Potential forecast skill of ensemble prediction, and spread and skill distributions of the ECMWF Ensemble Prediction System

Roberto Buizza
European Centre for Medium-Range Weather Forecasts
Reading, Berkshire, UK

ABSTRACT

Ensemble forecasting is a feasible method to integrate a deterministic forecast with an estimate of the probability distribution of atmospheric states. At the European Centre for Medium-Range Weather Forecasts (ECMWF) the Ensemble Prediction System (EPS) comprises 32 perturbed and one unperturbed non-linear integrations at T63 spectral triangular truncation and with 19 vertical levels. The perturbed initial conditions are generated using the most unstable directions growing over a 48-hour time period, computed at T42L19 resolution.

This work describes the performance of the ECMWF EPS during the first 18 months of daily operation, focusing on the 500 hPa geopotential height fields.

First, the EPS is described, and the validation approach followed throughout this work is discussed. In particular, spread and skill distribution functions are introduced to define a more integral validation methodology for ensemble prediction.

Then, the potential forecast skill of ensemble prediction is estimated considering one ensemble member as verification (perfect ensemble assumption). In particular the ratio between ensemble spread and control error is computed, and the potential correlation between ensemble spread and control forecast skill is evaluated. The results obtained within the perfect ensemble hypothesis give estimates of the limits of forecast skill to be expected by ensemble prediction.

Finally, the EPS is validated against analysis fields, and potential and real forecast skill are compared. Results indicate that the EPS spread is smaller than the distance between the control forecast and the analysis. Considering ensemble spread/control skill scatter diagrams, a predictability index is introduced to estimate the percentage of wrongly predicted cases with small spread/high control skill. Results suggest that there is a good correspondence between small ensemble spread and high control skill. Nevertheless, the percentage of analysis values lying outside the EPS forecast range is still high.

1. INTRODUCTION

Simmons et al (1995) analyzed in great detail the error growth of the 10-day forecast of the European Centre for Medium-Range Weather Forecasts's (ECMWF) operational model from 1 December 1980 to 31 May 1994, and concluded that the accuracy has been improved substantially over the first half of the forecast range, but that there has been little reduction of error in the late forecast range. While this applies on average, it is also true that there has been improvement in the skill of the good forecasts. In other words, good forecasts have higher skill now than in the past. The problem is that it is still difficult to assess a-priori whether a forecast will be skilful or unskilful, only using deterministic prediction. By contrast, ensemble prediction has the capability of estimating the forecast skill of a deterministic forecast. This is because ensemble prediction can integrate a deterministic forecast with an estimate of the probability distribution function of atmospheric states.

Since December 1992, both the US National Meteorological Centre (NMC) and ECMWF have integrated their deterministic high-resolution prediction with medium-range ensemble prediction (*Tracton and Kalnay*, 1993; *Palmer et al*, 1993). The development follows the theoretical and experimental work of *Epstein* (1969), *Gleeson* (1970), *Fleming* (1971a-b) and *Leith* (1974).

Both centres follow the same strategy of providing an ensemble of forecasts computed with the same model, one started with unperturbed initial conditions referred to as the 'control' forecast and an ensemble with initial conditions defined adding small perturbations to the control initial condition. Apart from differences in the ensemble size and the fact that at NMC a combination of lagged forecasts is used, the NMC and the ECMWF approaches to ensemble prediction differ substantially in the definition of the perturbations added to the control initial conditions to generate the initial conditions of the perturbed forecast. We refer the reader to *Toth and Kalnay* (1993) for the description of the 'breeding' method applied at NMC, and to *Buizza and Palmer* (1995) for a thorough discussion of the singular vector approach followed at ECMWF.

The first part of this paper briefly describes the ECMWF Ensemble Prediction System (EPS) and lists the major modifications introduced since its implementation on 19 December 1992 (the reader is referred to *Molteni et al*, 1996, for a more complete description of the EPS). The successful implementation of the ECMWF EPS follows early experiments by *Hollingsworth* (1980), who demonstrated that a sparse random sampling of phase space does not produce a realistic distribution of forecast states, and ensemble forecasting experiments in which unstable singular vectors computed from a 3-level quasi-geostrophic model were used to generate the initial perturbations (*Mureau et al*, 1993; *Molteni and Palmer*, 1993).

The second part of this paper has been inspired by the work of *Lorenz* (1982), who introduced the concept of potential improvement in forecast skill for a deterministic prediction, and estimated it by comparing deterministic forecasts from subsequent days (see also *Simmons et al*, 1995). Our aim is to estimate the potential forecast skill of ensemble prediction, by considering a perfect ensemble system where the verifying analysis is substituted by a randomly chosen perturbed ensemble member. Working within this hypothesis, we evaluate the characteristics of a perfect ensemble in terms of ensemble spread (i.e. the average distance of the ensemble members from the control forecast), skill of the control forecast and of the ensemble-mean, correlation between spread and skill, and percentage of analysis values lying outside the ensemble forecast range. In particular, we focus on three requirements: *i) the ensemble spread should be comparable to the error of the control forecast, ii) small spread should indicate high probability of a skilful control forecast*, and *iii) the verifying analysis should be included within the range covered by the ensemble forecasts*. The work is based on 18 months of daily ensemble prediction, from 1 May 1994 (when daily EPS started being run at ECMWF) to 31 October 1995.

Finally, in the third part of the paper, we consider the ECMWF EPS and, focusing on the three requirements mentioned above, we compare the potential with the real ensemble forecast skill. In particular, ensemble spread and control skill distributions are compared, and scatter diagrams of spread and control skill are analyzed. An index of predictability computed from contingency tables of ensemble spread and control skill is introduced to quantify the difference between the potential and the real skill of the EPS. The skill of the ensemble-mean is also discussed.

The paper is organized as follows. After this Introduction, Section 2 describes the ECMWF EPS and lists the major modifications introduced since its implementation. The methodology and the validation technique applied in this paper are discussed in Section 3. The potential forecast skill of ensemble prediction is studied in Section 4, while the comparison between potential and real forecast skill of the EPS is discussed in Section 5. Conclusions are drawn in Section 6.

2. THE ECMWF ENSEMBLE PREDICTION SYSTEM

The ECMWF EPS comprises, at the moment of writing, 32 perturbed and one unperturbed (control) non-linear integrations of a T63L19 Eulerian version of the ECMWF model (*Simmons et al*, 1989 and *Courtier et al*, 1991). The initial conditions of the 32 perturbed members are created by adding perturbations to the control initial conditions. The initial perturbations are defined using the singular vectors (*Buizza and Palmer*, 1995) of a linear approximation of the ECMWF model. A brief description of the EPS is reported hereafter, and a (randomly chosen) case study is discussed to illustrate the main steps of the EPS performed daily. The reader is referred to *Molteni et al* (1996) for a more detailed description of the EPS.

2.1 Singular vectors computation

Denote by $\boldsymbol{x}$ a perturbation, and by

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

the solution of the linear model equations

$$\frac{\partial \boldsymbol{x}}{\partial t} = \boldsymbol{A}_l \boldsymbol{x}, \tag{2}$$

where $\boldsymbol{L}$ is the forward tangent propagator, and $\boldsymbol{A}_l$ in Eq (2) is an approximation of the tangent linear version of the ECMWF model $\boldsymbol{A}$.

The total energy of the perturbation $\boldsymbol{x}$ at time t can be computed as

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

where $\boldsymbol{L}^*$ is the adjoint of the linear propagator $\boldsymbol{L}$ with respect to the total energy scalar product (..;..) [*Buizza*

et al, 1993, Eq (5.1)], and $\boldsymbol{T}$ is the self-adjoint local projection operator (*Buizza and Palmer*, 1995). 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 $\phi \geq 30°N$.

The singular vectors $\boldsymbol{v}_i$ of the propagator $\boldsymbol{L}$ are the perturbations with maximum energy growth over the optimisation time interval t. They are computed solving the eigenvalue problem

$$\boldsymbol{L}^{*}\boldsymbol{T}^{2}\boldsymbol{L}\boldsymbol{v}_i = \sigma_i^2 \boldsymbol{v}_i. \tag{4}$$

Specifically, they are the eigenvectors $\boldsymbol{v}_i$ with maximum eigenvalues σ_i^2 of the operator $\boldsymbol{L}^{*}\boldsymbol{T}^{2}\boldsymbol{L}$. The square root of an eigenvalue σ_i^2 is named a singular value.

At the moment of writing, the singular vectors are computed at T42L19 resolution with $t = 48\ h$, following a time evolving trajectory computed applying the complete ECMWF physical package, but using only a linear surface drag and vertical diffusion scheme (*Buizza*, 1994a) when computing linear forward and adjoint integrations (Table 1).

Figure 1 shows the singular values computed for 05.11.95 initial conditions (the singular vectors are ranked with respect to the singular values). For this case, 38 singular vectors have been computed numerically after performing 70 integrations of the forward/adjoint models, using a Lanczos algorithm (*Strang*, 1986). The singular vectors have very localized structures, and grow in the regions of maximum instability of the atmosphere. This was shown by *Buizza and Palmer* (1995), who also pointed out that there is a very strong relation between the singular vectors' localization and a simple measure of both barotropic and baroclinic energy growth given by the growth rate of the most unstable Eady mode (*Hoskins and Valdes*, 1990).

Figure 2 shows three singular vectors for the 95.11.05 case, at initial and optimisation time. The first singular vector is growing across the eastern border of the Asian continent, a region characterized by a very intense and rapid development (Fig 3). By contrast, the third and the sixth singular vectors are amplifying in relatively less unstable regions, as is reflected by their smaller singular values. The different flow characteristics of the Asian, Pacific and European regions influence not only the singular values, but also the vertical structure of the singular vectors, especially at optimisation time. In fact, while the total energy of the first singular vector has double maxima in the vertical, with a predominant low-level growth, the third and the sixth singular vectors have a more common vertical profile peaking at optimisation time around model level 9 (Fig 4).

Horizontal resolution:	T42 spectral triangular truncation
Vertical resolution:	19 vertical levels
Optimisation time interval:	48 hours
Diabatic schemes in direct:	full physics
Diabatic schemes in forward/adjoint:	linear physics and horizontal diffusion
Local Projection Operator:	$\lambda \geq 30°N$

Table 1 Characteristics of the singular vector computation used at the time of writing.

	Northern Hemisphere		Europe	
	day 2	day 5 to 7	day 2	day 5 to 7
01.05.94-31.07.94	1.43 (0.99)	1.56 (0.96)	1.30 (0.99)	1.50 (0.97
01.08.94-31.10.94	1.10 (1.03)	1.22 (0.97)	0.96 (1.00)	1.22 (1.00)
01.11.94-31.01.95	1.06 (1.03)	1.27 (0.98)	0.97 (1.03)	1.19 (0.99)
01.02.95-03.05.95	1.06 (1.01)	1.23 (1.02)	0.96 (0.99)	1.16 (1.03)
01.05.95-31.07.95	1.31 (1.02)	1.38 (1.00)	1.29 (0.98)	1.32(1.00)
01.08.95-31.10.95	1.28 (1.03)	1.28 (1.01)	1.11 (1.05)	1.16 (1.02)

Table 2 Ratio among the (seasonal average) rms error of the control and the (seasonal average) rms spread, for NH and Europe, at forecast day 2 and averaged from day 5 to 7. Values in parentheses refer to a perfect ensemble (see text).

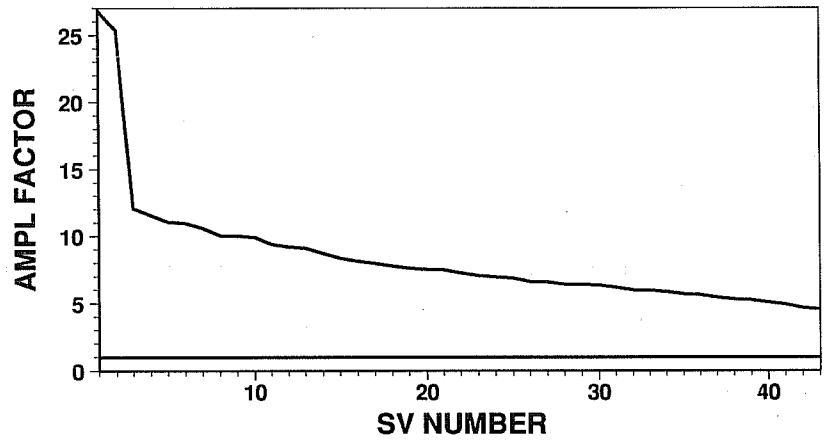

Fig 1 Singular values σ_i for 05.11.95 (the singular vectors are ranked with respect to their amplification rate).

