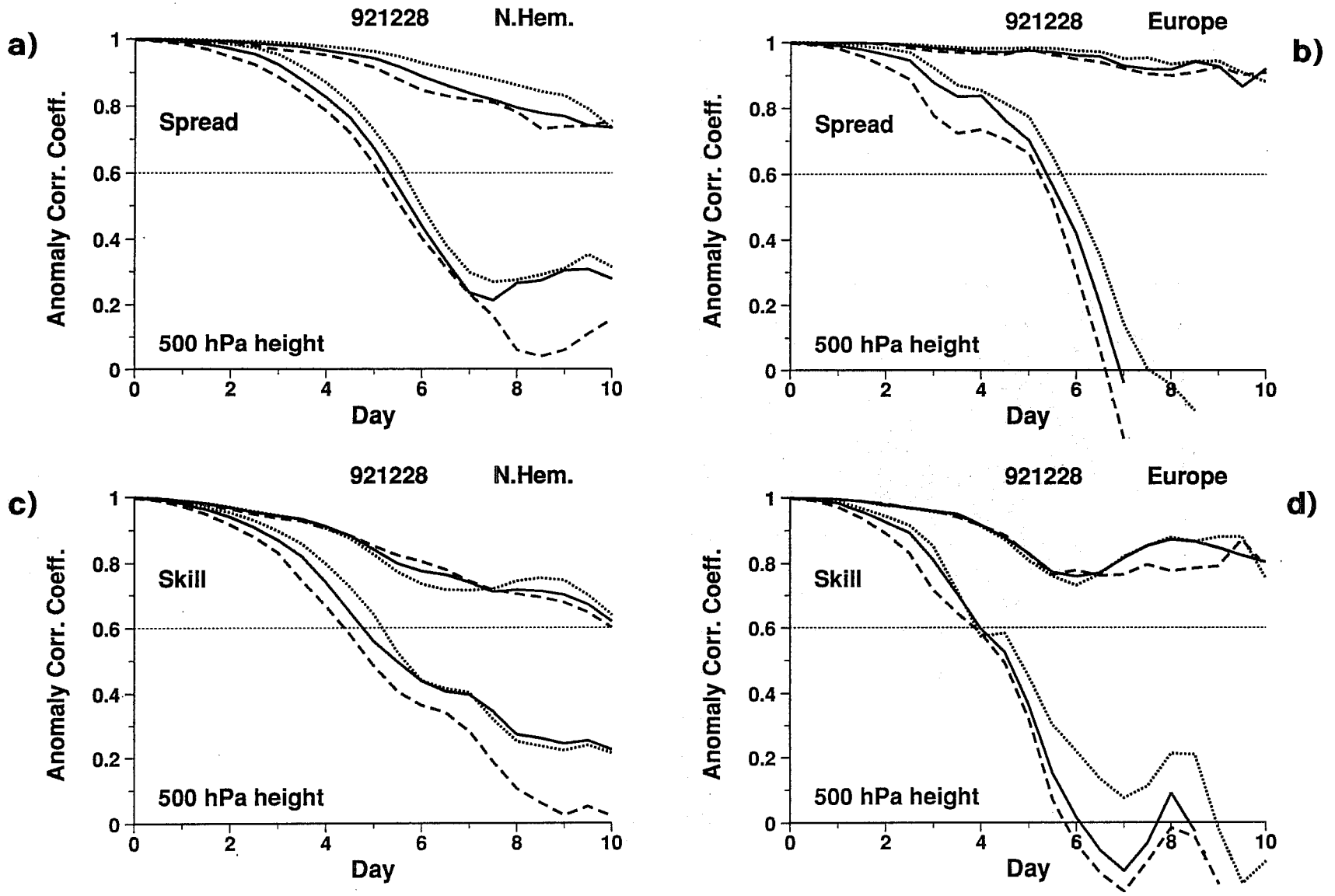

Fig 15 Extreme spread and skill score curves relative to the ensembles started on 92.12.28. Panel a): NH smallest and largest spread curves for the normal-amplitude (solid lines), large-amplitude (dashed lines) and small-amplitude (dotted lines) ensembles. Panel b): as panel a) but for Europe. Panel c): NH best-skill-score and worst-skill-score curves for the normal-amplitude (solid lines), large-amplitude (dashed lines) and small-amplitude (dotted lines) ensembles. Panel d): as panel c) but for Europe.

