

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, United Kingdom

(Manuscript received 6 February 1996, in final form 11 June 1996)

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 1 unperturbed nonlinear 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-h time period, computed at T42L19 resolution.

This work describes the performance of the ECMWF EPS during the first 21 months of daily operation, from 1 May 1994 to 31 January 1996, 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 for the ECMWF EPS.

Finally, the EPS is validated against analysis fields, and the EPS skill is compared with the skill of the perfect ensemble. 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 so-called faulty index is introduced to estimate the percentage of wrongly predicted cases with small spread/high control skill. Results suggest that there is some correspondence between small ensemble spread and high control skill. Considering the 500-hPa geopotential height field over the Northern Hemisphere at forecast day 7, approximately 20% (45%) of the perturbed ensemble members have anomaly correlation skill higher than 0.6 during warm (cold) seasons, respectively. The percentage of analysis values lying outside the EPS forecast range is thought to be still too 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 Forecasts (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 skillful or unskillful, using only deterministic prediction. By contrast, ensemble prediction has the ca-

pability 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 National Centers for Environmental Prediction (NCEP, formerly the National Meteorological Center) 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 centers 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 the others with initial conditions defined adding small perturbations to the control initial condition. Apart from differences in the ensemble size and the fact that at NCEP a combination of lagged forecasts is used, the NCEP and the ECMWF approaches to ensemble prediction differ sub-

Corresponding author address: Dr. Roberto Buizza, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG12 9AX, United Kingdom.
E-mail: buizza@ecmwf.int

stantially 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 NCEP, 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 [see 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 three-level quasigeostrophic 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). A methodology is proposed to estimate the potential forecast skill of ensemble prediction, by defining a “perfect ensemble” system with the same characteristics as the EPS in terms of ensemble size and type of initial perturbations (i.e., with 32 perturbed members with initial conditions generated using singular vectors). The perfect ensemble is defined as an ensemble of integrations of a perfect model, which includes the analysis within the range of the ensemble forecasts. These two hypotheses are fulfilled by considering a randomly chosen perturbed member as the verifying analysis. Working within these hypotheses, we evaluate the characteristics of the 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 the following three requirements: (i) the ensemble spread should be comparable to the error of the control forecast, (ii) small spread should indicate a skillful control forecast, and (iii) the verifying analysis should be included within the range covered by the ensemble forecasts. The work is based on 21 months of daily ensemble prediction, from 1 May 1994 (when daily EPS started being run at ECMWF) to 31 January 1996.

Finally, in the third part of the paper, we consider the ECMWF EPS and, focusing on the three requirements mentioned above, we compare the performance of the ECMWF EPS and the perfect ensemble. 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 1 unperturbed (control) nonlinear integrations of a T63L19 Eulerian version of the ECMWF model (Simmons et al. 1989; 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.

a. Singular vectors computation

Denote by \mathbf{x} a perturbation, and by

$$\mathbf{x}(t) = \mathbf{L}\mathbf{x}_0 \quad (1)$$

the solution of the linear model equations is

$$\frac{\partial \mathbf{x}}{\partial t} = \mathbf{A}_1 \mathbf{x}, \quad (2)$$

where \mathbf{L} is the forward linear propagator of Eq. (2), and \mathbf{A}_1 is an approximation of the tangent linear version of the ECMWF model \mathbf{A} .

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

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

where \mathbf{L}^* is the adjoint of the linear propagator \mathbf{L} with respect to the total energy scalar product $(;)$ [Buizza et al. 1993, their Eq. (5.1)], and \mathbf{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 $\phi \geq 30^\circ\text{N}$. It is worth pointing

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

Horizontal resolution	T42 spectral triangular truncation
Vertical resolution	19 vertical levels
Optimization time interval	48 h
Diabatic schemes in direct	Full physics
Diabatic schemes in forward/adjoint	Linear physics and horizontal diffusion
Local projection operator	$\lambda \geq 30^\circ\text{N}$

out that the implementation of the local projection operator \mathbf{T} was necessary for the EPS to compute, with a limited number of iterations, perturbations also growing in the NH during the NH warm seasons (Buizza 1994b).

The singular vectors \mathbf{v}_i of the propagator \mathbf{TL} are the perturbations with localized maximum energy growth over the optimization time interval t . They are computed solving the eigenvalue problem

$$\mathbf{L}^*\mathbf{T}^2\mathbf{L}\mathbf{v}_i = \sigma_i^2\mathbf{v}_i. \quad (4)$$

Specifically, they are the eigenvectors \mathbf{v}_i with maximum eigenvalues σ_i^2 of the operator $\mathbf{L}^*\mathbf{T}^2\mathbf{L}$. 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 a randomly chosen case, 5 November 1995 (the singular vectors are ranked with respect to the singular values). For this initial date, 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 of the 5 November 1995 singular vectors at initial and optimization time. The first singular vector (Figs. 1a,b) is growing across the eastern border of the Asian continent, a region characterized by a very intense and rapid development, as can be detected both at 500 and 1000 hPa (Fig. 3). By contrast, the third (Figs. 2c,d) and the sixth (Figs. 2e,f) singular vectors are amplifying in relatively less unstable regions, as is reflected by their smaller singular values (Fig. 1). 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 (Fig. 4), especially at optimization time. In fact, while

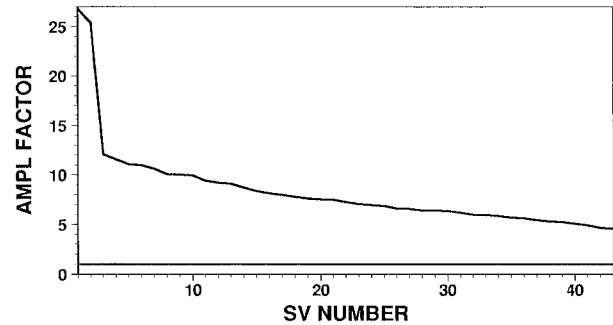


FIG. 1. Singular values σ_i for 5 November 1995 (the singular vectors are ranked with respect to their amplification rate).

the total energy of the first singular vector (Fig. 4a) has double maxima in the vertical, with a predominant low-level growth, the third (Fig. 4b) and the sixth (Fig. 4c) singular vectors have a (more commonly detected) vertical profile peaking at optimization time around model level 9.

These 3 singular vectors are among the 16 selected from the 38 computed singular vectors. The selection criteria are such that the first four singular vectors are always chosen, and each subsequent singular vector (from the fifth onward) is selected only if half of its total energy lies outside the regions where the singular vectors already selected are localized.

b. Generation of 16 perturbations

Once the 16 singular vectors have been selected, an orthogonal rotation in phase space and a final rescaling are performed to construct the ensemble perturbations. The purpose of the phase-space rotation is to generate perturbations with the same globally averaged energy as the singular vectors, but smaller local maxima and a more uniform spatial distribution. Moreover, unlike the singular vectors, the rotated singular vectors are characterized by similar amplification rates (at least up to 48 h). Thus, the rotated singular vectors diverge, on average, equally from the control forecast. 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 rescaling 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 rescaling procedure can be seen by comparing the singular vectors of Fig. 2 with three initial perturbations shown in Fig. 5.

c. Nonlinear integrations

The 16 initial perturbations are specified in terms of the spectral coefficients of the three-dimensional vorticity, divergence, and temperature fields (no perturba-

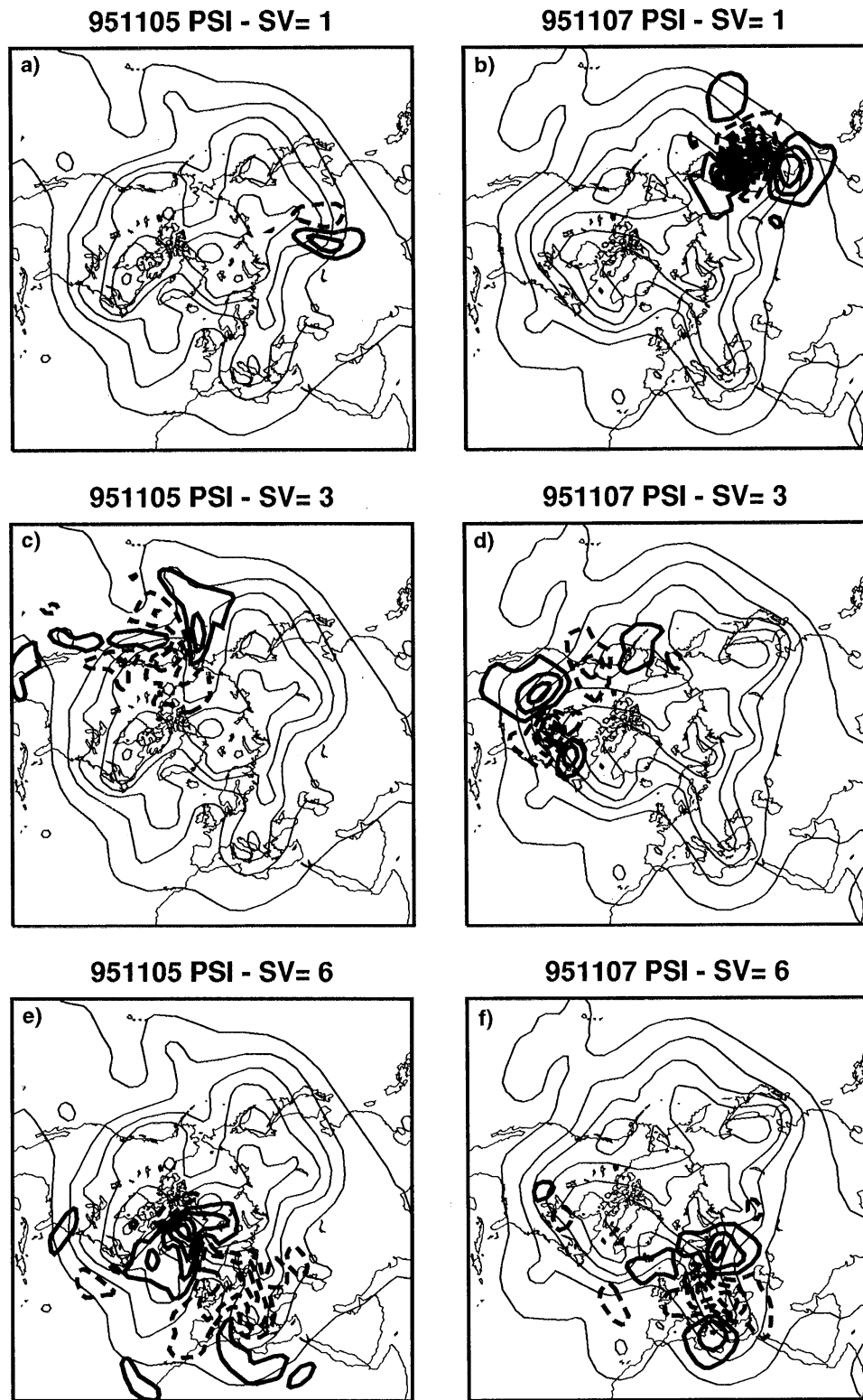


FIG. 2. Singular vectors numbers 1 (top panels), 3 (middle panels), and 6 (bottom panels) at initial (left panels) and optimization time (right panels). Each panel shows the singular vector streamfunction at model level 11 (approximately 500 hPa), superimposed to the trajectory 500-hPa geopotential height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