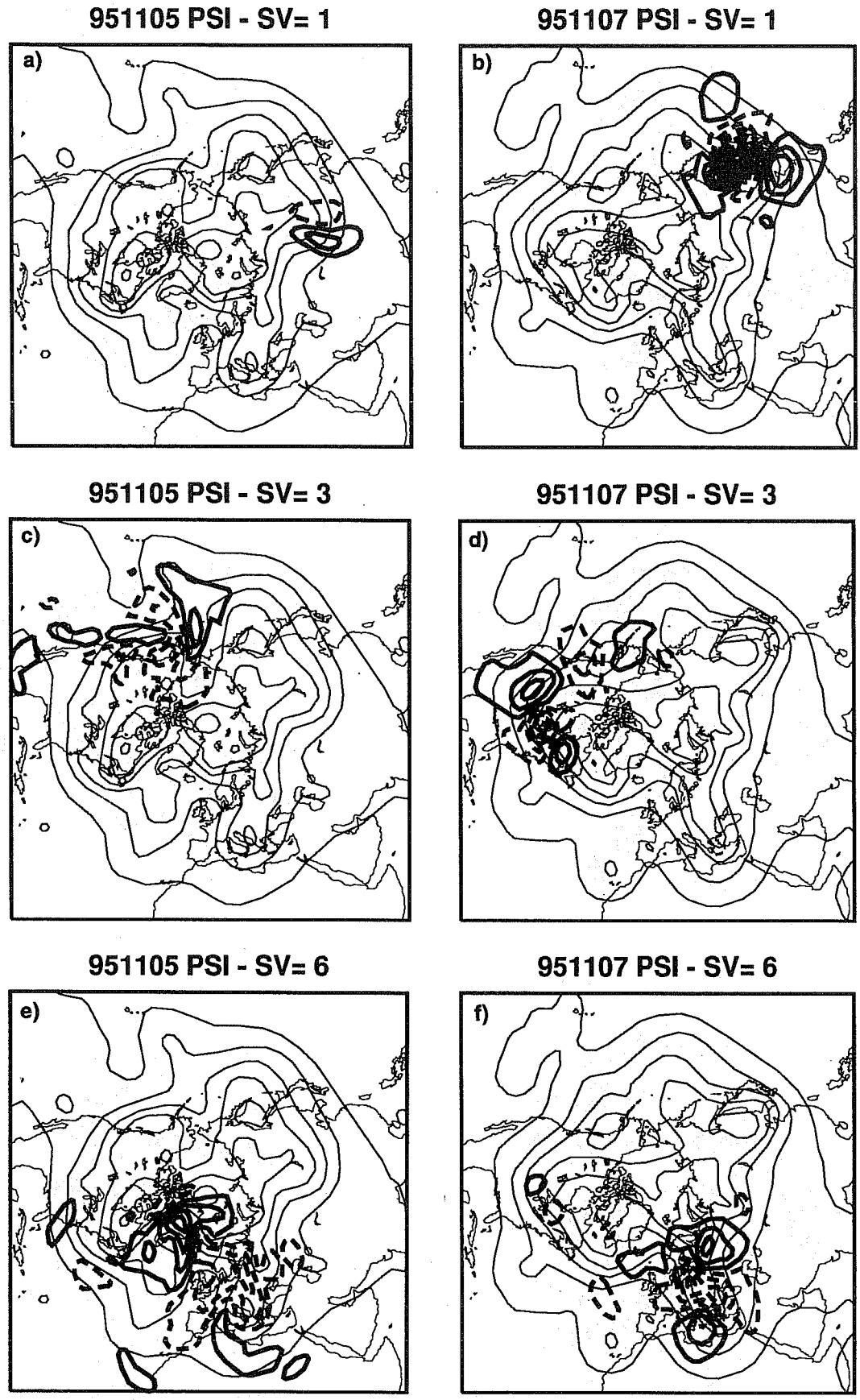

Fig 2 Singular vectors number 1 (top panels), 3 (middle panels) and 6 (bottom panels) at initial (left panels) and optimisation time (right panels). Each panel shows the singular vector streamfunction at model level 11 (approximately 500 hPa), superimposed to the trajectory 500 hPa geopotential height field. Streamfunction contour interval 0.5 x 10^{-8} m^2s^{-1} for left panels and 20 times larger for right panels; geopotential height contour interval 80 m.

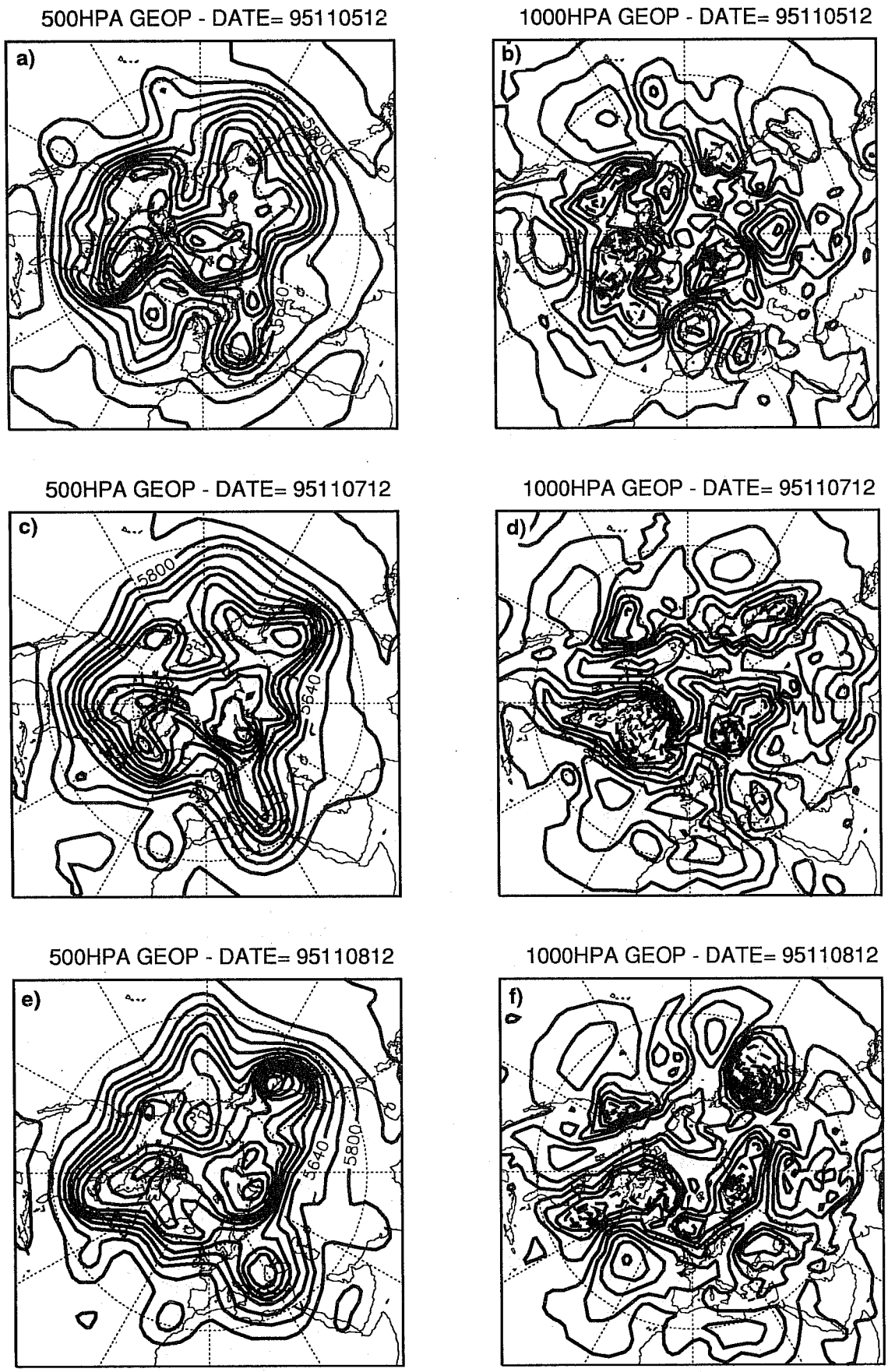

Fig 3 500 hPa (left panels) and 1000 hPa (right panels) geopotential height (left panels) trajectory for 5, 7 and 8 November 1995. Contour intervals 80 m for left panels and 40 m for right panels.

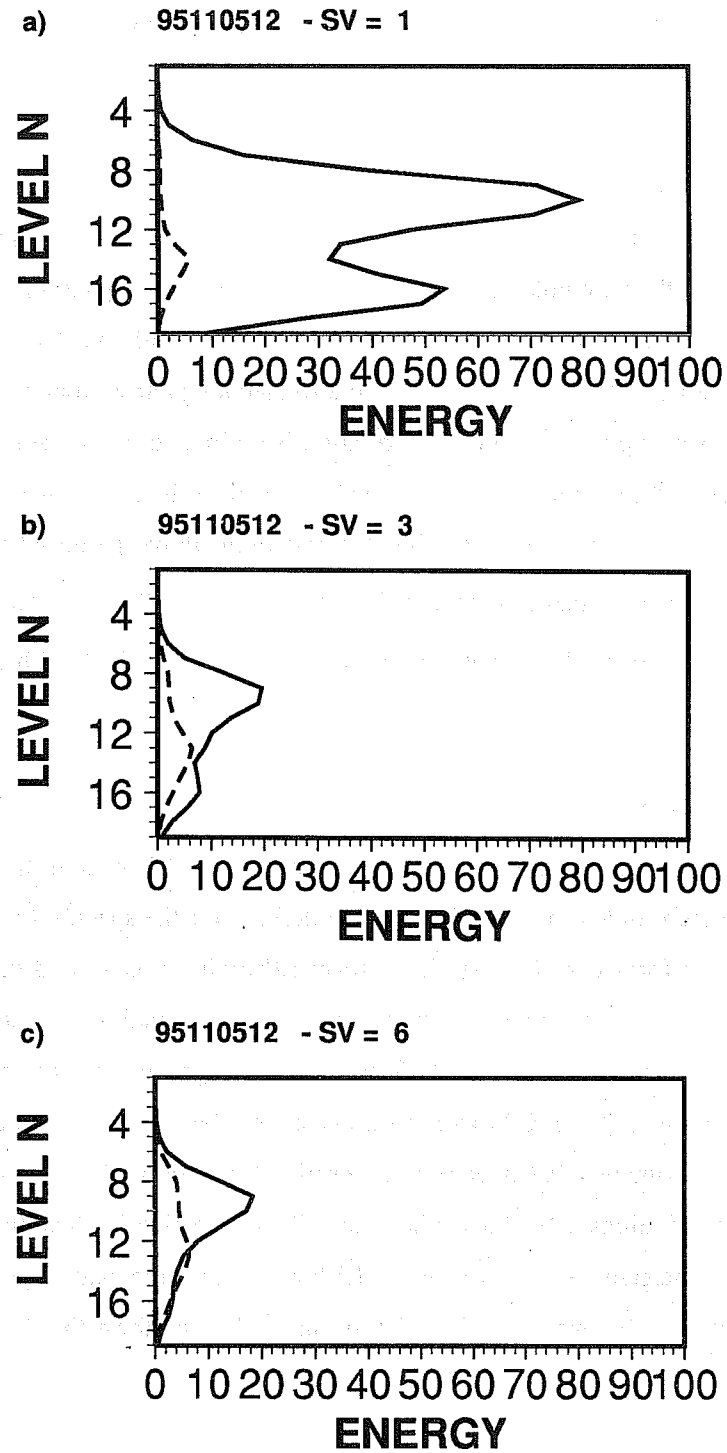

Fig 4 Total energy (m^2s^{-2}) vertical profile of (a) 1st, (b) 3rd and (c) 6th singular vector of 05.11.95, at initial (dash line, values multiplied by 100) and optimisation (solid line) time. (Note that singular vectors are normalized to have unit initial total energy norm.)

These three singular vectors are among the 16 selected from the 38 computed singular vectors. The selection criteria are such that the first 4 singular vectors are always chosen and each subsequent singular vector (from the 5th onwards) is selected only if half of its total energy lies outside the regions where the singular vectors already selected are localized.

2.2 Generation of 16 perturbations

Once the 16 singular vectors have been selected, an orthogonal rotation in phase-space and a final re-scaling are performed to construct the ensemble perturbations. The purpose of the phase-space rotation is to generate perturbations which have the same globally-averaged energy as the singular vectors, but smaller local maxima and a more uniform spatial distribution. Moreover, the rotated singular vectors are characterized by similar amplification rates (at least up to 48 hours). The rotation is defined to minimize the local ratio between the perturbation amplitude and the amplitude of the analysis error estimate given by the ECMWF Optimum Interpolation procedure. At the moment of writing, the re-scaling allows perturbations to have local maxima up to $\alpha = \sqrt{1.5}$ larger than the local maxima of the analysis error estimate. The effect of the phase-space rotation and re-scaling procedure can be seen by comparing the singular vectors of Fig 2 with three initial perturbations shown in Fig 5.