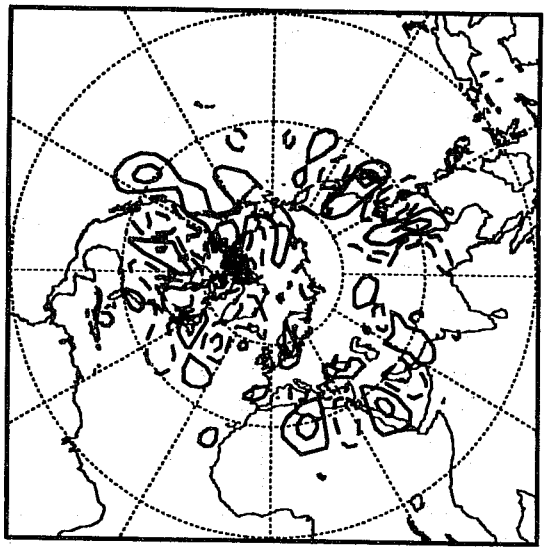

Fig 16 Initial perturbation added to the normal-amplitude ensemble member 8, in terms of 500 hPa geopotential height. Contour isolines 5 m, starting from 2.5 m.

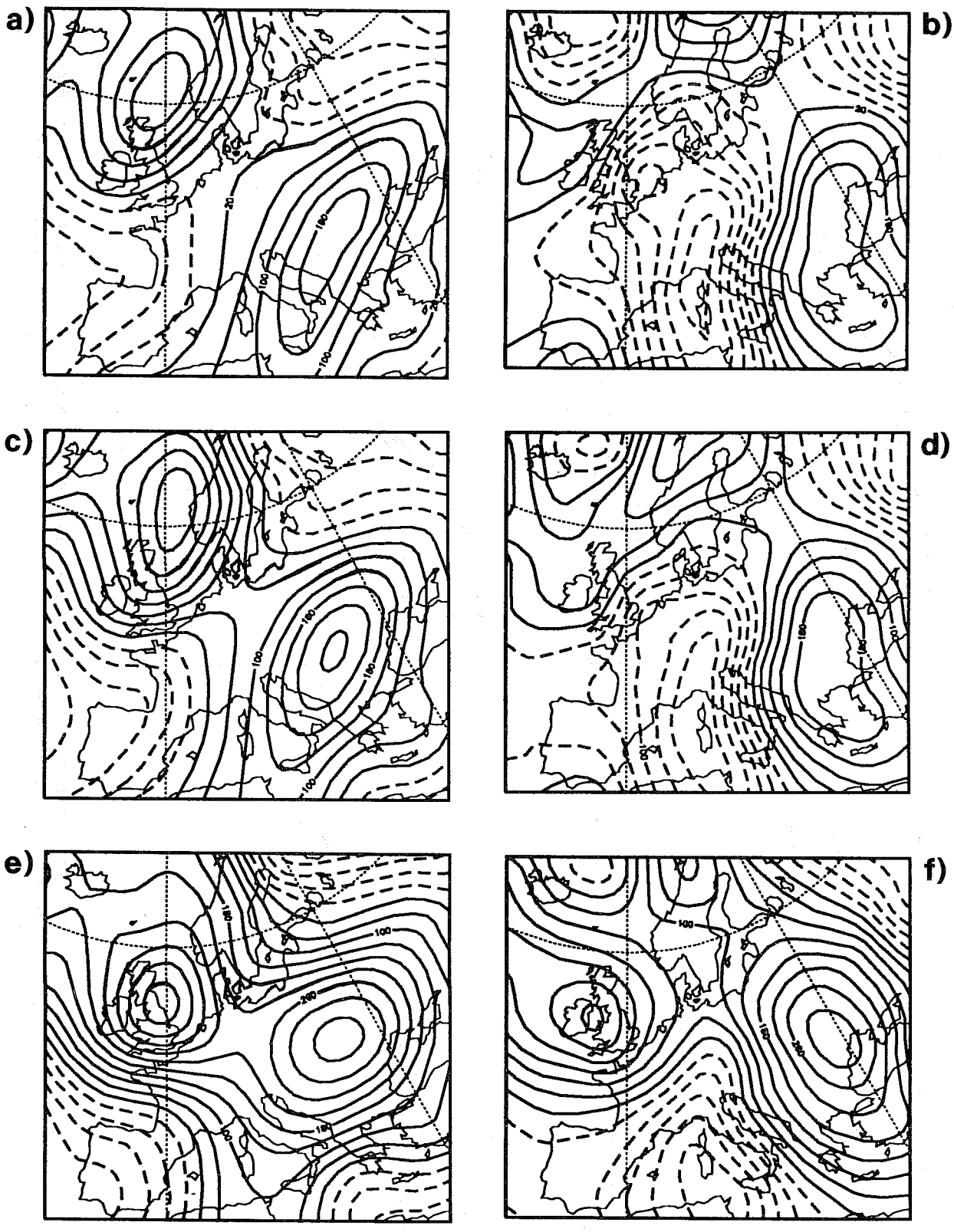

Fig 17 Difference between the control and the perturbed forecast (left panels) and difference between the analysis and the perturbed forecast (right panels), in terms of 500 hPa geopotential height, at forecast day 7, of members number 8 of the small-amplitude ensemble (panels a) and b)), of the reference ensemble normal-amplitude (panels c) and d)) and of the large-amplitude ensemble (panels e) and f)). Contour isolines 40 m, starting from 20 m.