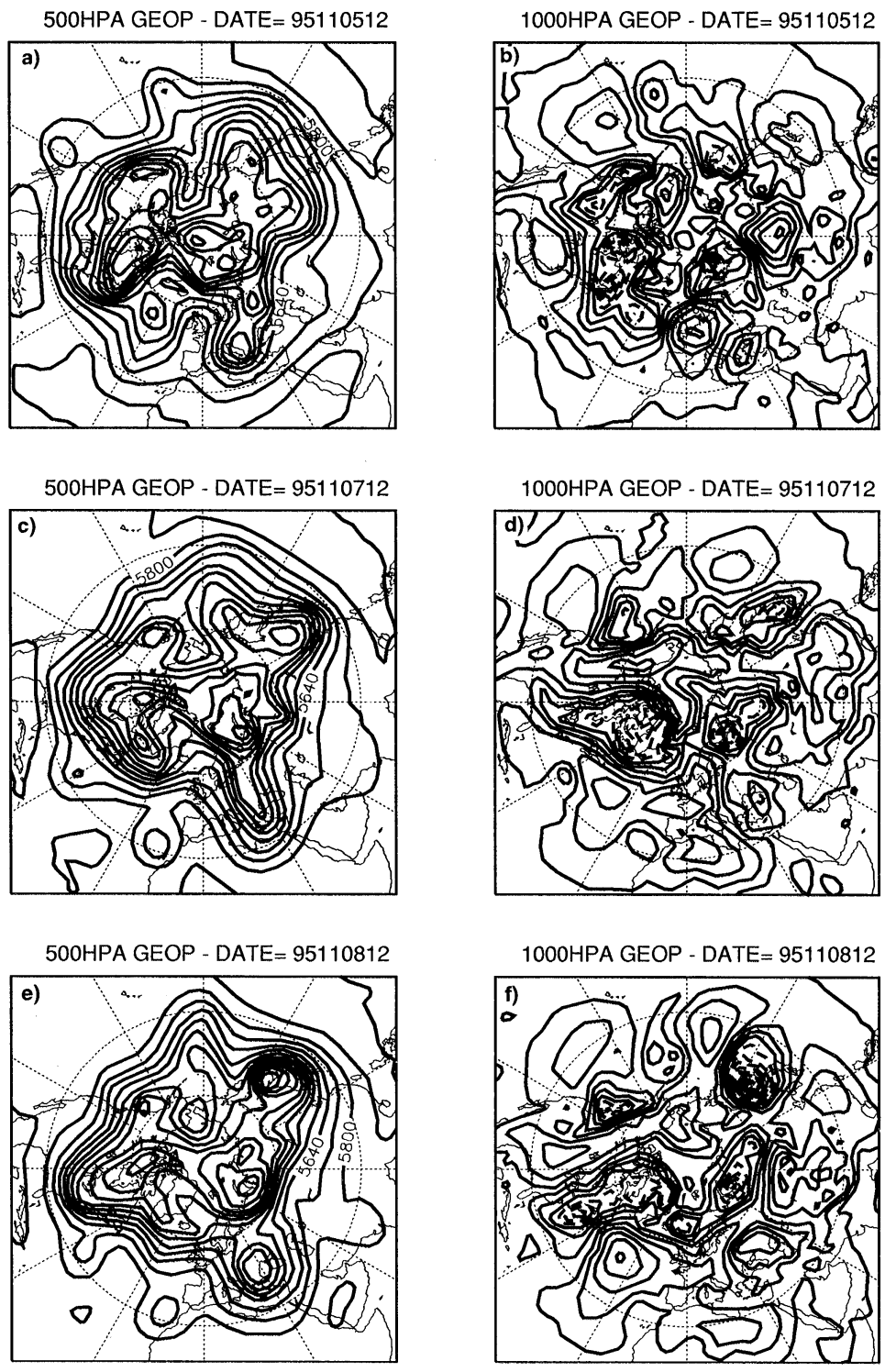


FIG. 3. The 500-hPa (left panels) and 1000-hPa (right panels) geopotential height (left panels) trajectory for 5, 7, and 8 November 1995. Contour intervals are 80 m for left panels and 40 m for right panels.

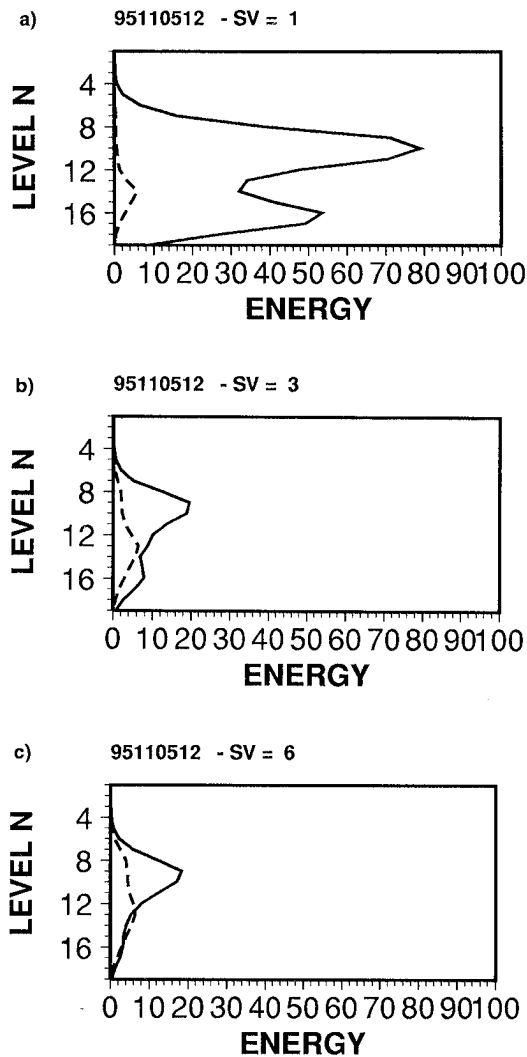


FIG. 4. Total energy ($\text{m}^2 \text{s}^{-2}$) vertical profile of the (a) first, (b) third, and (c) sixth singular vector of 5 November 1995, at initial (dashed line, values multiplied by 100) and optimization (solid line) time. (Note that singular vectors are normalized to have unit initial total energy norm.)

tions are defined for the specific humidity since the singular vector computation is performed with a dry linear forward/adjoint model), and of the two-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 nonlinear 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 h (phase described in section 2a), the elapsed time needed for generating the initial perturbations is 0.1 h (phase described in section 2b), and the elapsed time needed to perform the 33 nonlinear integrations is approximately 1.8 h. Thus, the total elapsed time is approximately 2.7 h, which is

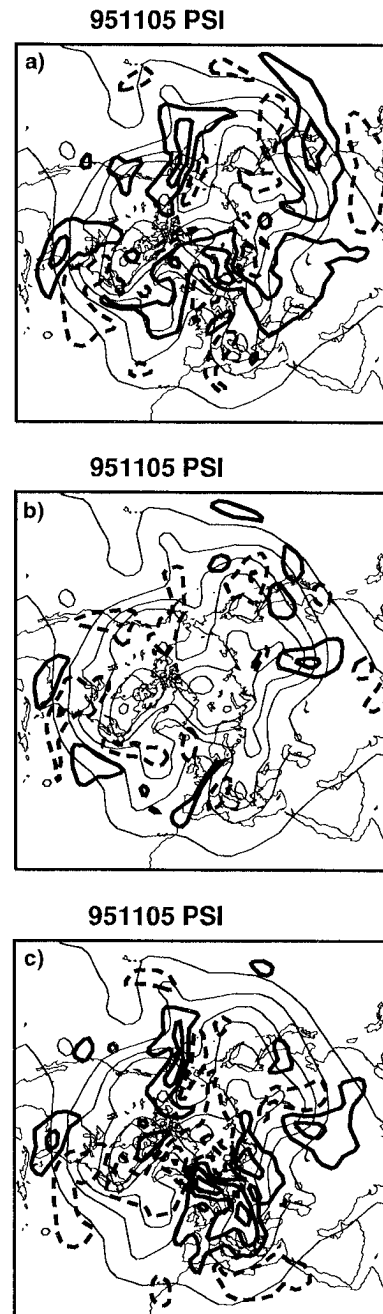


FIG. 5. As in Fig. 2 but for three perturbations (streamfunction) generated by applying a phase-space rotation to the selected singular vectors, superimposed to the trajectory 500-hPa geopotential height field. Contour interval is $0.25 \times 10^{-8} \text{ m}^2 \text{ s}^{-1}$ for streamfunction and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

about 1.3 times the elapsed time needed to performed the 10-day T213L31 ECMWF operational forecast.