2.3 Non-linear integrations

The 16 initial perturbations are specified in terms of the spectral coefficients of the 3-dimensional vorticity, divergence and temperature fields (no perturbations are defined for the specific humidity since the singular vector computation is performed with a dry linear forward/adjoint model), and the 2-dimensional surface pressure field. They are added and subtracted to the control initial conditions to define 32 perturbed initial conditions. Then, 32+1 (control) 10-day T63L19 non-linear integrations are performed. With the current ECMWF computer facilities (CRAY C90 with 16 processors), the elapsed time needed to compute 35-40 singular vectors is approximately 0.8 hour (phase described in section 2.1), the elapsed time needed for generating the initial perturbations is 0.1 hour (phase described in section 2.2) and the elapsed time needed to perform the 33 non-linear integrations is approximately 1.8 hours. Thus, the total elapsed time is approximately 2.7 hours, which is about 1.3 times the elapsed time needed to perform the 10-day T213L31 ECMWF operational forecast.

A first way of verification of the EPS performance is an analysis of the spread and the skill characteristics over different areas (see Fig 6 for the 05.11.95 case over NH). The spread of a perturbed ensemble member is defined as the anomaly correlation or root-mean-square distance between the perturbed ensemble member and the control, and the skill of a forecast is defined as the anomaly correlation or the root-mean-square distance between the forecast and the analysis (see Section 3)

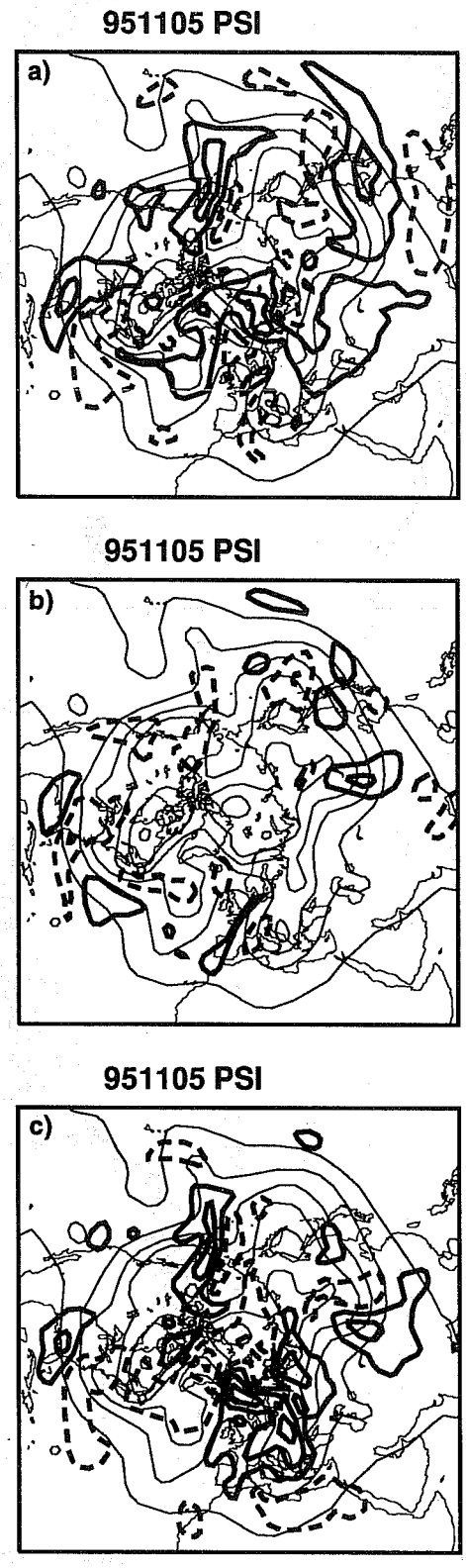

Fig 5 As Fig 2 but for three perturbations (streamfunction) generated applying a phase-space rotation to the selected singular vectors, superimposed to the trajectory 500 hPa geopotential height field. Contour interval 0.25 x 10^{-8} m^2s^{-1} for streamfunction and 80 m for geopotential height.

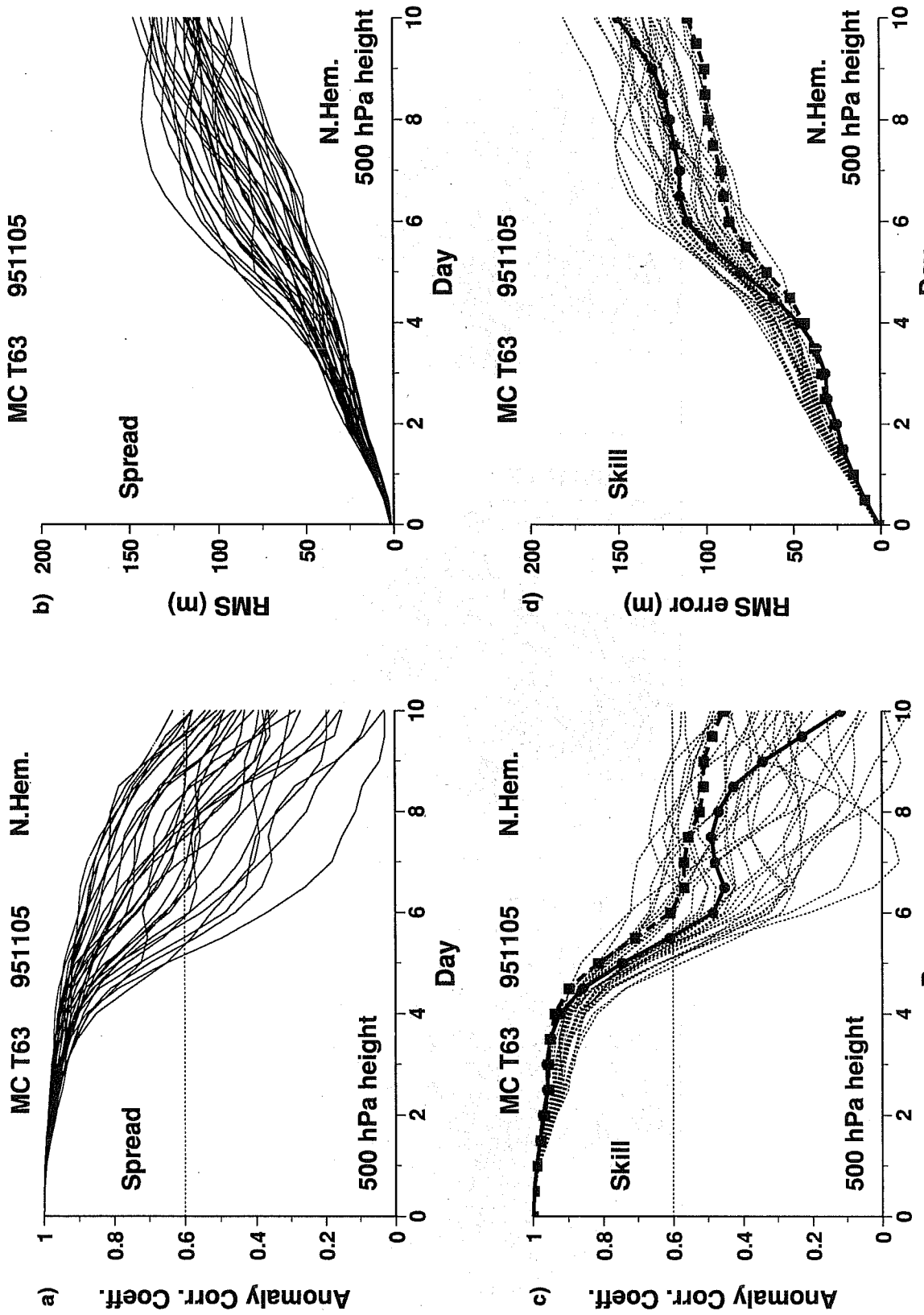

Fig 6 Spread (left panels) and skill (right panels) of the EPS run with starting date 05.11.95, computed for the 500 hPa geopotential height over NH. Top panels refer to acc and bottom panels to rms values.

2.4 Major modifications of the ECMWF EPS

Both model modifications and changes in the configuration used to generate the initial perturbations alter the EPS. The major model changes since EPS started (19.12.92) occurred on 04.08.93, when the new ECMWF surface and boundary layer scheme was introduced (*Viterbo and Beljaars*, 1995), and on 04.04.95, when the new ECMWF prognostic cloud scheme (*Tiedtke*, 1993; *Jacob*, 1994) and a new scheme for the representation of the sub-grid scale orography (*Lott and Miller*, 1995) were implemented. While it is beyond the scope of this paper to describe the impact of these changes on the model performance, we focus on the changes of the configuration used to generate the perturbed initial conditions.

19.12.92: EPS starts three days a week

After a month of trial, on 19.12.92 ECMWF started running the EPS every Saturday, Sunday and Monday. The singular vectors used to generate the perturbed initial conditions were computed globally at T21L19 resolution. After comparing singular vectors optimized over time intervals ranging from 12 to 72 hours, the optimisation time interval was set to 36 hours (*Buizza*, 1994a). The perturbation initial amplitude was set by $\alpha = \sqrt{2}$. The computer facilities allowed the ensemble size to be set to 32 perturbed and one unperturbed 10-day T63L19 non-linear integration.

20.03.93: local projection operator

Experimentation showed that during the NH warm seasons 70 iterations were not enough to compute at least 16 singular vectors growing in the NH. As a consequence, very similar if not identical perturbed forecasts would be generated. To avoid this problem, the so-called 'local projection operator' (*Buizza*, 1994b) was introduced in the singular vector computation, with $\phi \geq 30°N$. The impact of the local projection operator was proved to be almost negligible during the NH winter season. A further effect of its implementation was a small increase in the ensemble spread over the NH.

01.05.94: daily EPS

From this date onwards the EPS has been run daily.

23.08.94: larger initial amplitude and longer optimisation time

Analysis of the EPS performance suggested that the ensemble spread was too small. Thus, it was decided to increase the perturbation initial amplitude (i.e. $\alpha = \sqrt{4}$).

During winter 1994-95, so-called sensitivity experiments (*Rabier et al*, 1996) were performed at ECMWF. The basic idea was to identify the smallest initial perturbation that, added to the unperturbed initial conditions, was able to most reduce the day 2 forecast error. This smallest initial perturbation, named 'sensitivity field', was

computed using the adjoint model version. It was immediately realized that the comparison between the singular vectors and the sensitivity fields could give indications of the role of the singular vectors in explaining forecast error growth (*Buizza et al*, 1995). To ease the comparison, it was decided to increase the singular vector optimisation time interval to 48 hours.

It is worth noting that one of the hypotheses on which the ECMWF EPS is based is that perturbations with initial size of typical analysis errors evolve almost linearly up to the optimisation time. In other words, an upper boundary to the optimisation time interval used when computing the singular vectors is given by the time limit of validity of the linear approximation. This aspect was studied by *Buizza* (1995), who concluded that the time evolution of perturbations defined using T21L19 singular vectors optimized over 48 hours, and scaled with $\alpha \leq \sqrt{4}$, was linear to a good extent.

14.03.95: T42 singular vectors with smaller initial amplitude

The analysis of the total energy spectra of the singular vectors showed that the T21 truncation limit was too close to where the singular vector initial energy was peaking (*Hartmann et al*, 1995). Moreover, a systematic comparison of forecast errors, sensitivity fields and singular vectors confirmed that T42 was a more appropriate resolution to capture dynamically relevant directions in phase-space (*Buizza et al*, 1995). Since T42 singular vectors are characterized by larger amplification rates, the initial amplitude was reduced to keep the average root-mean-square spread comparable to the control root-mean-square error by setting $\alpha = \sqrt{1.5}$.

3. VALIDATION OF THE ECMWF EPS

Molteni et al (1996) listed the range of EPS products available at ECMWF, and presented different ways of validating its performance. This paper focuses on spread and skill relations considering the 500 hPa geopotential height field. Specifically, two types of verification are considered: spread and skill relations, and probability of analysis values lying outside the EPS forecast range. Six 92-day seasons are analyzed, starting from 01.05.94 until 31.10.95. The reader is referred to *Strauss and Lanzinger* (1995) for a description of other means of validation of the EPS (e.g. reliability diagrams).

3.1 Spread and skill

Consider two 500 hPa geopotential height fields $f(t) = f(\boldsymbol{x}_g, t)$ and $h(t) = h(\boldsymbol{x}_g, t)$, defined for each grid point value $\boldsymbol{x}_g$ inside a region Σ. Define the inner product

$$\langle f(t); h(t) \rangle_g \equiv \sum_{\boldsymbol{x}_g \in \Sigma} w(\boldsymbol{x}_g) f(\boldsymbol{x}_g, t) h(\boldsymbol{x}_g, t), \qquad (5)$$

where $w(\boldsymbol{x}_g) \equiv \cos(\phi_{\boldsymbol{x}_g})$, $\phi_{\boldsymbol{x}_g}$ being the latitude of $\boldsymbol{x}_g$. Denote by $\|..\|_g$ the norm associated to the inner product defined in Eq (5).