APPENDIX A: SPREAD AND SKILL SCORE DEFINITION

The spread of a perturbed forecast is defined as its distance from the control forecast, and the skill of a forecast as its distance from the analysis. The spread $p(t)$ of each perturbed forecast ***ens*** is measured either using the anomaly correlation coefficient (ACC) between the perturbed forecast and the control ***con***

$$p_{ACC}(t) = \frac{<(ens(t)-cli(t));(con(t)-cli(t))>_g}{\|(ens(t)-cli(t))\|_g \; \|(con(t)-cli(t))\|_g}, \tag{A.1}$$

or computing the rms distance between the two forecasts

$$p_{RMS} = \|ens(t) - con(t)\|_g. \tag{A.2}$$

The grid-point scalar product $<..;..>_g$ and its associated norm $\|..\|_g$ are computed in the physical space (including $\cos(\lambda)$ weights, λ being latitude).

The skill of each forecast ***fc*** (with ***fc*** identifying either a perturbed forecast, or the control ***con***) is defined either as the ACC between the forecast and the analysis

$$k_{ACC}(t) = \frac{<(fc(t)-cli(t));(ana(t)-cli(t))>_g}{\|(fc(t)-cli(t))\|_g \; \|(ana(t)-cli(t))\|_g}, \tag{A.3}$$

or as the rms distance between a forecast and the analysis

$$k_{RMS} = \|fc(t) - ana(t)\|_g. \tag{A.4}$$

In all these eqs. ***cli*** identifies the climatology and ***ana*** the observed field (analysis).

REFERENCES

Buizza, R, J Tribbia, F Molteni and T Palmer, 1993. Computation of optimal unstable structures for a numerical weather prediction model. Tellus, 45A, 388-407.

Buizza, R and T N Palmer, 1994. The singular-vector structure of the atmospheric general circulation. J Atmos Sci, submitted.

Buizza, R, 1994a. Localization of optimal perturbations using a projection operator. In print by the Q J R Meteorol Soc.

Buizza, R, 1994b. Sensitivity of optimal unstable structures. Q J R Meteorol Soc, 120, 429-451.

Courtier, P, C Freydier, J-F Geleyn, F Rabier and M Rochas, 1991. The Arpege project at Météo France. In: Seminar on Numerical Methods in Atmospheric Models, Vol. 2, 9-13 September 1991, ECMWF, Shinfield Park, Reading, UK, 193-231.

Epstein, E S, 1969. Stochastic dynamic prediction. Tellus, 21, 739-759.

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

Farrell, B F, 1988. Optimal excitation of baroclinic waves. J Atmos Sci, 46 (9), 1193-1206.

Lacarra, J-F and O Talagrand, 1988. Short range evolution of small perturbations in a barotropic model. Tellus, 40A, 81-95.

Leith, C E, 1974. Theoretical skill of Monte Carlo forecast. Mon Wea Rev, 102, 409-418.

Molteni, F, R Buizza, T N Palmer and T Petroliagis, 1994. The ECMWF Ensemble Prediction System: methodology and validation. In: Fourth Workshop on Meteorological Operational Systems, 22-26 Nov 1993, ECMWF, Shinfield Park, Reading, UK, 33-79.

Mureau, R, F Molteni and T N Palmer, 1993. Ensemble prediction using dynamically-conditioned perturbations. Q J R Meteorol Soc, 119, 299-323.

Palmer, T N, F Molteni, R Mureau, R Buizza, P Chapelet and J Tribbia, 1993. Ensemble prediction. In: Validation of Models over Europe, Vol. 1, 7-11 September 1992, ECMWF, Shinfield Park, Reading, UK, 21-66.

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

Toth, Z and E Kalnay, 1993. Ensemble forecasting at NMC: the generation of perturbations. Bull Am Met Soc, 74 (12), 2317-2330.

Tracton, S and E Kalnay, 1993. Operational ensemble prediction at the National Meteorological Center: practical aspects. Weather and Forecasting, 8, 379-398.

Vukicevic, T, 1991. Non-linear and linear evolution of initial forecast errors. Mon Wea Rev, 119, 1602-1611.