A first way of verifying the EPS is to analyze the spread and the skill characteristics over different areas (see Fig. 6 for the 5 November 1995 case over the NH). The spread of a perturbed ensemble member is defined as the anomaly correlation or root-mean-square distance

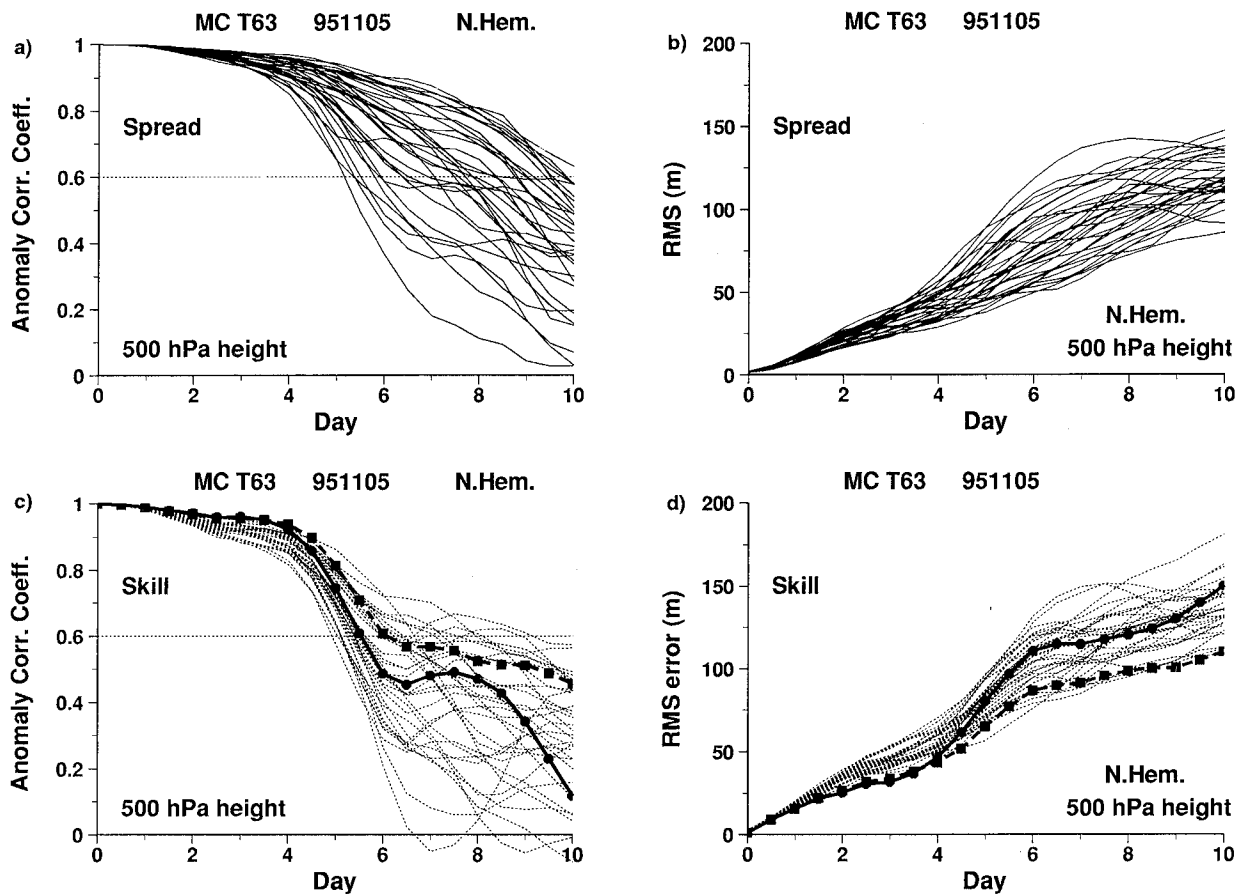


FIG. 6. Spread (top panels) and skill (bottom panels) of the EPS run with starting date 5 November 1995, computed for the 500-hPa geopotential height over the NH. Left panels refer to acc and right panels to rms values. Bottom panels: dotted lines refer to perturbed members, bold solid lines with dots refer to the control forecast, and bold dashed lines with squares refer to the ensemble mean.

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).

d. 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 December 1992) occurred on 4 August 1993, when the new ECMWF surface and boundary layer scheme was introduced (Viterbo and Beljaars 1995), and on 4 April 1995, when the new ECMWF prognostic cloud scheme (Tiedtke 1993; Jacob 1994) and a new scheme for the representation of the subgrid-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.

1) 19 DECEMBER 1992: EPS STARTS THREE DAYS A WEEK

After a month of trial, on 19 December 1992 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 h, the optimization time interval was set to 36 h (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 1 unperturbed 10-day T63L19 nonlinear integrations.

2) 20 MARCH 1993: 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 $\phi \geq 30^\circ N$ (due to the lack of physics in the linear forward and adjoint system, we decided not to set the southern boundary to lower latitudes). The impact of the local projection operator was proven to be almost negligible during the NH winter season. A further effect of its implementation was a small increase of the ensemble spread over the NH.

3) 1 MAY 1994: DAILY EPS

From this date onward the EPS has been run daily.

4) 23 AUGUST 1994: LARGER INITIAL AMPLITUDE AND LONGER OPTIMIZATION 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 reduce the day 2 forecast error the most. 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 optimization time interval to 48 h.

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 optimization time. In other words, an upper boundary to the optimization time interval used when computing the singular vectors is given by the time limit of the 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 h, and scaled with $\alpha \leq \sqrt{4}$, was linear to a good extent.

5) 14 MARCH 1995: 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 at optimization time, 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. Seven 92-day seasons are analyzed, from 1 May 1994 to 31 January 1996. The reader is referred to Strauss and Lanzinger (1995) for a description of other means of validation of the EPS (e.g., reliability diagrams).

a. Spread and skill

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

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

where $w(\mathbf{x}_g) \equiv \cos(\phi_{\mathbf{x}_g})$, $\phi_{\mathbf{x}_g}$ being the latitude of \mathbf{x}_g . Let $\| \cdot \|_g$ denote 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,’’

$$d_{\text{acc}}(f, h; t) \equiv \frac{\langle [f(t) - \text{cli}(t)]; [h(t) - \text{cli}(t)] \rangle_g}{\| [f(t) - \text{cli}(t)] \|_g \| [h(t) - \text{cli}(t)] \|_g}, \quad (6)$$

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

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

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 acc values are computed after applying a so-called Fisher z transform (see, e.g., Ledermann 1984). Specifically, given the set of acc values $d_{\text{acc}}(f_j, h; t)$, with $j = 1, \dots, N$, first, the transformed values z_j are defined as

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

second, the average among the z_j is computed as

$$\bar{z} \equiv \frac{1}{N} \sum_{j=1}^N z_j; \quad (8b)$$

and finally the average acc $d_{\text{acc}}(f, h; t)$ is computed by applying the reverse Fisher z transform

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

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

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

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 rms distances as

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

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

where f_j is the j th perturbed ensemble member, “con” is the control forecast, and Eqs. (8a)–(8c) and (9) have been applied.

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

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

The *skill* of a forecast f , f being 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 by 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

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

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

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

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

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

and the probability distribution function of the ensemble rms error as

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

Similarly, considering the spread, we denote by

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

and

$$P_{rms}^{sp}(x), \quad 0 \leq x, \quad (14b)$$

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

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

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 and high/low skill will be analyzed, with the categories defined by the average values (with this choice the tables will not be necessarily symmetric).

b. Probability of the analysis lying outside the EPS forecast range

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

$$da_1 \equiv a(\mathbf{x}, t), \quad \text{if } -\infty \leq a(\mathbf{x}, t) < f_1, \quad (15a)$$

$$da_j \equiv a(\mathbf{x}, t); \quad \text{if } f_{j-1} \leq a(\mathbf{x}, t) < f_j, \\ j = 2, \dots, 32, \quad (15b)$$

$$da_{33} \equiv a(\mathbf{x}, t), \quad \text{if } f_{32} \leq a(\mathbf{x}, t) \leq f_{33}, \quad (15c)$$

$$da_{34} \equiv a(\mathbf{x}, t), \quad \text{if } f_{33} < a(\mathbf{x}, t) \leq \infty. \quad (15d)$$

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, da_1 and da_{34} .

In the perfect model, following Eqs.(15a)–(15d), the probability of the analysis being inside the first and the last intervals should be nil, since by definition of perfect ensemble, there are always two perturbations that are guaranteed to produce the largest possible deviation from the control at all points and all time ranges. Moreover, the probability of being inside each of the other 32 categories when considering a sufficiently large number of cases, should be identical, that is, 1/32 in case of a 32 + 1 EPS.

Instead of this, in case of the EPS being verified against the analysis field and when considering a sufficiently large number of cases, one should expect the probability of the analysis being inside each of the 34 categories to be 1/34.