The distance of two fields f and h can be computed in terms of their anomaly correlation coefficient (acc) with respect to the climate cli

$$d_{acc}(f,h;t) \equiv \frac{<(f(t)-cli(t));(h(t)-cli(t))>_g}{\|(f(t)-cli(t)\|_g \|h(t)-cli(t)\|_g}, \tag{6}$$

or it can be computed simply as the root-mean-square (rms) distance

$$d_{rms}(f,h;t) \equiv \|f(t)-cli(t)\|_g. \tag{7}$$

Let us now consider an ensemble of N values f_j (e.g. the 32 EPS perturbed forecasts). Since the anomaly correlation is not a normally distributed variable, averages among anomaly correlation coefficients are computed after applying a so-called Fisher-z transform (see e.g. *Ledermann*, 1984). Specifically, given the set of anomaly correlation values $d_{acc}(f_j,h;t)$, with $j = 1,..,N$, first the transformed values z_j are defined

$$z_j \equiv 1.513 \log_{10} \frac{1+d_{acc}(f_j,h;t)}{1-d_{acc}(f_j,h;t)}, \tag{8a}$$

secondly the average among the z_j is computed

$$\bar{z} \equiv \frac{1}{N} \sum_{j=1}^{N} z_j, \tag{8b}$$

and finally the average acc $\bar{d}_{acc}(f_j,h;t)$ is computed applying the reverse Fisher-z transform

$$\bar{d}_{acc}(f_j,h;t) \equiv \frac{10^{\frac{\bar{z}}{1.513}}-1}{10^{\frac{\bar{z}}{1.513}}+1}. \tag{8c}$$

For similar reasons, averages among rms values $d_{rms}(f_j,h;t)$ are computed as

$$\bar{d}_{rms}(f_j,h;t) \equiv \sqrt{\frac{1}{N} \sum_{j=1}^{N} [d_{rms}(f_j,h;t)]^2}. \tag{9}$$

We define the **spread** of an ensemble of forecasts as the average distance of the perturbed ensemble members from the control, computed either in terms of anomaly correlation or root-mean-square distances

$$sp_{acc}(t) \equiv \bar{d}_{acc}(f_j, con; t), \tag{10a}$$

$$sp_{rms}(t) \equiv \bar{d}_{rms}(f_j, con; t), \tag{10b}$$

Here f_j is the j-th perturbed ensemble member, *con* is the control forecast and Eqs (8a-c) and (9) have been applied.

The ensemble-mean of an ensemble of N forecasts is defined as the average field

$$\bar{f} \equiv \frac{1}{N} \sum_{j=1}^{N} f_j. \tag{11}$$

The **skill** of a forecast f given by an ensemble member, the T213L31 or the ensemble-mean, is computed either in terms of acc or rms distance between the forecast and the analysis. Thus, it is defined applying Eqs (6) or (7), with $f(t)$ being the forecast and $h(t)$ being the analysis.

Let us now consider the skill of the control forecast and the spread of the EPS during a season. Probability distribution functions of the control skill and of the EPS spread can be constructed and compared. Specifically, considering the control skill, we denote by

$$P_{acc}^{sk(con)}(x), \ -1 \le x \le 1, \tag{12a}$$

the probability that the control acc skill has a value x. Analogously,

$$P_{rms}^{sk(con)}(x), \ 0 \le x, \tag{12b}$$

identifies the probability that the control rms error has a value x.

Considering an ensemble of N forecasts f_j, the probability distribution function of the ensemble acc skill is defined as

$$P_{acc}^{sk(ens)}(x) \equiv \frac{1}{N} \sum_{j=1}^{N} P_{acc}^{sk(f_j)}(x), \tag{13a}$$

and the probability distribution function of the ensemble rms error as

$$P_{rms}^{sk(ens)}(x) \equiv \frac{1}{N} \sum_{j=1}^{N} P_{rms}^{sk(f_j)}(x). \tag{13b}$$

Similarly, considering the spread, we denote by

$$P_{acc}^{sp}(x), \ -1 \le x \le 1, \tag{14a}$$

and

$$P^{sp}_{rms}(x),\ 0 \le x, \tag{14b}$$

the probabilities that the acc and rms spread have, respectively, values x.

Seasonal mean of skill and spread values can be computed either averaging the daily values applying Eqs (8a-c) and (9), or by computing the first-order momentum of the probability distribution functions defined in Eqs (12a-b), (13a-b) and (14a-b).

Spread and control skill relations are analyzed in this paper using scatter diagrams and computing correlation coefficients between the two time series. Moreover, contingency tables for high/low spread, high/low skill will be analyzed, with the categories defined by the average values (with this choice the tables will not be necessarily symmetric).

3.2 Probability of the analysis lying outside the EPS forecast range

Following an idea of *O Talagrand* (personal communication, 1994, see also *Strauss and Lanzinger*, 1995), let us consider the grid point values of the 500 hPa geopotential height predicted by the 32+1 EPS members $f_j(\boldsymbol{x},t)$, ranked so that $-\infty \le f_1 \le .. \le f_{33} \le \infty$. They define 34 intervals:

$$dh_1 \equiv (h(\boldsymbol{x},t)\ if\ -\infty \le h(\boldsymbol{x},t) < f_1)), \tag{15a}$$

$$dh_j \equiv (h(\boldsymbol{x},t)\ if\ f_{j-1} \le h(\boldsymbol{x},t) < f_j),\ j=2,..,32, \tag{15b}$$

$$dh_{33} \equiv (h(\boldsymbol{x},t)\ if\ f_{32} \le h(\boldsymbol{x},t) \le f_{33}), \tag{15c}$$

$$dh_{34} \equiv (h(\boldsymbol{x},t)\ if\ f_{33} < h(\boldsymbol{x},t) \le \infty). \tag{15d}$$

Note that the first 32 intervals are closed to the left and opened to the right, while interval number 33 is closed on both sides and interval number 34 is opened to the left and closed to the right. Following this interval definition, the probability of the analysis lying outside the EPS forecast range is defined as the sum of the probabilities of the analysis being in the two extreme categories, dh_1 and dh_{34}.

In a perfect model, because of the way each interval is defined, the probability of the analysis being inside the first and the last intervals should be nil, while the probability of being inside each of the other 32 categories should be identical, i.e. 1/32 in case of a 32+1 EPS.

4. POTENTIAL FORECAST SKILL OF ENSEMBLE PREDICTION

i) The EPS distribution of spread around the control should be similar to the control skill distribution (e.g. seasonal average rms spread should be similar to seasonal average rms error of the control).

ii) Small spread (around the control) should indicate high probability of a skilful control forecast.

iii) The verifying analysis should be included within the range covered by the ensemble forecasts.

All these requirements should be verified in the hypothesis of a perfect model, and, in the hypothesis that the analysis error lies in the sub-space of the phase-space of the system described, the 16 selected singular vectors. In fact, the spread and control skill distributions should be similar for the cloud of perturbed members to include the analysis. If the rms spread is smaller than the control rms error, for example, none of the perturbed forecast will be able to diverge enough from the control to get closer to the analysis. Small spread indicates that, whatever perturbation is added to the control initial conditions, very similar forecasts are originated. In such a situation, one should expect that the probability of analysis errors not damaging the skill of the control forecast should be high (very predictable case).

In this section we work in a perfect ensemble hypothesis, and we verify whether requirements *i)-iii)* are true for the six seasons analyzed in this paper. The 'perfect ensemble' hypothesis has been fulfilled by considering a randomly chosen ensemble member as the verifying analysis. This guarantees that the verifying analysis always lies in the sub-space of the phase space of the system described by the 16 selected singular vectors, that the spread is large enough and clearly that model errors are not taken into consideration.

Figures 7a-b show the 5-day running mean of the skill of the control forecast, the skill of the ensemble-mean and the average spread during a warm and a cold season (acc values are shown instead of rms values since they are less seasonally dependent). The ensemble spread and the skill of the control forecasts are comparable. This is confirmed by the ratio between the rms error of the control and the average rms spread at forecast day 2, and by the ratio averaged from day 5 to 7 (Table 2, values in parentheses refer to perfect ensemble), computed not only over NH but also over Europe.

The correspondence between spread and skill is further confirmed by the comparison of the spread distribution $P_{acc}^{sp}(x)$ and the control skill distribution $P_{acc}^{sk(con)}(x)$, for the two seasons (Fig 8). It is worth noting that comparable spread and control skill do not imply that the ensemble spread distribution $P_{acc}^{sp}(x)$ and the ensemble skill distribution $P_{acc}^{sk(ens)}(x)$ are comparable (see, e.g., Fig 9 for forecast day 7 over NH). In fact, by construction, all ensemble members are more distant from the analysis than the control up to forecast day 2, time limit up to which the initial perturbations dynamics can be linearly approximated (*Buizza*, 1995). After this time limit, non-linearity could possibly bring some ensemble members closer to the analysis than the control, but would probably bring most of them further away.

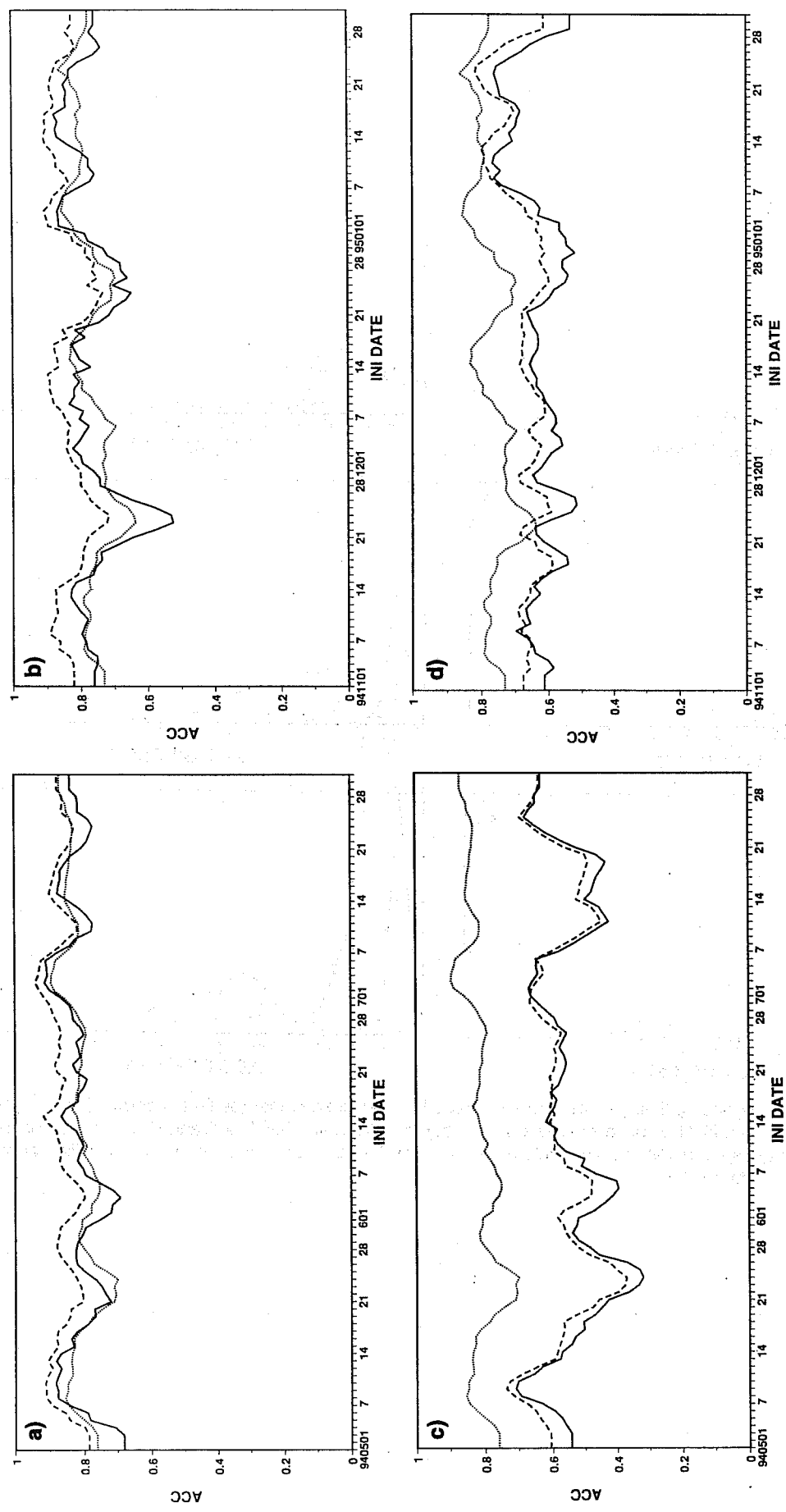

Fig 7 Control skill (solid), ensemble-mean skill (dash) and average spread (dot), at forecast day 7, for:
a) 01.05.94-31.07.94: perfect ensemble;
b) 01.11.94-31.01.95: perfect ensemble;
c) as a) but for the EPS verified against the analysis;
d) as b) but for the EPS verified against the analysis.
Values refer to 500 hPa geopotential height over NH.

