

## Verification of the ECMWF Ensemble Prediction System Forecasts: A Study of Large-Scale Patterns

PIERO A. CHessa\* AND FRANÇOIS LALAURETTE

*European Centre for Medium Range Weather Forecasts, Reading, United Kingdom*

(Manuscript received 25 September 2000, in final form 20 March 2001)

### ABSTRACT

The aim of this work is the evaluation of some basic measures of skill of the Ensemble Prediction System used operationally at the European Centre for Medium-Range Weather Forecasts. Probabilistic forecasts for the occurrence of predefined flow patterns are analyzed and validated throughout the 10-day forecast range. The weather regimes used as fixed patterns are those defined by Vautard for the 700-hPa geopotential height and used during the field phase of the Fronts and Atlantic Storm-Track Experiment as a component in the planning of the intense observing periods.

The evaluation of the ensemble forecasts is performed for three cold seasons, from 1996 to 1999, using the Brier score and the related skill score. In this case the sample climate and a simple probabilistic forecast (poor man's ensemble) have been used as reference forecasts. The results show a positive skill with respect to the reference forecasts up to the 240-h time step.

### 1. Introduction

The Ensemble Prediction System (EPS) represents a practical way to predict the evolution of the probability distribution function of the atmospheric states (see, e.g., Murphy 1988; Molteni et al. 1996; Buizza and Palmer 1998). Since its operational implementation at the European Centre for Medium-Range Weather Forecasts (ECMWF) in 1992 much effort has been devoted to distilling concise information valuable for operational weather forecasts (e.g., Molteni et al. 1996; Buizza et al. 1999a; Atger 1999) and to validating the system by assessing the quality of its by-products (Palmer et al. 1992; Wilson 1995; Strauss and Lanzinger 1995; Molteni 1996; Buizza 1997; Buizza and Palmer 1998; Atger 1999; Richardson 2000).

Many present applications are focused on a direct use of the weather parameters and their postprocessing in terms of probabilistic quantities (Buizza et al. 1999b). However, information related to the upper-level fields is still valuable in operational environments. In fact, it provides useful information about the large-scale context in which the local weather forecast is embedded. This is of great importance in areas where the model

output for surface variables is strongly biased, such as mountainous regions. By making use of climate records of weather parameters, which are conditioned upon specific upper-level flow patterns, the direct model output can be improved and tailored to specific operational needs, notably in the case of extreme weather events; see Cacciamani et al. (1994), Eckert et al. (1996), Benzi et al. (1997), and Chessa et al. (1999) for illustrations of this approach.

The use of EPS to obtain information on possible alternative flow patterns has been addressed from the beginning of EPS's short history. In fact, the different scenarios are automatically clustered according to the relative distances of the EPS members over a given area and for a certain forecast range (Molteni et al. 1996). Verification of these products, however, is not straightforward as clusters on different days cannot be compared. This work will therefore try to assess the ability of EPS to provide useful probabilistic forecasts for a limited number of predefined and fixed flow patterns. Particular emphasis will be put on the late medium range when the overall skill of the system decreases and smaller scales of motion have lost their predictability.

Similar verification studies have been presented in such papers as Palmer et al. (1992) or Molteni et al. (1996), but the verification periods available at the time were quite short and the ECMWF EPS has since been substantially improved. Also some European meteorological services, such as the Swiss Meteorological Institute in Switzerland and the Royal Netherlands Meteorological Institute in Netherlands, are issuing flow-

\* Current affiliation: Servizio Agrometeorologico Regionale per la Sardegna, Sassari, Italy.

*Corresponding author address:* Dr. Piero A. Chessa, Servizio Agrometeorologico Regionale per la Sardegna, Viale Porto Torres 119, 07100 Sassari, Italy.  
E-mail: chessa@sar.sardegna.it

TABLE 1. Weights used to obtain control ensemble forecasts (see text for details). The time steps refer to the most recent of the forecasts employed.

Step	Weights		
	D-1	D-2	D-3
+72	0.75	0.15	0.10
+96	0.70	0.20	0.10
+120	0.60	0.25	0.15
+144	0.50	0.30	0.20
+168	0.40	0.35	0.25
+192	0.35	0.35	0.30

dependent operational forecasts based on the ECMWF EPS (see, e.g., Eckert et al. 1996), but so far verification has only focused on the weather parameter distributions obtained through them.