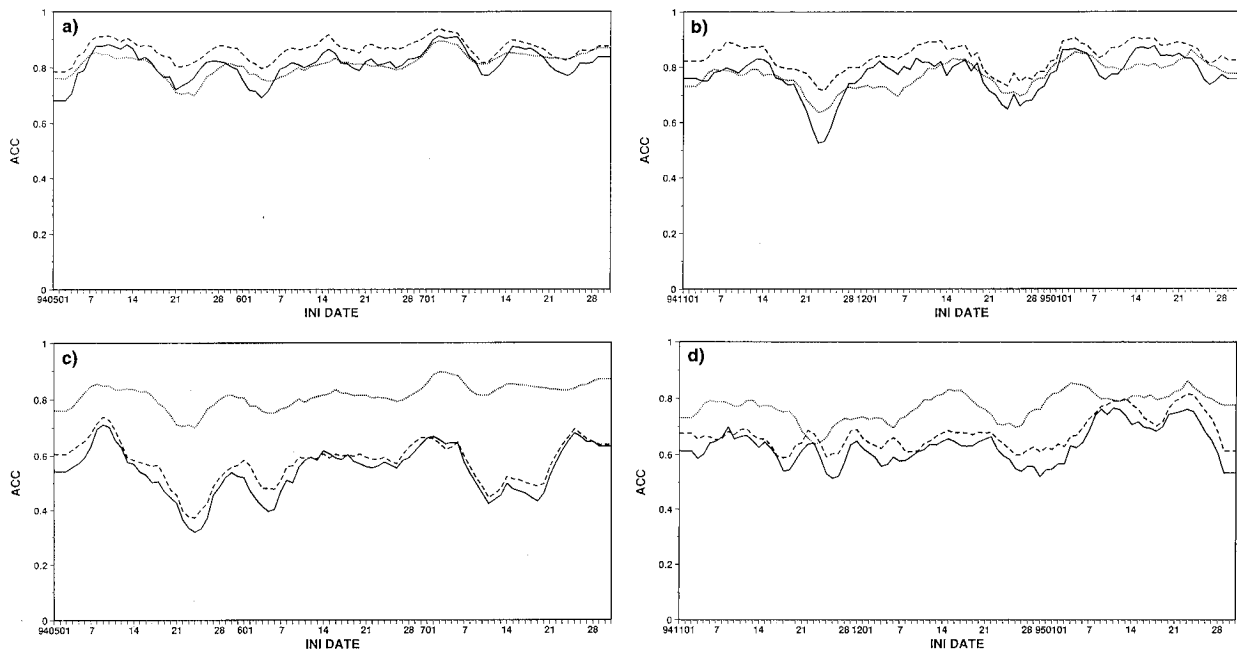


FIG. 7. Control skill (solid), ensemble mean skill (dashed), and average spread (dotted), at forecast day 7, for (a) 1 May 1994–31 July 1994, perfect ensemble; (b) 1 November 1994–31 January 1995, perfect ensemble; (c) as in (a) but for the EPS verified against the analysis; (d) as in (b) but for the EPS verified against the analysis. Values refer to 500-hPa geopotential height over the NH.

4. Potential forecast skill of ensemble prediction

Consider the following three requirements.

- 1) 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).
- 2) Small spread (around the control) should indicate a skillful control forecast.
- 3) The verifying analysis should be included within the range covered by the ensemble forecasts.

These requirements should be verified by the perfect ensemble defined in the introduction (i.e., by an ensemble of integrations of a perfect model, which includes the analysis within the forecast range). 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 verify whether requirements (i)–(iii) are true for the perfect ensemble. Let us remind the reader that the perfect ensemble is defined by considering a randomly chosen ensemble member as the verifying analysis. This guarantees that the verifying anal-

ysis always lies in the subspace 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 days 5 to 7 (Table 2, values in parenthesis refer to the perfect ensemble, computed not only over the 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 analyzed seasons (Fig. 8). It is worth noting that comparable spread and control skill does 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 the NH). In fact, by construction, all ensemble members are more distant from the analysis than the control up to forecast day 2, the time limit up to which the initial perturbations dynamics can be linearly approximated (Buizza 1995). After this time limit, nonlinearity could possibly bring some ensemble members closer to the analysis than the control, but would probably bring the most of them farther away.

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

	Northern Hemisphere		Europe	
	Day 2	Day 5 to 7	Day 2	Day 5 to 7
1 May 94–31 Jul 94	1.43 (0.99)	1.56 (0.96)	1.30 (0.99)	1.50 (0.97)
1 Aug 94–31 Oct 94	1.10 (1.03)	1.22 (0.97)	0.96 (1.00)	1.22 (1.00)
1 Nov 94–31 Jan 95	1.06 (1.03)	1.27 (0.98)	0.97 (1.03)	1.19 (0.99)
1 Feb 95–3 May 95	1.06 (1.01)	1.23 (1.02)	0.96 (0.99)	1.16 (1.03)
1 May 95–31 Jul 95	1.31 (1.02)	1.38 (1.00)	1.29 (0.98)	1.32 (1.00)
1 Aug 95–31 Oct 95	1.28 (1.03)	1.28 (1.01)	1.11 (1.05)	1.16 (1.02)
1 Nov 95–31 Jan 96	1.20 (1.02)	1.34 (1.01)	1.23 (1.10)	1.01 (1.04)

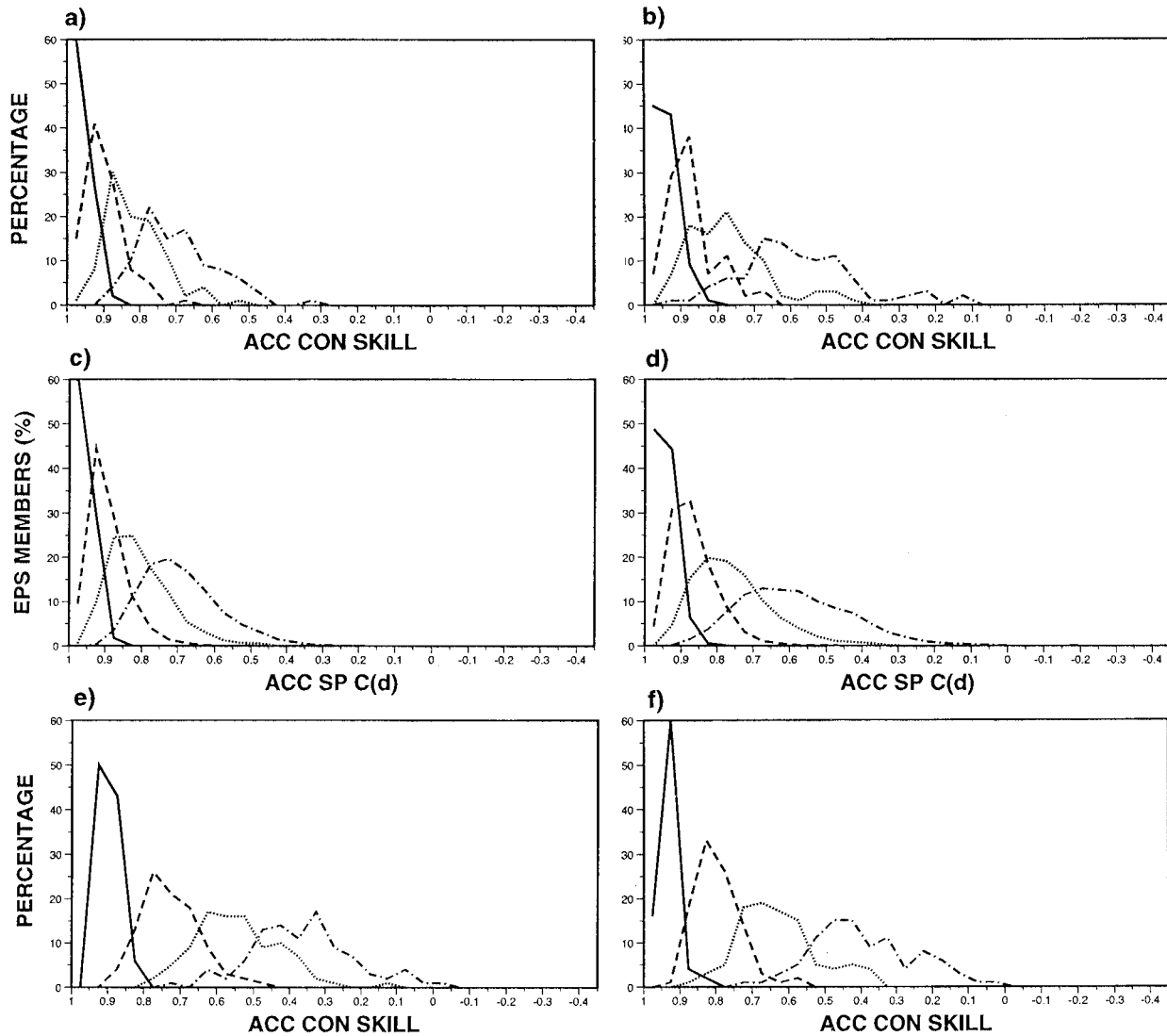


FIG. 8. Distribution of the control skill verified against a perturbed ensemble member (perfect ensemble) for (a) 1 May 1994–31 July 1994 and (b) 1 November 1994–31 January 1995 at forecast day 3 (solid), day 5 (dashed), day 7 (dotted), and day 10 (chain dashed); (c), (d) as in (a), (b) but for the spread distributions; (e), (f) as in (a), (b) but for the distribution of the control skill verified against the analysis. Values refer to the NH.

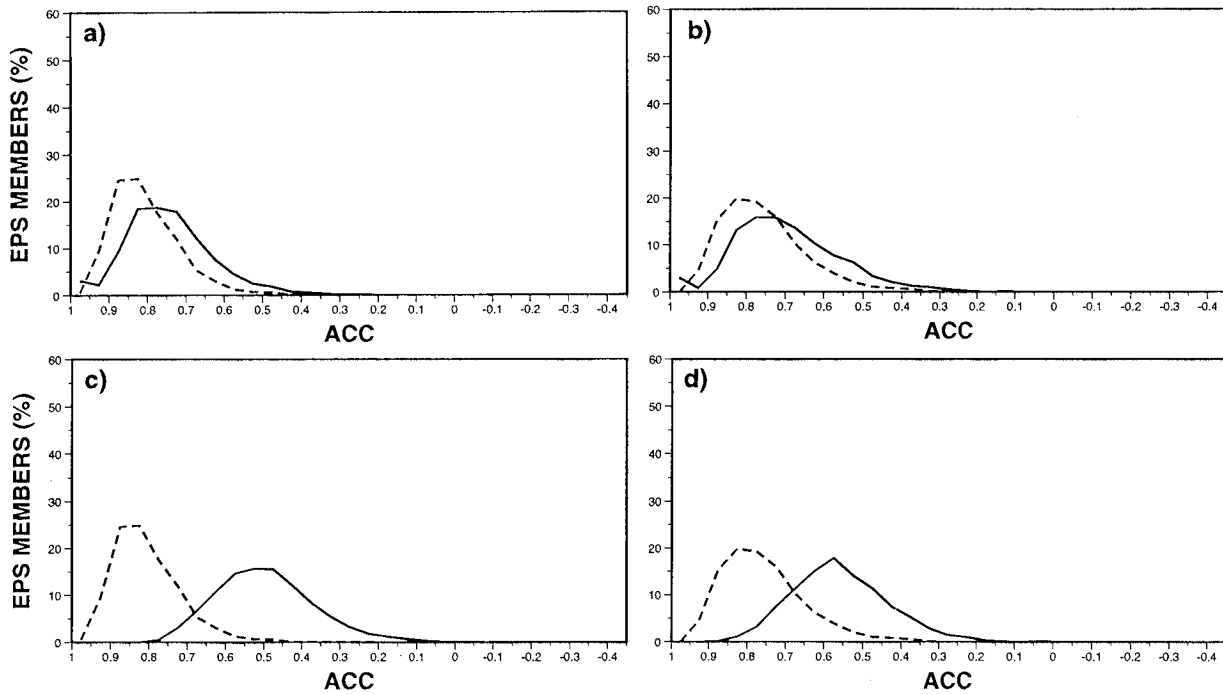


FIG. 9. Perfect ensemble spread distribution (solid) and skill distribution (dashed) for (a) 1 May 1994–31 July 1994 and (b) 1 November 1994–31 January 1995; (c),(d) EPS spread and skill distribution, with skill verified against the analysis. Values refer to the NH at forecast day 7.