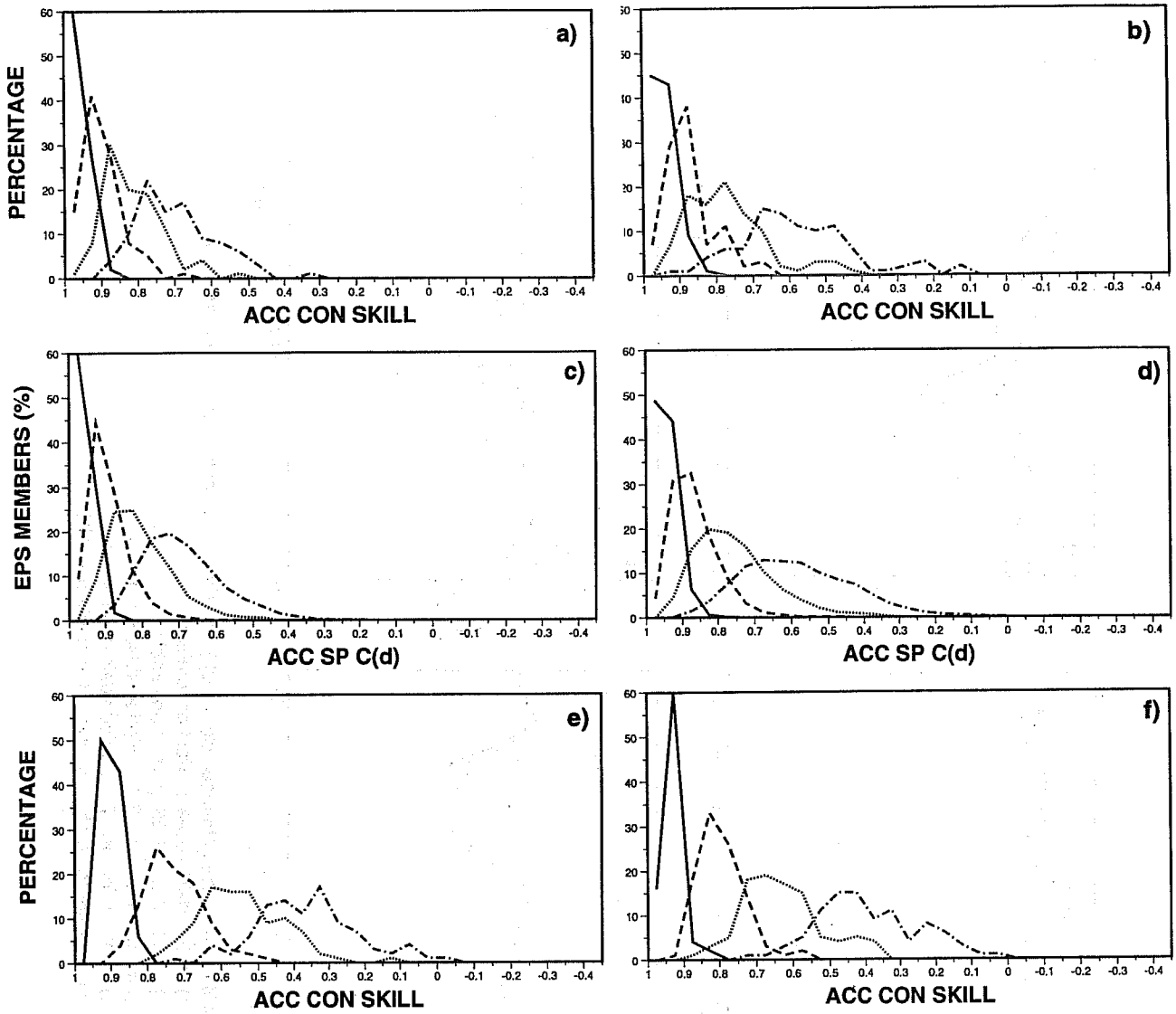

Fig 8 Distribution of the control skill verified against a perturbed ensemble member (perfect ensemble), for a) 01.05.94-31.07.94 and b) 01.11.95-31.01.95, at forecast day 3 (solid), day 5 (dash), day 7 (dot) and day 10 (chain dash). c-d) as a-b) but for the spread distributions. e-f): as a-b) but for the distribution of the control skill verified against the analysis. Values refer to NH.

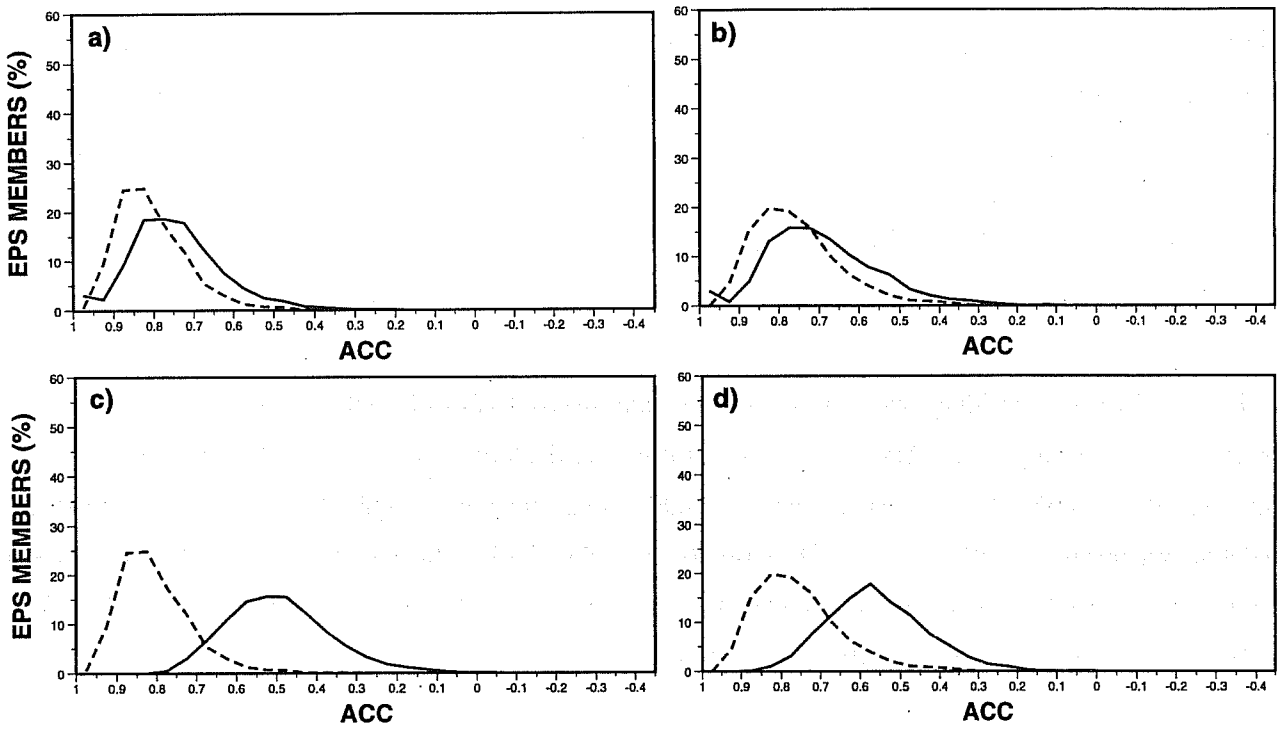

Fig 9 Perfect ensemble spread distribution (solid) and skill distribution (dash) for a) 01.05.94-31.07.94 and b) 01.11.94-31.1.95. c-d): EPS spread and skill distribution, with skill verified against the analysis. Values refer to NH at forecast day 7.

Figures 7a-b also show that there is temporal correlation between spread and skill. Tables 3-4 report correlation coefficients for NH and Europe at forecast day 7, for the six seasons analyzed in this paper (values in parentheses). Correlation coefficients range between 0.50 and 0.65 for NH, and between 0.31 and 0.86 for Europe. Note that the category small spread/low skill is not empty, and thus one should expect a certain percentage of small spread values not corresponding to skilful control forecast also with a perfect ensemble.

Considering the percentage of the analysis values lying outside the ensemble forecast range, nil percentages are found for all seasons, as expected.

It is worth discussing the impact of releasing the hypothesis that the analysis is included in the ensemble forecast range, by not including the ensemble member used for the verification in the perfect ensemble. One case study has been analyzed (Fig 10). The changes in the ensemble spread are almost negligible, and thus the spread/skill relation remains practically the same. The ensemble skill distribution changes slightly, and the percentage of analysis values lying outside the ensemble forecast range increases from nil to 4% (3%) and 8% (19%), respectively, at forecast day 7 and 10 over NH (Europe).

5. EPS VALIDATION FROM 01.05.94 TO 31.10.95

Let us first consider the two 92-day periods starting on 01.05.94 and 01.11.94 already discussed in detail in the previous section, and let us compare the performances of the EPS and the perfect ensemble. Note that in reality requirement *i)* formulated at the beginning of section 4 may be released or not, depending on whether small initial perturbations are asked to compensate for model errors. Until estimates of the relative size of model with respect to analysis errors are not known, we can argue that in the early forecast range, since model errors should not dominate the influence of analysis errors, requirement *i)* should be fulfilled, while we would allow the spread to be smaller than the distance between the control forecast and the analysis in the late forecast range.

Figures 7c-d are analogous to Figs 7a-b but using reality as verification. The control skill is lower than the average spread (high values of acc spread indicates highly correlated ensemble members, or in other words small differences among the ensemble forecasts). This is confirmed by the ratio between the rms error of the control and the average rms spread (Table 2), and by the comparison of the spread distribution (Fig 8c-d) and the control skill distribution verified against the analysis (Fig 8e-f).

Table 2 shows that the lack of spread is particularly noticeable during the NH warm seasons. The fact that the difference is enhanced during the NH warm seasons could be due to the EPS initial perturbations being computed with a dry linear forward and adjoint model, and to moist processes playing a more important role

Northern Hemisphere day 7			
01.05.94-31.07.94	cc=0.50 (0.65) Small spread Large spread i_{pred}= 0.23	Low skill 17 (13) 28 (33)	High skill 31 (35) 16 (11)
01.08.94-31.10.94	cc=0.40 (0.56) Small spread Large spread i_{pred}= 0.08	Low skill 12 (11) 27 (40)	High skill 28 (29) 25 (12)
01.11.94-31.01.95	cc=0.23 (0.64) Small spread Large spread i_{pred}= 0.31	Low skill 22 (15) 24 (35)	High skill 25 (32) 21 (10)
01.02.95-03.05.95	cc=0.27 (0.50) Small spread Large spread i_{pred}= 0	Low skill 17 (17) 29 (31)	High skill 30 (30) 16 (14)
01.05.95-31.07.95	cc=0.26 (0.55) Small spread Large spread i_{pred}= 0.22	Low skill 18 (12) 29 (37)	High skill 25 (31) 20 (12)
01.08.95-31.10.95	cc=0.39 (0.55) Small spread Large spread i_{pred}= 0.29	Low skill 17 (12) 31 (34)	High skill 30 (35) 14 (11)

Table 3 Seasonal contingency tables for small/large spread, low/high skill (computed in terms of acc), for NH at forecast day 7. The categories are defined by the average values. For each contingency table, the correlation coefficients *cc* between spread and skill and the predictability index i_{pred} [see Eq (16)] are reported. For each contingency table, the values in parentheses refer to a perfect ensemble (see text).

Northern Hemisphere day 7			
01.05.94-31.07.94	cc=0.47 (0.86) Small spread Large spread i_{pred}= 0.46	Low skill 15 (8) 39 (41)	High skill 25 (32) 13 (11)
01.08.94-31.10.94	cc=0.18 (0.31) Small spread Large spread i_{pred}= 0.06	Low skill 15 (14) 25 (30)	High skill 29 (30) 25 (18)
01.11.94-31.01.95	cc=0.45 (0.50) Small spread Large spread i_{pred}= 0.18	Low skill 16 (13) 33 (33)	High skill 31 (34) 12 (12)
01.02.95-03.05.95	cc=0.43 (0.43) Small spread Large spread i_{pred}= 0.26	Low skill 15 (11) 32 (35)	High skill 26 (30) 19 (16)
01.05.95-31.07.95	cc=0.46 (0.69) Small spread Large spread i_{pred}= 0.42	Low skill 19 (11) 28 (35)	High skill 26 (34) 19 (12)
01.08.95-31.10.95	cc=0.31 (0.49) Small spread Large spread i_{pred}= 0.30	Low skill 20 (14) 30 (34)	High skill 25 (31) 17 (13)

Table 4 As Table 3 but for Europe.

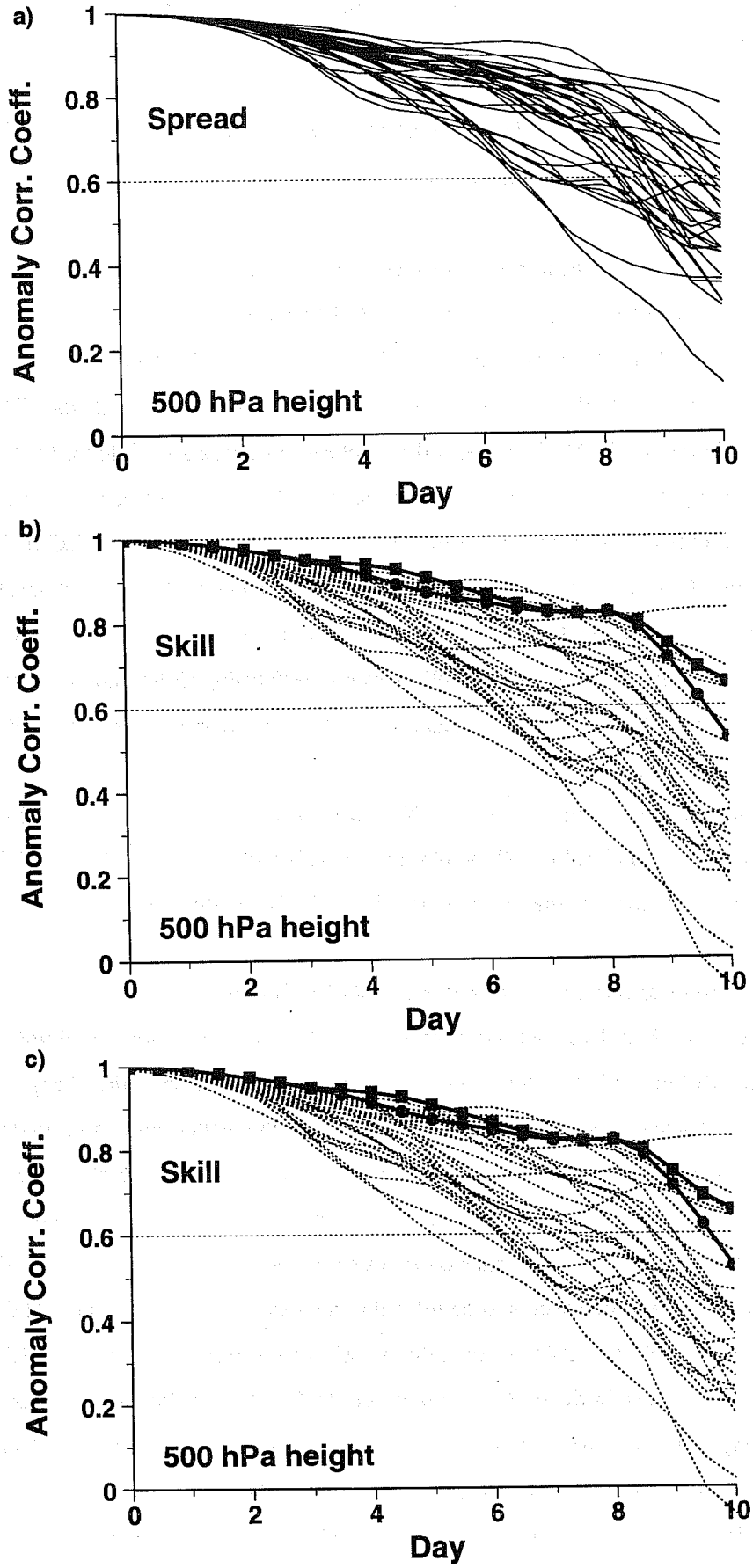

Fig 10 a) Spread of each member of a perfect ensemble. b) Skill of each member of a perfect ensemble including the ensemble member used as the analysis for verification (dot), skill of the control (solid with dots) and of the ensemble-mean (solid with squares). c) as b) but for a perfect ensemble without the ensemble member used as the analysis for verification. Values refer to the NH 500 hPa geopotential height for 02.12.94.

during the NH warm than cold seasons [*Errico and Ehrendorfer* (1995) showed that the inclusion of moist processes in the singular vector computation changes singular vector growth rates and structures.]

Table 2 also shows that the ratio between rms control error and rms spread is larger between day 5 and 7. This can be explained by the energetic characteristics of the T63L19 version of the ECMWF model, compared to the atmosphere or to the T213L31 model version itself. As *Simmons et al* (1995) pointed out, the level of transient activity in the forecast model should be similar to the transient activity in the analysis. *Tibaldi et al* (1990), comparing earlier versions of the ECMWF model at different resolutions, showed that a T63L19 resolution was not able to ensure the right level of change during the whole 10-day forecast period. Recent investigations confirmed that this problem is still present in the current T63L19 version of the ECMWF model (*Anders Persson*, personal communication 1995). The lack of model activity could be cured, at least in part, by using a higher resolution model version when performing the non-linear integrations. Note that, in fact, while the T106L19 model version used by *Tibaldi et al* (1990) was not performing significantly better than the T63L19 version, the current T106L19 model version is characterized by a more realistic model activity.

A complete picture of the EPS performance over the NH for the six seasons is given in Fig 11. Concerning the ensemble-mean skill, Fig 11 and Table 5 show that the ensemble-mean field is more skilful than the control forecast after forecast day 5, but that the improvement in reality is smaller than its potential value.

i) Ensemble spread distribution versus control skill distribution

Both Figs 11a-b and Table 2 confirm that the increase in the initial perturbation amplitude that was made on 23.08.94 reduced the difference between spread and control error, but despite the changes implemented on 14.03.95 (perturbations generated using T42 singular vectors but with smaller initial amplitude, see section 2.4) the spread is still too small. When the 14.03.95 modifications were introduced, the perturbation initial amplitude was set to have slightly more spread than with the previous system. Since T42 singular vectors are more unstable than T21 singular vectors, this was achieved with a net reduction of perturbation initial amplitude. The lack of spread measured afterwards seems to be related to the change implemented on 04.04.95 (new model version, see introduction of Section 2.4). In fact, the model version introduced on 04.04.95 seems to be less active and thus less able to sustain the perturbation growth (*A Simmons*, personal communication, 1995; this is confirmed by comparison of seasonal integrations of new and old model versions, Č Brancović, personal communication, 1995, and by results obtained by R Gelaro, personal communication, 1995, who found a reduction in the model sensitivity fields between the new and the old model versions).

A more thorough description of the EPS performance is given by the seasonally averaged ensemble skill distributions $P_{acc}^{sk(ens)}(x)$ (Fig 12). Probability distribution functions can be used to evaluate the percentage of

	Northern Hemisphere		Europe	
	day 7	day 10	day 7	day 10
01.05.94-31.07.94	0.03 (0.05)	0.05 (0.10)	0.02 (0.04)	0.05 (0.10
01.08.94-31.10.94	0.05 (0.07)	0.10 (0.17)	0.04 (0.08)	0.11 (0.21)
01.11.94-31.01.95	0.04 (0.06)	0.10 (0.14)	0.05 (0.05)	0.14 (0.14)
01.02.95-03.05.95	0.02 (0.06)	0.10 (0.14)	0.05 (0.06)	0.11 (0.14)
01.05.95-31.07.95	0.02 (0.07)	0.06 (0.16)	0.03 (0.05)	0.04 (0.15)
01.08.95-31.10.95	0.03 (0.08)	0.08 (0.18)	0.01 (0.09)	0.13 (0.19)

Table 5 Difference between the (seasonal average) skill of the ensemble-mean and the (seasonal average) skill of the control forecast, in terms of acc. Values in parentheses refer to a perfect ensemble.

	Northern Hemisphere		Europe	
	day 5	day 7	day 5	day 7
01.05.94-31.07.94	30	26	36	28
01.08.94-31.10.94	20	16	23	16
01.11.94-31.01.95	17	15	13	11
01.02.95-03.05.95	19	16	10	16
01.05.95-31.07.95	32	28	48	34
01.08.95-31.10.95	27	23	38	25

Table 6 Seasonal average of the percentage of analysis values lying outside the EPS forecast range.

	NH forecast day 7		Europe forecast day 7	
	cc	i_{pred}	cc	i_{pred}
01.05.94-31.07.94	0.50 (0.65)	0.11	0.55 (0.50)	0.33
01.08.94-31.10.94	0.27 (0.53)	0.30	0.04 (0.19)	0.46
01.11.94-31.01.95	0.24 (0.55)	0.45	0.35 (0.47)	0.10
01.02.95-03.05.95	0.32 (0.36)	0.27	0.51 (0.39)	0.16
01.05.95-31.07.95	0.42 (0.63)	0.31	0.44 (0.77)	0.53
01.08.95-31.10.95	0.42 (0.56)	0.06	0.31 (0.46)	0.45

Table 7 Correlation coefficients and predictability indices computed from scatter diagrams of the skill of the ensemble-mean versus the ensemble spread around the ensemble-mean, at forecast day 7 over NH and Europe. Values in parentheses refer to a perfect ensemble.

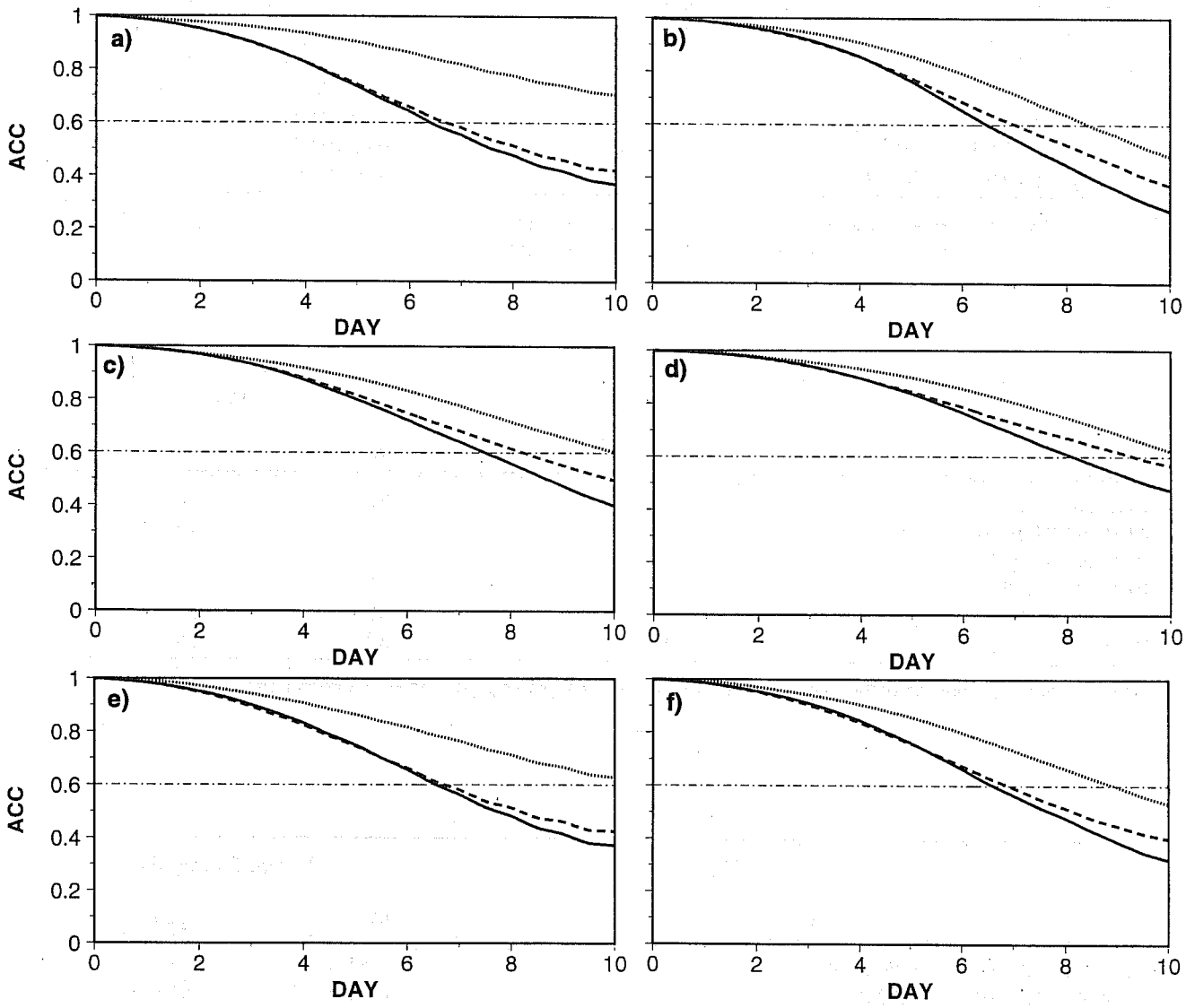

Fig 11 Control skill (solid), ensemble-mean skill (dash) and average spread (dot) for the 92-day seasons starting a) 01.05.94, b) 01.08.94, c) 01.11.94, d) 01.02.95, e) 01.05.95 and f) 01.08.95. Values refer to acc computed for the 500 hPa geopotential height over NH.