Figures 7a,b also show that there is temporal correlation between spread and skill. Tables 3 and 4 report correlation coefficients for the NH and Europe at forecast day 7 for all the seasons analyzed in this paper (values in parenthesis refer to the perfect ensemble). Correlation coefficients range between 0.50 and 0.65 for the 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 skillful control forecast also with the 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. One case study has been analyzed. Two ensembles have been generated, one with all 32 perturbed members and one with only 31 members, defined by neglecting the ensemble member used for the verification in the perfect ensemble. Figure 10a shows the spread of the 32 ensemble members around the control unperturbed forecast. Figure 10b shows the skill of the 32 perturbed members (dotted lines), the control (bold solid lines with dots), and the ensemble mean (bold solid lines with squares) of the 32-member ensemble. (The ensemble member used as verifying analysis is the one with the acc skill of 1.) By contrast, Fig. 10c shows the skill of the 31 perturbed members, the control, and the ensemble mean of the 31-member ensemble. Since the

ensemble spreads of the 32-member and the 31-member ensembles are almost identical, the spread/skill relations are practically the same. As a consequence, the ensemble skill distributions differ only slightly. The percentages of analysis values lying outside the ensemble forecast range increase from nil to 4% (3%) and 8% (19%), respectively, at forecast days 7 and 10 over the NH (Europe).

5. EPS validation from 1 May 1994 to 31 January 1996

Let us first consider the two 92-day periods starting on 1 May 1994 and 1 November 1994, already discussed in detail in the previous section, and let us compare the performances of the EPS and the perfect ensemble. Note that 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 error 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 they use the analysis instead of a randomly chosen ensemble member as verification. The control skill (solid lines)

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

Northern Hemisphere—Forecast day 7			
		Low skill	High skill
1 May 94–31 Jul 94	<i>cc</i> = 0.50 (0.65) small spread large spread <i>i_{faulty}</i> = 0.23	17 (13) 28 (33)	31 (35) 16 (11)
1 Aug 94–31 Oct 94	<i>cc</i> = 0.40 (0.56) small spread large spread <i>i_{faulty}</i> = 0.08	12 (11) 27 (40)	28 (29) 25 (12)
1 Nov 94–31 Jan 95	<i>cc</i> = 0.23 (0.64) small spread large spread <i>i_{faulty}</i> = 0.31	22 (15) 24 (35)	25 (32) 21 (10)
1 Feb 95–3 May 95	<i>cc</i> = 0.27 (0.50) small spread large spread <i>i_{faulty}</i> = 0	17 (17) 29 (31)	30 (30) 16 (14)
1 May 95–31 Jul 95	<i>cc</i> = 0.26 (0.55) small spread large spread <i>i_{faulty}</i> = 0.22	18 (12) 29 (37)	25 (31) 20 (12)
1 Aug 95–31 Oct 95	<i>cc</i> = 0.39 (0.55) small spread large spread <i>i_{faulty}</i> = 0.29	17 (12) 31 (34)	30 (35) 14 (11)
1 Nov 95–31 Jan 96	<i>cc</i> = 0.52 (0.68) small spread large spread <i>i_{faulty}</i> = 0.21	14 (11) 35 (35)	32 (35) 11 (11)

is lower than the average spread (dotted lines). High values of acc spread (around 0.8 and 0.75, respectively, during the seasons started on 1 May 1994 and 1 November 1994) indicate 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 (Figs. 8c,d) and the control skill distribution verified against the analysis (Figs. 8e,f).

Table 2 shows that the lack of spread is particularly noticeable during the NH warm seasons (cf. the value of 1.43 for the season started on 1 May 1994 against the value of 1.06 for the season started on 1 November 1994 over the NH at day 2). 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 during the NH warm seasons than during the cold seasons. Errico and Ehrendorfer (1995) showed that the inclusion of moist

TABLE 4. As in Table 3 but for Europe.

Europe—Forecast day 7			
		Low skill	High skill
1 May 94–31 Jul 94	<i>cc</i> = 0.47 (0.86) small spread large spread <i>i_{faulty}</i> = 0.46	15 (8) 39 (41)	25 (32) 13 (11)
1 Aug 94–31 Oct 94	<i>cc</i> = 0.18 (0.31) small spread large spread <i>i_{faulty}</i> = 0.06	15 (14) 25 (30)	29 (30) 23 (18)
1 Nov 94–31 Jan 95	<i>cc</i> = 0.45 (0.50) small spread large spread <i>i_{faulty}</i> = 0.18	16 (13) 33 (33)	31 (34) 12 (12)
1 Feb 95–3 May 95	<i>cc</i> = 0.43 (0.43) small spread large spread <i>i_{faulty}</i> = 0.26	15 (11) 32 (35)	26 (30) 19 (16)
1 May 95–31 Jul 95	<i>cc</i> = 0.46 (0.69) small spread large spread <i>i_{faulty}</i> = 0.42	19 (11) 28 (35)	26 (34) 19 (12)
1 Aug 95–31 Oct 95	<i>cc</i> = 0.31 (0.49) small spread large spread <i>i_{faulty}</i> = 0.30	20 (14) 30 (34)	25 (31) 17 (13)
1 Nov 95–31 Jan 96	<i>cc</i> = 0.50 (0.68) small spread large spread <i>i_{faulty}</i> = 0.38	13 (8) 35 (38)	27 (35) 14 (11)

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 days 5 and 7 than at day 2. Note that this is true also for the cold seasons (cf. the value of 1.27 against 1.06 for the season started on 1 November 1994). 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. 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) compared earlier versions of the ECMWF model at different resolutions and showed that a T63L19 resolution was not able to ensure the right level of transient activity 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 (A. Persson 1995, personal communication). The lack of model activity could be cured, at least in part, by using the higher resolution T106L19 model version when performing the nonlinear 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 all the seasons is discussed hereafter.

a. 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 August 1994 reduced the difference between spread and control error, but despite the changes implemented on 14 March 1995 (perturbations generated using T42 singular vectors but with smaller initial amplitude; see section 2d) the spread is still too small. When the 14 March 1995 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 afterward seems to be related to the change implemented on 4 April 1995 (new model version; see the first paragraph of section 2d). In fact, the model version introduced on 4 April 1995 seems to be less active and thus less able to sustain the perturbation growth [A. Simmons 1995, personal communication; this is confirmed by comparison of seasonal integrations of new and old model versions (Č. Brancović 1995, personal communication) and by results obtained by R. Gelaro (1995, personal communication), 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). If we consider, for example, the season started on 1 November 1994, Fig. 12c shows that at forecast day 5 (dashed curve) the perturbed members have acc skill between 0.95 and 0.35, and that the skill distribution peaks around 0.75. Probability distribution functions can be used to evaluate the percentage of EPS members with a skill score higher than a predefined threshold—for example, the percentage of EPS members having acc skill higher than 0.8 or 0.6 (Fig. 13). If we consider again the season started on 1 November 1994, Fig. 13c shows that at forecast day 5, for example, around 20% of the ensemble members (i.e., 6 members) have acc skill higher than 0.8 (solid curve) and that around 90% of them have acc skill higher than 0.6 (dashed curve).

b. Correspondence between small spread and high skill of the control forecast

Consistent with the results discussed above, the correlation coefficients between spread and control skill in the EPS are smaller than the correlation values of the perfect ensemble. Tables 3 and 4 report the values at forecast day 7 for the NH and Europe, respectively. Considering, for example, the season started on 1 No-

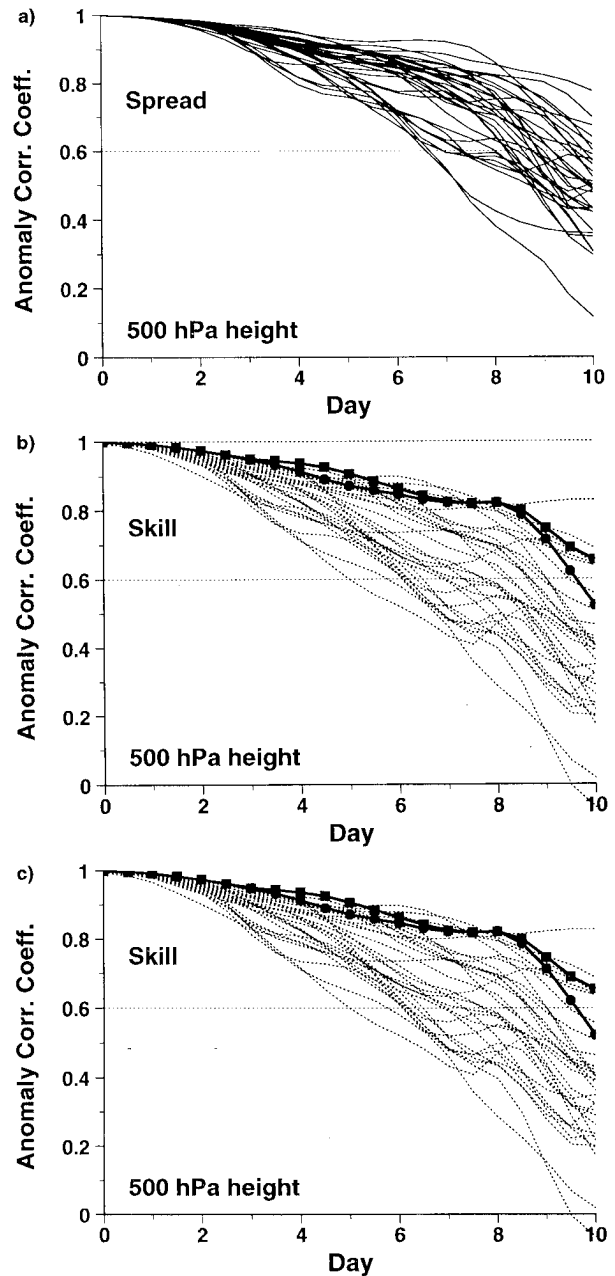


FIG. 10. (a) Spread of each member of the perfect ensemble. (b) Skill of each member of the perfect ensemble, including the ensemble member used as the analysis for verification (dotted), and the skill of the control (solid with dots) and of the ensemble mean (solid with squares). (c) As in (b) but for the perfect ensemble without the ensemble member used as the analysis for verification. Values refer to the NH 500-hPa geopotential height for 2 December 1994.