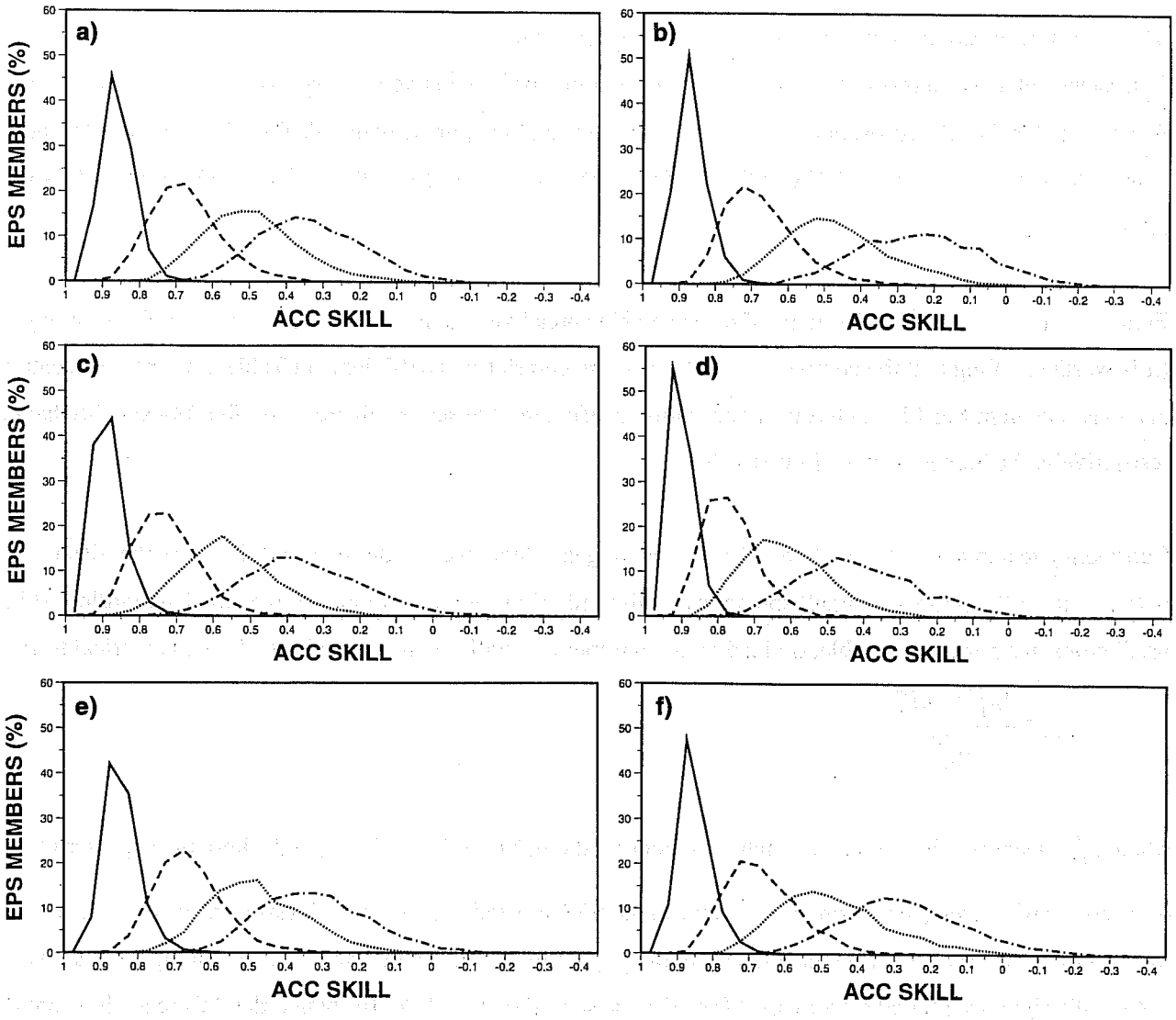

Fig 12 Ensemble skill distributions $P_{acc}^{sk(ens)}(x)$, at forecast day 3 (solid), day 5 (dash), day 7 (dot) and day 10 (chain dash), for the 92-day seasons starting a) 01.05.94, b) 01.08.94, c) 01.11.94, d) 01.02.95, e) 01.05.95 and f) 01.08.95 Values refer to acc computed for the 500 hPa geopotential height over NH.

EPS members with a skill score higher than a predefined threshold, e.g the percentage of EPS members having acc skill higher than 0.8 or 0.6 (Fig 13).

ii) Correspondence between small spread and high skill

Consistent with the results discussed above, the correlation coefficients between spread and control skill in the EPS are smaller than the correlation values in the perfect ensemble approximation (Tables 3-4), with differences ranging from zero (for the period 01.02.95-03.05.95 over Europe) to 0.41 (for the period 01.11.94-31.01.95 over NH).

Figure 14 shows the scatter diagrams of the ensemble spread versus the control skill for NH at forecast day 7 (it is worth recalling that the contingency tables and the correlation coefficients in Table 3 refer to the scatter diagrams shown in Fig 14). In terms of correlation coefficient, the scatter diagrams in Figs 14a and 14c have, respectively, the highest and the lowest value.

Considering requirement *ii)* of Section 4, for each contingency table we can define a skill index as the difference between the EPS number of small spread/low skill predictions and the number of the small spread/low skill predictions in a perfect ensemble, divided by the number of small spread predictions. In mathematical terms

$$i_{pred} \equiv \frac{|n_{1,1}^{EPS} - n_{1,1}^{per}|}{n_{1,1}^{EPS}}, \qquad (16)$$

where n_{ij}^{EPS} identifies the contingency table element on the *i-th* row, *j-th* column, and where the superscript ^{EPS} or ^{per} refers, respectively, to the contingency table of the EPS verified against the analysis or against an ensemble member (perfect ensemble). [The differences in the populations of the actual and perfect contingency tables is statistically significant for all seasons apart from the one started on 01.11.94, for which the difference has (only) a 65% probability of being significant according to a chi-square test.]

Since the index i_{pred} quantifies the capability of the EPS to identify the predictable cases, we will call i_{pred} the 'predictability index' (in simple words, i_{pred} gives the percentage of cases when predicted small spread did not correspond to high skill of the control forecast). Predictability indices have been computed for all seasons (Tables 3-4): values close to zero indicate that the EPS values of small spread/low skill is very close to the value given by a perfect ensemble. Predictability indices range from zero (for the period 01.02.95-03.05.95 over NH) to 0.46 (for the period 01.05.94-31.07.94 over Europe). Note that periods with very high correlation between spread and skill do not usually correspond to periods with low predictability indices. For example, the scatter diagram for the period 01.02.95-03.05.95 (Fig 13d) is characterized by a very skilful predictability index, but

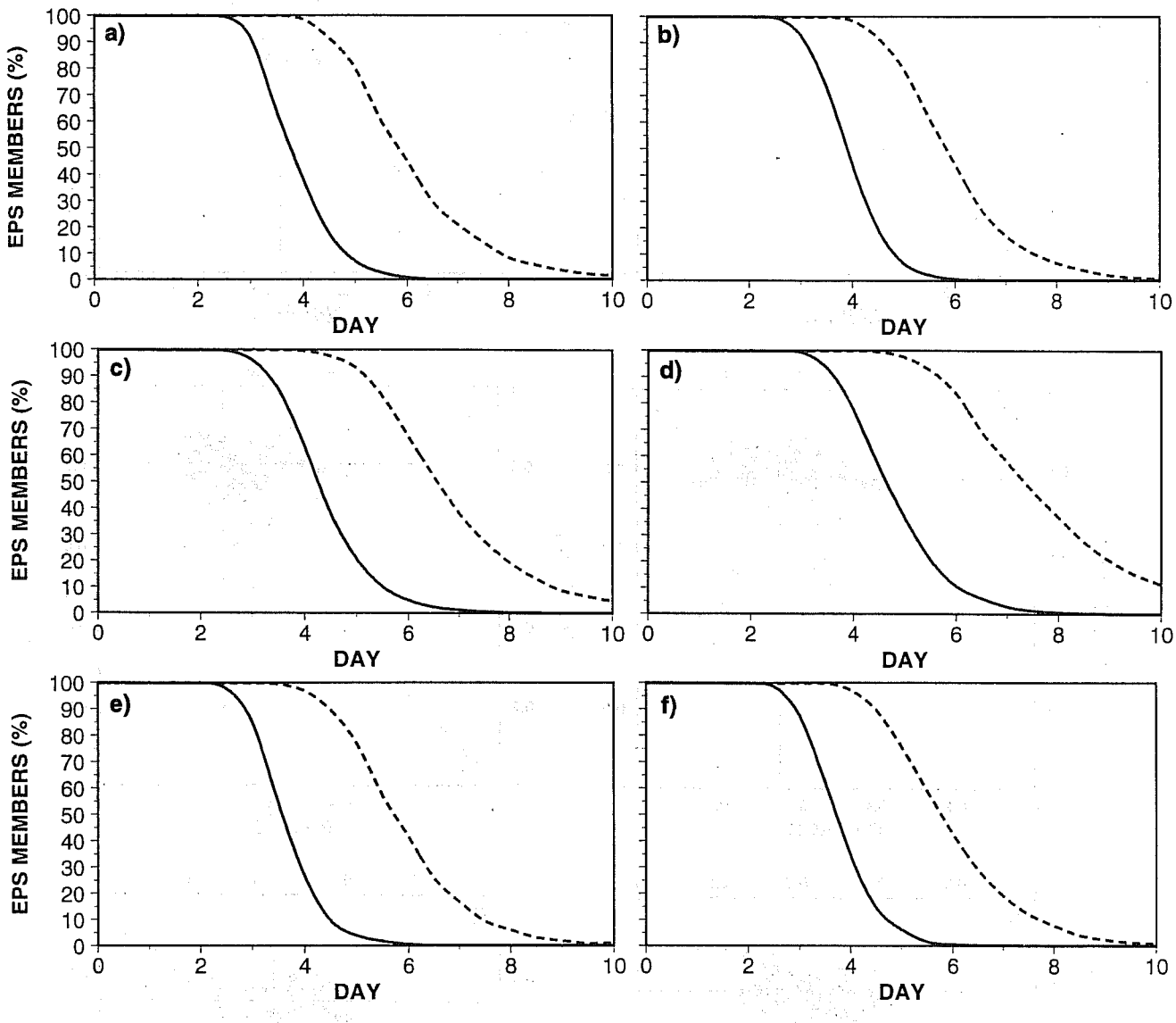

Fig 13 Percentage of EPS members with acc skill higher than 0.8 (solid) and 0.6 (dash), for the 92-day seasons starting a) 01.05.94, b) 01.08.94, c) 01.11.94, d) 01.02.95, e) 01.05.95 and f) 01.08.95 Values refer to acc computed for the 500 hPa geopotential height over NH.

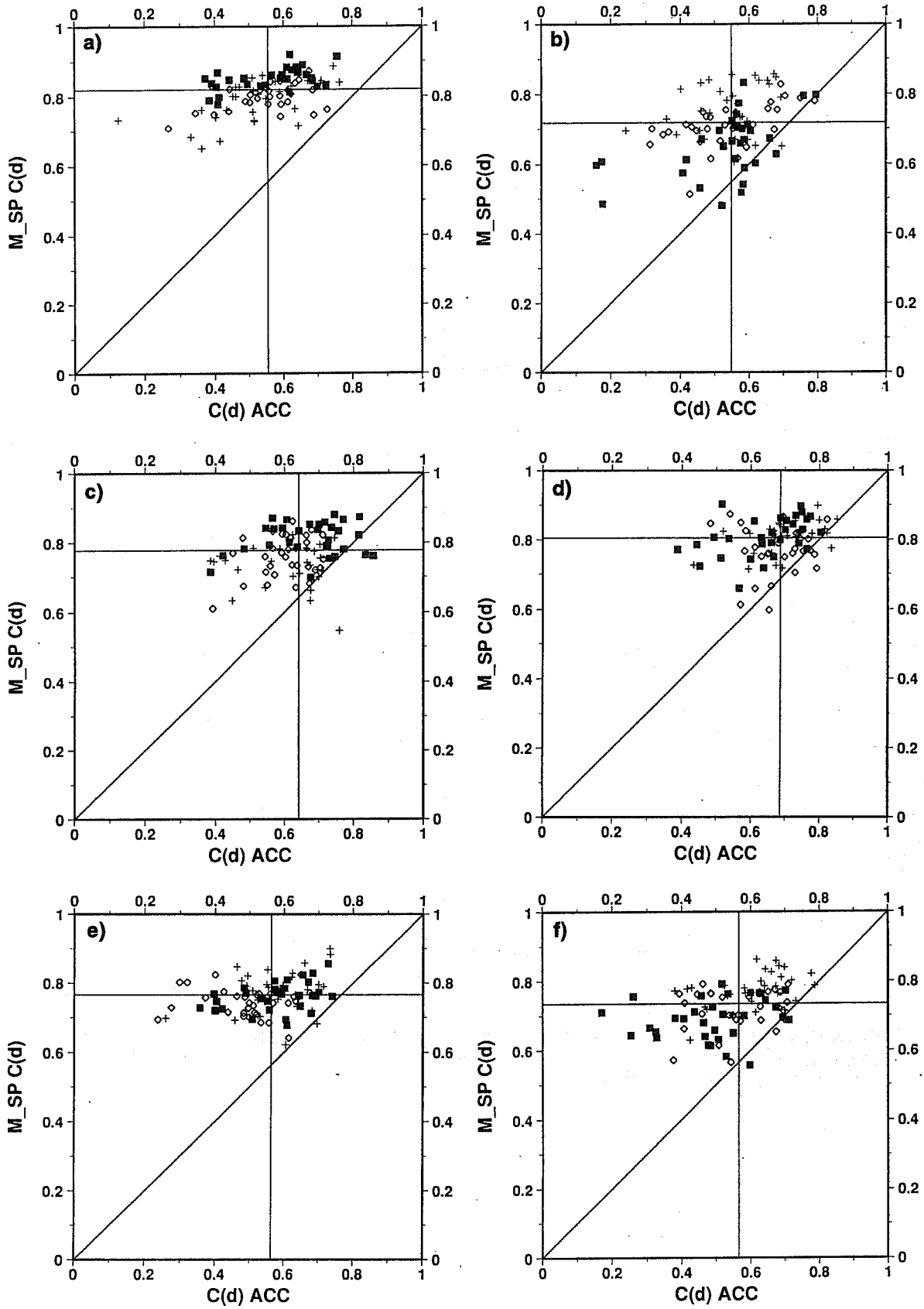

Fig 14 Scatter diagrams of EPS spread versus control skill for the 92-day seasons starting a) 01.05.94, b) 01.08.94, c) 01.11.94, d) 01.02.95, e) 01.05.95 and f) 01.08.95. Values refer to the 500 hPa geopotential height over NH. For each season, plus-diamond-square markers in the scatter diagram identify cases of the first-second-third month of the analyzed season.

by one of the lower correlation coefficients. This indicates that the two types of verification capture different aspects of the EPS, and should be used jointly to have a more complete EPS validation.

iii) Percentage of analysis values lying outside the EPS forecast range

Table 6 lists the percentage of analysis values lying outside the EPS forecast range, for NH and Europe at forecast days 5 and 7. The seasonal variability of the percentages reflects the ratio between control error and ensemble spread, with seasons with larger percentages corresponding to seasons with higher ratios (see Tables 2 and 6).

iv) Skill score of the ensemble-mean, and its correlation with ensemble spread around the ensemble-mean

Figure 11 shows that the ensemble-mean is more skilful than the control during NH cold seasons, but that the difference in skill is rather small during NH warm seasons. This is also confirmed by Table 5, which also shows that the difference between the potential and the actual difference is not very large.

It is worth investigating whether ensemble prediction can be used to predict the forecast skill of the ensemble-mean, by comparing the ensemble-mean skill with the spread of the ensemble defined with respect to the ensemble-mean instead of to the control forecast. Correlation coefficients between the skill of the ensemble-mean and the spread around the ensemble-mean, and predictability indices of the corresponding contingency tables have been computed. Table 7 reports the values computed using acc spread and skill for NH and Europe at forecast day 7. Compared with the results obtained using the control forecast (Tables 3-4), the correlation coefficients are similar, while the predictability indices relative to the ensemble-mean skill are higher, indicating an overall reduction of correspondence between small spread/high skill.

6. CONCLUSIONS

Ensemble prediction through multiple integrations of a deterministic model estimates the probability distribution of atmospheric states. At ECMWF, the Ensemble Prediction System (EPS) comprises one 10-day low resolution (T63L19) integration starting from the analysis, and 32 10-day T63L19 integrations starting from perturbed initial conditions. The perturbed initial conditions are generated using the singular vectors of a T42L19 linear approximation of the ECMWF primitive equation model. In the first part of the paper the EPS has been described, and the validation methodology applied throughout the paper has been presented. In particular, ensemble spread and control skill distribution functions have been defined.

In the second part of the paper the potential forecast skill of ensemble prediction has been discussed. A perfect ensemble system has been considered, by substituting the verifying analysis with an ensemble member, and verified for six 92-day seasons from 1 May 1994, the starting date of daily operational EPS. It has been shown

that a perfect ensemble fulfils three requirements: *i) the ensemble spread is comparable to the skill of the control forecast, ii) small spread indicates a high probability of a skilful control forecast* and *iii) the verification is included within the ensemble forecasts range.*

In the third part of this paper we compared the potential and the real forecast skill of the ECMWF EPS. Results show that the ensemble spread is still too small, especially in the second half of the 10-day forecast period. It is worth mentioning that preliminary results obtained by using a T106L19 model version indicate that the initial perturbation growth is more sustained in the T106L19 than in the T63L19 model version. Moreover, recent comparison of T42L19 and T63L19 singular vectors shows that the T63L19 singular vector growth is approximately 20% larger than the T42L19 growth. These results suggest that an increase of resolution both in the singular vector computation and in the non-linear model integration should reduce the difference between ensemble spread and control skill, without any further increase in the perturbation initial amplitude.

Considering the correlation between small spread around the control forecast and the control forecast skill, a 'predictability index' i_{pred} has been defined and computed using ensemble spread/control skill contingency tables. Both temporal correlations and predictability indices quantify limits of forecast skill that should be expected from ensemble prediction. Correlation coefficients of the EPS verified using the analysis have been proved to be smaller than their potential counterparts. Nevertheless, predictability indices indicate that there is a good correspondence between small ensemble spread and high control skill.

Consistent with the too small ensemble spread, the percentage of analysis values lying outside the EPS forecast range is still not negligible. This could indicate that the sub-space of the phase space of the system spanned by the ensemble perturbations does not include the analysis. It is again worth mentioning that preliminary results of the comparison of ensemble experiments with 128 instead of 32 perturbed members suggest that 32 members is too small an ensemble size. Thus, at least some of the problems should be related to the ensemble size, although other problems could be related to only part of the analysis error lying in the sub-space spanned by the initial perturbation.

The skill of the EPS ensemble-mean has been compared with the skill of the control forecast. Results show that the ensemble-mean is slightly more skilful, with acc differences at forecast day 7 for NH (Europe) between 0.02 (0.01) and 0.05 (0.05). The analysis of the correlation between the ensemble spread with respect to the ensemble-mean and the skill of the ensemble-mean has demonstrated that the correspondence between small spread/high ensemble-mean skill is smaller than the correspondence measured using the control forecast.

Concluding, the analysis of the first 18 months of daily EPS has highlighted the potential of ensemble prediction, together with the weaknesses of the present system. The results reported in this paper should provide the users with an updated figure, so that the possible economic value of the quantification of forecast uncertainty could be estimated (see, for example, *Wilks and Hamill*, 1995, who examined the potential economic value of ensemble-based forecasts of surface weather elements).

Current experimentation of a new ensemble configuration with a larger size, and based on a higher resolution forecast model, indicates that at least some of the problems highlighted in this paper should be cured in the near future.

ACKNOWLEDGEMENT

Appreciation goes to all ECMWF staff and consultants who contributed to the development of the ECMWF Integrated Forecasting System, onto which the ECMWF Ensemble Prediction System is based. Acknowledgement goes to Franco Molteni, Robert Mureau and Joe Tribbia, who contributed technically and scientifically to the implementation of the Ensemble Prediction System, and to Ron Gelaro, whose work contributed to the understanding of the relationship between singular vectors and sensitivity fields. Tim Palmer is also acknowledged for the inspiring discussions we had on very different aspects of ensemble prediction, and Adrian Simmons for carefully revising a first version of this paper.

REFERENCES

Buizza, R, 1994a. Sensitivity of optimal unstable structures. QJR Meteorol Soc, **120**, 429-451.

Buizza, R, 1994b. Localization of optimal perturbations using a projection operator. QJR. Meteorol Soc, **120**, 1647-1681.

Buizza, R, 1995. Optimal perturbation time evolution and sensitivity of ensemble prediction to perturbation amplitude. QJR Meteorol Soc, **121**, 1705-1738.

Buizza, R and T N Palmer, 1995. The singular vector structure of the atmospheric general circulation. J Atmos Sci, **52**, 9, 1434-1456.

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

Buizza, R, R Gelaro, F Molteni and T N Palmer, 1995. Predictability studies using high resolution singular vectors. QJR Meteorol Soc, submitted.

Courtier, P, C Freyder, J F Geleyn, F Rabier and M Rochas, 1991. The Arpege project at Météo-France. Pp 192-231 in Proceedings of the ECMWF seminar on Numerical methods in atmospheric models, ECMWF, Shinfield Park, Reading RG2 9AX, 9-13 September 1991, Vol 2.

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

Errico, E R and M Ehrendorfer, 1995. Moist singular vectors in a primitive-equation regional model. American Meteorological prep-prints of the Tenth conference on atmospheric and oceanic waves and stability, June 5-9, 1995, Big Sky, Montana, 272 pp.

Fleming, R J, 1971a. On stochastic dynamic prediction. I: the energetics of uncertainty and the question of closure. Mon Wea Rev, **99**, 851-872.

Fleming, R J, 1971b. On Stochastic dynamic prediction. II: predictability and utility. Mon Wea Rev, **99**, 927-938.

Gleeson, T A, 1970. Statistical-dynamical predictions. J Appl Meteorol, **9**, 333-344.

Hartmann, D L, R Buizza and T N Palmer, 1995. Singular vectors: the effect of spatial scale on linear growth of disturbances. J Atmos Sci, **52**, 22, 3885-3894.

Hollingsworth, A, 1980. An experiment in Monte Carlo forecasting procedure. ECMWF workshop on Stochastic dynamic prediction, ECMWF, Shinfield Park, Reading RG2 9AX, UK. 99 pp.

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

Jacob, C, 1994. The impact of the new cloud scheme on ECMWF's Integrated Forecasting System (IFS). Proceedings of the ECMWF/GEWEX workshop on Modelling, validation and assimilation of clouds, ECMWF, Shinfield Park, Reading RG2 9AX, November 1994.

Ledermann, 1984. Handbook of applicable mathematics. Statistics, volume VI, part B. J Wiley and sons, 942 pp.

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

Lorenz, E N, 1982. Atmospheric predictability experiments with a large numerical model. Tellus, **34**, 505-513.

Lott, F, and M Miller, 1995. A new sub-grid scale orographic drag parametrization: its formulation and testing. ECMWF Research Department Technical Memorandum n. 218, ECMWF, Shinfield Park, Reading RG2 9AX. Also Q J R Meteorol Soc , submitted.

Molteni, F and T N Palmer, 1993. Predictability and finite-time instability of the northern winter circulation. Q J R Meteorol Soc, **119**, 1088-1097.

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

Mureau, R, F Molteni and T N Palmer, 1993. Ensemble prediction using dynamically-conditioned perturbations. Q J R Meteorol Soc, **119**, 269-298.

Palmer, T N, F Molteni, R Mureau, R Buizza, P Chapelet and J Tribbia, 1993. Ensemble prediction. ECMWF Seminar Proceedings on Validation of models over Europe: Vol. 1, ECMWF, Shinfield Park, Reading RG2 9AX, UK, 285 pp.

Rabier, F, E Klinker, P Courtier and A Hollingsworth, 1995. Sensitivity of forecast error to initial conditions. Q J R Meteorol Soc, **122**, 121-150.

Simmons, A J, D M Burridge, M Jarraud, C Girard and W Wergen, 1989. The ECMWF medium-range prediction models development of the numerical formulations and the impact of increased resolution. Meteorol Atmos Phys, **40**, 28-60.

Simmons, A J, R Mureau and T Petroliagis, 1995. Error growth and predictability estimates for the ECMWF forecasting system. ECMWF Research Department Technical Memorandum No. 201, ECMWF, Shinfield Park, Reading RG2-9AX. Also Q J R Meteorol Soc, submitted.

Strang, G, 1986. Introduction to applied mathematics. Wellesley-Cambridge Press, 758 pp.

Strauss, B and A Lanzinger, 1995. Validation of the ECMWF EPS. Proceedings of the ECMWF Seminar on Predictability, 4-8 September 1995, ECMWF, Shinfield Park, Reading, RG2 9AX, UK.

Tibaldi, S, T N Palmer, Č Brancović and U Cubasch, 1990. Extended-range predictions with ECMWF models: influence of horizontal resolution on systematic error and forecast skill. Q J R Meteorol Soc, **116**, 835-866.

Tiedtke, M, 1993. Representation of clouds in large-scale models. Mon Wea Rev, **121**, 11, 3040-3060.

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

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

Viterbo, P and A C M Beljaars, 1995. An Improved land surface parametrization scheme in the ECMWF model and its validation. J Clim, **8**, 2716-2748.

Wilks, D S and T M Hamill, 1995. Potential economic value of ensemble-based surface weather forecasts. Mon Wea Rev, **123**, 3656-3575.