vember 1994, for Europe (Table 4), the correlation coefficient of the EPS is 0.45, while the value for the perfect ensemble is 0.50. Generally speaking, differences range from zero (for the period 1 February–3 May 1995 over Europe) to 0.41 (for the period 1 November 1994–31 January 1995 over the NH).

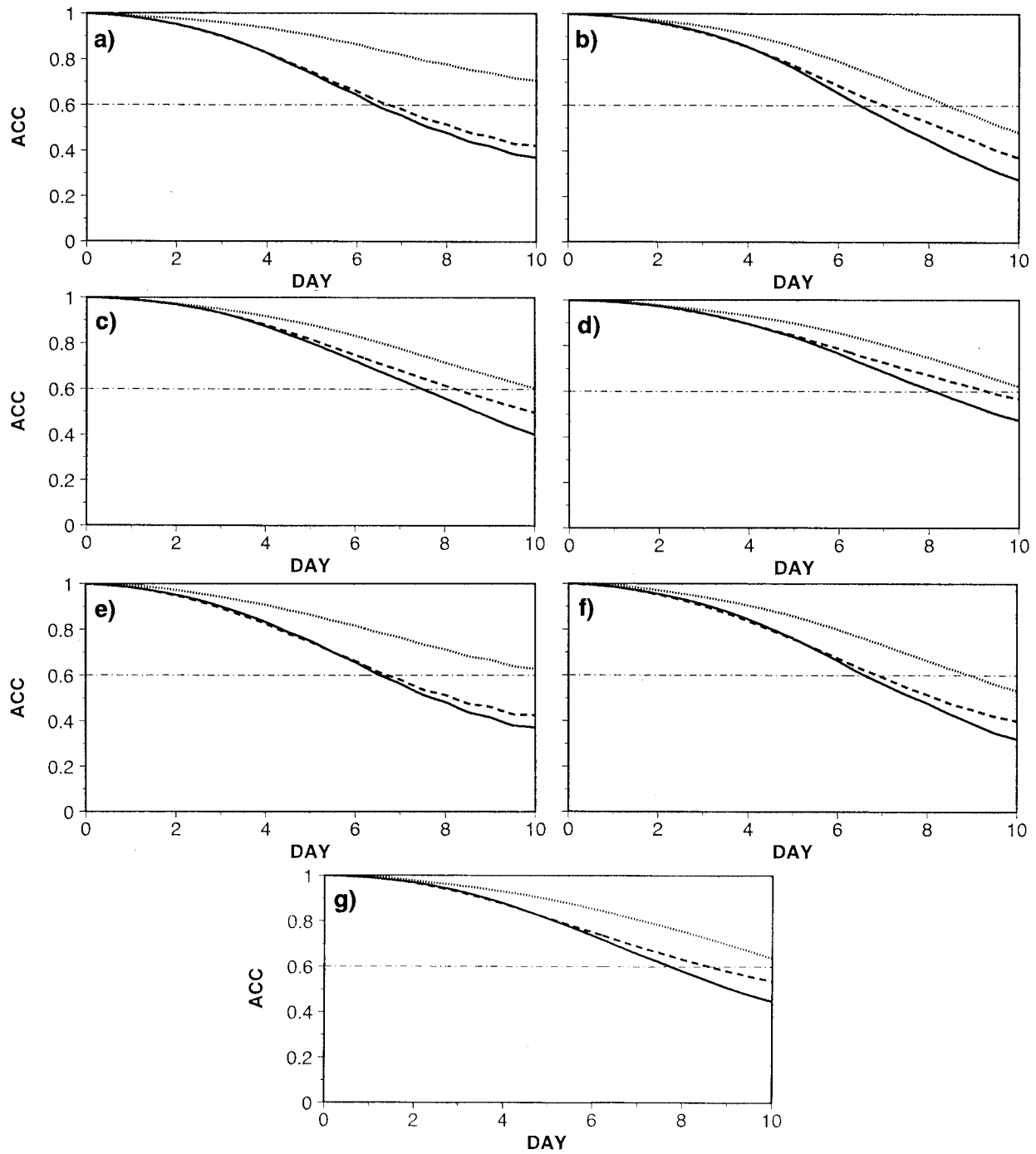


FIG. 11. Control skill (solid), ensemble-mean skill (dashed), and average spread (dotted) for the 92-day seasons starting (a) 1 May 1994, (b) 1 August 1994, (c) 1 November 1994, (d) 1 February 1995, (e) 1 May 1995, (f) 1 August 1995, and (g) 1 November 1995. Values refer to acc computed for the 500-hPa geopotential height over the NH.

Figure 14 shows the scatter diagrams of the ensemble spread versus the control skill for the 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 corre-

lation coefficient, the scatter diagrams in Figs. 14a and 14c have the highest and the lowest value, respectively.

Considering requirement (ii) of section 4, for each contingency table we can define a “faulty index” as the difference between the EPS number of small spread/low

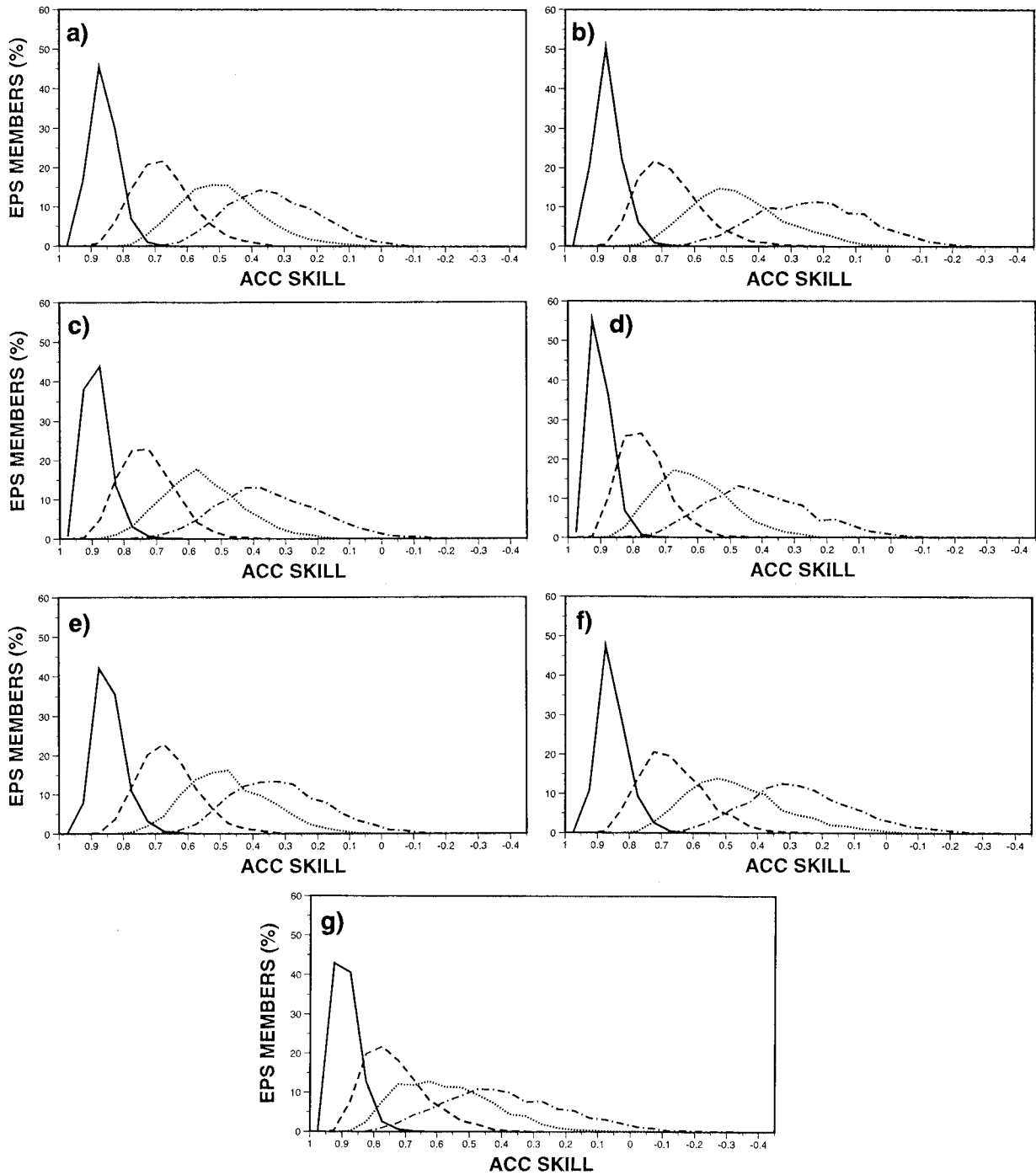


FIG. 12. Ensemble skill distributions $P_{acc}^{sk(ens)}(x)$, at forecast day 3 (solid), day 5 (dashed), day 7 (dotted), and day 10 (chain dashed), for the 92-day seasons starting (a) 1 May 1994, (b) 1 August 1994, (c) 1 November 1994, (d) 1 February 1995, (e) 1 May 1995, (f) 1 August 1995, and (g) 1 November 1995. Values refer to acc computed for the 500-hPa geopotential height over the NH.

skill predictions and the number of the small spread/low skill predictions in the perfect ensemble, divided by the number of small spread predictions. In mathematical terms

$$i_{\text{faulty}} \equiv \frac{|n_{1,1}^{\text{EPS}} - n_{1,1}^{\text{per}}|}{n_{1,1}^{\text{EPS}}}, \quad (16)$$

where n_{ij}^{EPS} identifies the contingency table element on

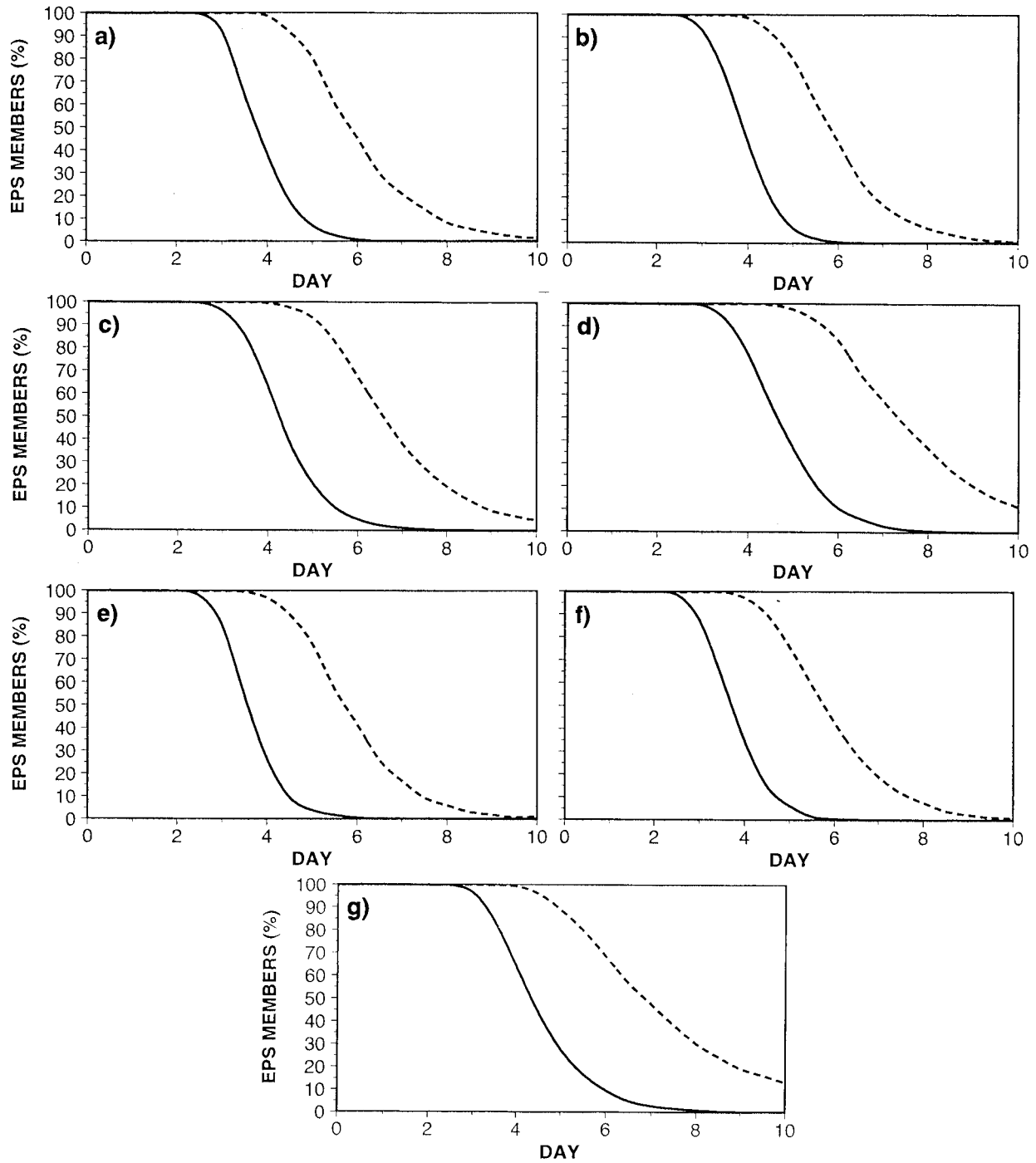


FIG. 13. Percentage of EPS members with acc skill higher than 0.8 (solid) and 0.6 (dashed) for the 92-day seasons starting (a) 1 May 1994, (b) 1 August 1994, (c) 1 November 1994, (d) 1 February 1995, (e) 1 May 1995, (f) 1 August 1995, and (g) 1 November 1995. Values refer to acc computed for the 500-hPa geopotential height over the NH.

the i th row, j th column, and where the superscript EPS or “per” refers, respectively, to the contingency table of the EPS and the perfect ensemble. [The differences in the populations of the actual and perfect contingency tables are statistically significant for all seasons apart from the one started on 1 November 1994, for which

the difference has (only) a 65% probability of being significant according to a chi-square test.]

In simple words, i_{faulty} gives the percentage of cases when forecasted high predictability did not correspond to actual high skill of the control forecast—that is, when spread is not a good predictor of control skill. Faulty

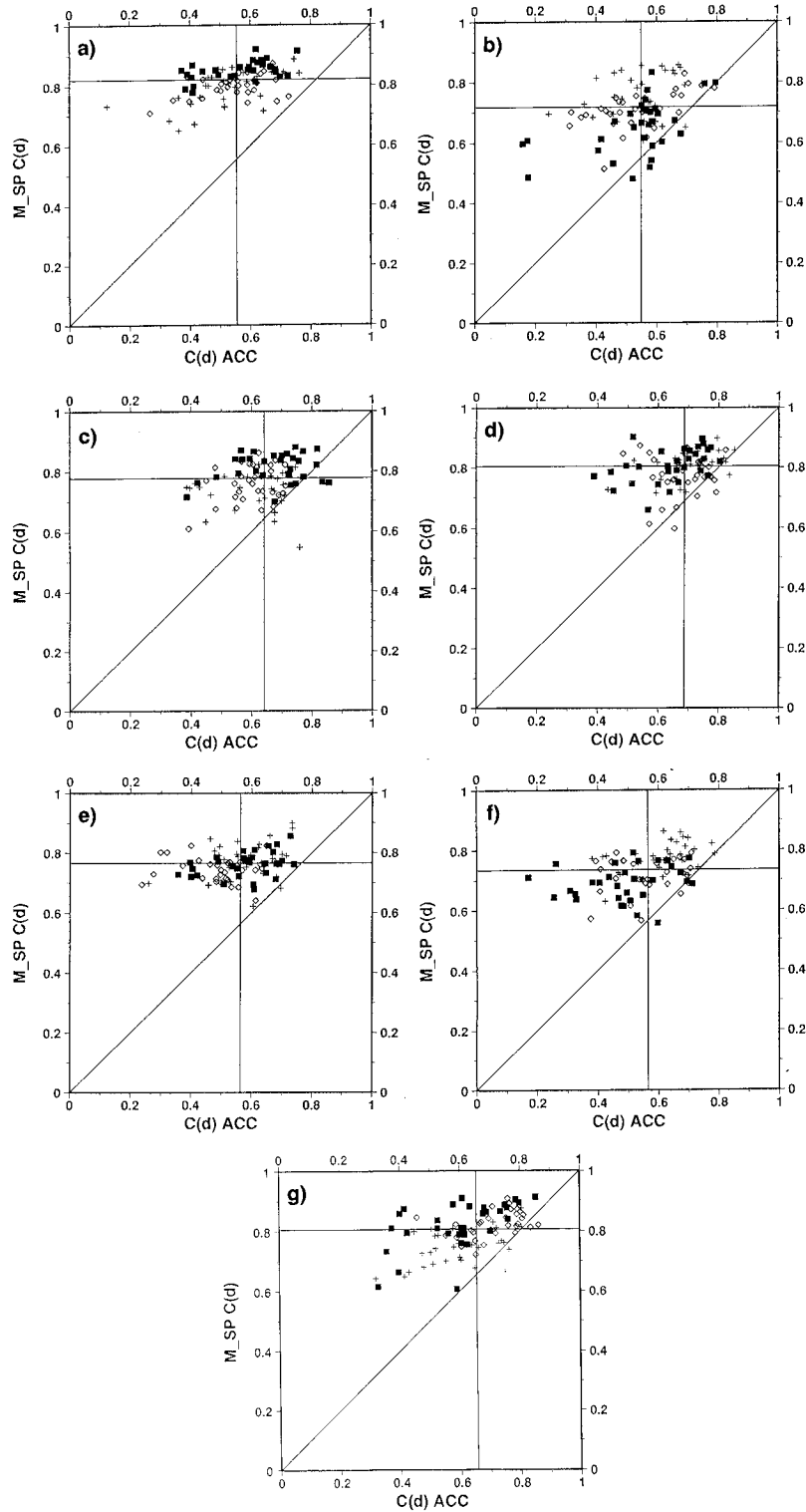


FIG. 14. Scatter diagrams of EPS spread versus control skill for the 92-day seasons starting (a) 1 May 1994, (b) 1 August 1994, (c) 1 November 1994, (d) 1 February 1995, (e) 1 May 1995, (f) 1 August 1995, and (g) 1 November 1995. Values refer to the 500-hPa geopotential height over the NH at forecast day 7. For each season plus, diamond, and square markers in the scatter diagram identify cases of the first, second, and third months of the analyzed season, respectively.

TABLE 5. Seasonal average of the percentage of analysis values lying outside the EPS forecast range.

	Northern Hemisphere		Europe	
	Day 5	Day 7	Day 5	Day 7
1 May 94–31 Jul 94	30	26	36	28
1 Aug 94–31 Oct 94	20	16	23	16
1 Nov 94–31 Jan 95	17	15	13	11
1 Feb 95–3 May 95	19	16	10	16
1 May 95–31 Jul 95	32	28	48	34
1 Aug 95–31 Oct 95	27	23	38	25
1 Nov 95–31 Jan 96	31	27	36	24

indices have been computed for all seasons (Tables 3 and 4). They range from zero (for the period 1 February 1995–3 May 1995 over the NH) to 0.46 (for the period 1 May 1994–31 July 1994 over Europe). Note that periods with very high correlation between spread and skill do not usually correspond to periods with low faulty indices. For example, the scatter diagram for the period 1 February 1995–3 May 1995 (Fig. 14d) is characterized by a very low faulty index, but 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.

c. Percentage of analysis values lying outside the EPS forecast range

Table 5 lists the percentage of analysis values lying outside the EPS forecast range for the NH and Europe at forecast days 5 and 7. All values are higher than the expected value of $100 \times (2/34)$ for the case of 34 categories defined by a 32+1 member ensemble (see section 3b). Considering, for example, the NH for the season started on 1 November 1994, 17% and 15% of the analysis values were outside the EPS forecast range at forecast day 5 and 7, respectively. 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 5).

TABLE 7. Correlation coefficients and faulty 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 the NH and Europe. Values in parentheses refer to the perfect ensemble.

	NH forecast day 7		Europe forecast day 7	
	cc	i_{faulty}	cc	i_{faulty}
1 May 94–31 Jul 94	0.50 (0.65)	0.11	0.55 (0.50)	0.33
1 Aug 94–31 Oct 94	0.27 (0.53)	0.30	0.04 (0.19)	0.46
1 Nov 94–31 Jan 95	0.24 (0.55)	0.45	0.35 (0.47)	0.10
1 Feb 95–3 May 95	0.32 (0.36)	0.27	0.51 (0.39)	0.16
1 May 95–31 Jul 95	0.42 (0.63)	0.31	0.44 (0.77)	0.53
1 Aug 95–31 Oct 95	0.42 (0.56)	0.06	0.31 (0.46)	0.45
1 Nov 95–31 Jan 96	0.50 (0.70)	0.45	0.54 (0.69)	0.10

d. Skill score of the ensemble mean and its correlation with ensemble spread around the ensemble mean

Figure 11 shows that the ensemble-mean field is more skillful than the control forecast after forecast day 5, especially during the cold seasons. Considering, for example, the season started on 1 November 1994 (Fig. 11c), the ensemble-mean skill curve (dashed) crosses the 0.6 threshold line almost 1 day after the control forecast (solid). Table 6 also reports the improvements, in terms of acc, of the ensemble mean with respect to the control forecast for the perfect ensemble. Indeed, Table 6 shows that the improvement of the EPS is smaller than its potential value (e.g., 0.10 instead of 0.14 at forecast day 10 over the NH, for the season started on 1 November 1994).

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 faulty indices of the corresponding contingency tables have been computed. Table 7 reports the values computed using acc spread and skill for the 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 faulty indices relative

TABLE 6. 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 the perfect ensemble.

	Northern Hemisphere		Europe	
	Day 7	Day 10	Day 7	Day 10
1 May 94–31 Jul 94	0.03 (0.05)	0.05 (0.10)	0.02 (0.04)	0.05 (0.10)
1 Aug 94–31 Oct 94	0.05 (0.07)	0.10 (0.17)	0.04 (0.08)	0.11 (0.21)
1 Nov 94–31 Jan 95	0.04 (0.06)	0.10 (0.14)	0.05 (0.05)	0.14 (0.14)
1 Feb 15–3 May 95	0.04 (0.06)	0.10 (0.14)	0.05 (0.06)	0.11 (0.14)
1 May 95–31 Jul 95	0.02 (0.07)	0.06 (0.16)	0.03 (0.05)	0.04 (0.15)
1 Aug 95–31 Oct 95	0.03 (0.08)	0.08 (0.18)	0.01 (0.09)	0.13 (0.19)
1 Nov 95–31 Jan 96	0.03 (0.06)	0.09 (0.14)	0.02 (0.06)	0.09 (0.17)

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. The ECMWF Ensemble Prediction System 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 with the same characteristics of the EPS (i.e., with 32 perturbed members with initial conditions generated using singular vectors) has been defined by substituting the verifying analysis with an ensemble member, and verified for seven 92-day seasons from 1 May 1994, the starting date of daily operational EPS.

In the third part of this paper we compared the performance of the ECMWF EPS with the skill of the perfect ensemble. Attention has been focused on three requirements: (i) the ensemble spread is comparable to the skill of the control forecast, (ii) small spread should indicate a skillful control forecast, and (iii) the verification is included within the ensemble forecasts range. 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 nonlinear 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 faulty index i_{faulty} has been defined and computed using ensemble spread/control skill contingency tables. Both temporal correlations and faulty 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, faulty indices indicate that there is some correspondence between small ensemble spread and high control skill.

Consistently with too small ensemble spread, the percentage of analysis values lying outside the EPS forecast range is still not negligible. 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 growing features of analysis error lying in the subspace 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 skillful, with acc differences at forecast day 7 for the 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 21 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. For example, Wilks and Hamill (1995) 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 very near future.

Acknowledgments. Appreciation goes to all ECMWF staff and consultants who contributed to the development of the ECMWF Integrated Forecasting System, on which the ECMWF Ensemble Prediction System is based. Acknowledgment 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, as is Adrian Simmons for carefully revising a first version of this paper. The author would also like to thank two anonymous reviewers and R. L. Elsberry for their very appropriate comments.

REFERENCES

- Buizza, R., 1994a: Sensitivity of optimal unstable structures. *Quart. J. Roy. Meteor. Soc.*, **120**, 429–451.

- , 1994b: Localization of optimal perturbations using a projection operator. *Quart. J. Roy. Meteor. Soc.*, **120**, 1647–1681.
- , 1995: Optimal perturbation time evolution and sensitivity of ensemble prediction to perturbation amplitude. *Quart. J. Roy. Meteor. Soc.*, **121**, 1705–1738.
- , and T. N. Palmer, 1995: The singular vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52**, 1434–1456.
- , J. Tribbia, F. Molteni, and T. N. Palmer, 1993: Computation of optimal structures for a numerical weather prediction model. *Tellus*, **45A**, 388–407.
- , R. Gelaro, F. Molteni, and T. N. Palmer, 1995: Predictability studies using high resolution singular vectors. ECMWF Research Department Tech. Memo. 219, 38 pp. [Available from ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom.]
- Courtier, P., C. Freyder, J. F. Geleyn, F. Rabier, and M. Rochas, 1991: The Arpege project at Météo France. *Proc. Numerical Methods in Atmospheric Models*, Shinfield Park, Reading, United Kingdom, ECMWF, 192–231.
- 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. Preprints, *10th Conf. on Atmospheric and Oceanic Waves and Stability*, Big Sky, MT, Amer. Meteor. Soc., 235–238.
- Fleming, R. J., 1971a: On stochastic dynamic prediction. Part I: The energetics of uncertainty and the question of closure. *Mon. Wea. Rev.*, **99**, 851–872.
- , 1971b: On stochastic dynamic prediction. Part II: Predictability and utility. *Mon. Wea. Rev.*, **99**, 927–938.
- Gleeson, T. A., 1970: Statistical-dynamical predictions. *J. Appl. Meteor.*, **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**, 3885–3894.
- Hollingsworth, A., 1980: An experiment in Monte Carlo forecasting procedure. *Proc. Workshop on Stochastic Dynamic Prediction*, Shinfield Park, Reading, United Kingdom, ECMWF, 65–86.
- 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). *Proc. Workshop on Modelling, Validation and Assimilation of Clouds*, Shinfield Park, Reading, United Kingdom, ECMWF, 277–294.
- Ledermann, W., 1984: *Statistics*. Vol. 6, *Handbook of Applicable Mathematics*, 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 Tech. Memo. 218, 34 pp. [Available from ECMWF, Shinfield Park, Reading RG2-9AX, United Kingdom.]
- Molteni, F., and T. N. Palmer, 1993: Predictability and finite-time instability of the northern winter circulation. *Quart. J. Roy. Meteor. Soc.*, **119**, 1088–1097.
- , R. Buizza, T. N. Palmer, and T. Petroligis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.
- Mureau, R., F. Molteni, and T. N. Palmer, 1993: Ensemble prediction using dynamically conditioned perturbations. *Quart. J. Roy. Meteor. Soc.*, **119**, 269–298.
- Palmer, T. N., F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia, 1993: Ensemble prediction. *Proc. Validation of Models over Europe*, Vol. 1, Shinfield Park, Reading, United Kingdom, ECMWF, 21–66.
- Rabier, F., E. Klinker, P. Courtier, and A. Hollingsworth, 1996: Sensitivity of forecast error to initial conditions. *Quart. J. Roy. Meteor. 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. *Meteor. Atmos. Phys.*, **40**, 28–60.
- , R. Mureau, and T. Petroligis, 1995: Error growth and predictability estimates for the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.*, **121**, 1739–1771.
- Strang, G., 1986: *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, 758 pp.
- Strauss, B., and A. Lanzinger, 1995: Validation of the ECMWF EPS. *Proc. Predictability*, Vol. 2, Shinfield Park, Reading, United Kingdom, ECMWF, 157–166.
- 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. *Quart. J. Roy. Meteor. Soc.*, **116**, 835–866.
- Tiedtke, M., 1993: Representation of clouds in large-scale models. *Mon. Wea. Rev.*, **121**, 3040–3060.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
- Tracton, M. S., and E. Kalnay, 1993: Operational ensemble prediction at the National Meteorological Center: Practical aspects. *Wea. Forecasting*, **8**, 379–398.
- Viterbo, P., and A. C. M. Beljaars, 1995: An improved land surface parameterization scheme in the ECMWF model and its validation. *J. Climate*, **8**, 2716–2748.
- Wilks, D. S., and T. M. Hamill, 1995: Potential economic value of ensemble-based surface weather forecasts. *Mon. Wea. Rev.*, **123**, 3656–3675.