The regimes illustrated by Vautard (1990) have been chosen for the present study since they are well-established patterns of the northern Atlantic wintertime circulation, supported both by observational evidence and by theories of quasi-equilibrium dynamics (Vautard and Legras 1988). They have also been used to identify highly transient cyclones, strongly dependent on the underlying weather regimes properties, during the preparation of the Fronts and Atlantic Storm-Track Experiment (FASTEX; Ayrault et al. 1995) and during its field phase, when probabilities of occurrence of weather regimes based on the EPS were derived (Joly et al. 1999).

A drawback associated with the choice of these flow patterns is that their relevance for the forecast over continental Europe may be limited. In this respect, the present study should be regarded as a first step in evaluating the EPS performance in terms of flow patterns.

The paper is organized as follow. In section 2 a short outline of the ECMWF EPS and of a simple ensemble used for comparison is given. In section 3 the North Atlantic weather regimes are briefly described while section 4 deals with the production and the verification of probabilistic forecasts. A discussion of the results will come in section 5 and some conclusions are presented in section 6.

## 2. The ECMWF EPS and the poor man's ensemble

### a. The Ensemble Prediction System

The ECMWF EPS has been produced operationally since December 1992, when it was composed of 31 members with the model in the T63L19 configuration [spectral triangular truncation T63 and 19 vertical levels; Palmer et al. (1993); Molteni et al. (1996)]. The present version of the system has 51 members and the model version is in the T<sub>L</sub>159L40 configuration (spectral triangular truncation T159 with a linear grid). Moreover a stochastic scheme for the representation of uncertainties associated to the model physics (Buizza et al. 1999c) is currently used.

Schematically, each EPS forecast  $e_j$  is generated by integrating the perturbed model equations,

$$\frac{\partial e_j}{\partial t} = A(e_j, t) + P'_j(e_j, t), \quad (1)$$

starting from perturbed initial conditions

$$e_j(t = 0) = e_0(t = 0) + \delta e_j(t = 0), \quad (2)$$

where  $A$  and  $P'$  identify the contribution to the equation tendency of the nonparameterized and parameterized physical processes. For each grid point, identified by its latitude  $\lambda$ , longitude  $\phi$ , and vertical hybrid coordinate  $\sigma$ , the parameterized tendency  $P'$  is defined as

$$P'_j(e_j, t) = [1 + \langle r_j(\lambda, \phi, t) \rangle_{D,T}] P(e_j, t). \quad (3)$$

Here  $P$  represents the unperturbed diabatic tendency, and  $\langle \cdot \rangle_{D,T}$  indicates that the same random number  $r_j$  has been used for all grid points inside a  $D^\circ \times D^\circ$  box and over  $T$  time steps. The random numbers are currently sampled uniformly in the interval  $[-0.5, 0.5]$ . The same random number is used inside  $10^\circ$  boxes ( $D = 10$ ) and the set of random numbers is updated every 6 h ( $T = 6$ ) (note that random numbers do not vary with the vertical coordinate).

In Eq. (2),  $e_0(t = 0)$  is the operational analysis at  $t = 0$ , while  $\delta e_j$  denotes the  $j$ th initial perturbation. For each  $d$ , the initial perturbations are defined using the singular vectors growing in the forecast range between day  $d$  and day  $d + 2$  at initial time, and the singular vectors that had grown in the past between day  $d - 2$  and day  $d$  at final time:

$$\delta e_j(t = 0) = \sum_{i=1}^{25} [\alpha_{i,j} v_i^{d,d+2}(t = 0) + \beta_{i,j} v_i^{d-2,d}(t = 48h)], \quad (4)$$

where  $v_i^{d,d+2}(t = 0)$  is the  $i$ th singular vector growing between day  $d$  and  $d + 2$  at time  $t = 0$ . The coefficients  $\alpha_{i,j}$  and  $\beta_{i,j}$  set the local, initial amplitude of the ensemble perturbations while keeping them in the same unstable space. The amplitude ratio is defined by comparing the singular vectors with estimates of analysis errors (Barkmeijer et al. 1999; Molteni et al. 1996).

### b. The poor man's ensemble

The production of the EPS forecasts is computationally very expensive and verification generally involves comparisons of its performance with different and less expensive systems (see, e.g., Atger 1999 and references therein).

The simple ensemble forecast used in this work makes use of the unperturbed EPS member (called the control forecast) and what we will refer to as the control ensemble (CE). Each day's CE is constructed using the control forecasts of three consecutive days verifying that day. These forecasts are linearly combined with appropriate weights whose choice (see Table 1) has been guid-

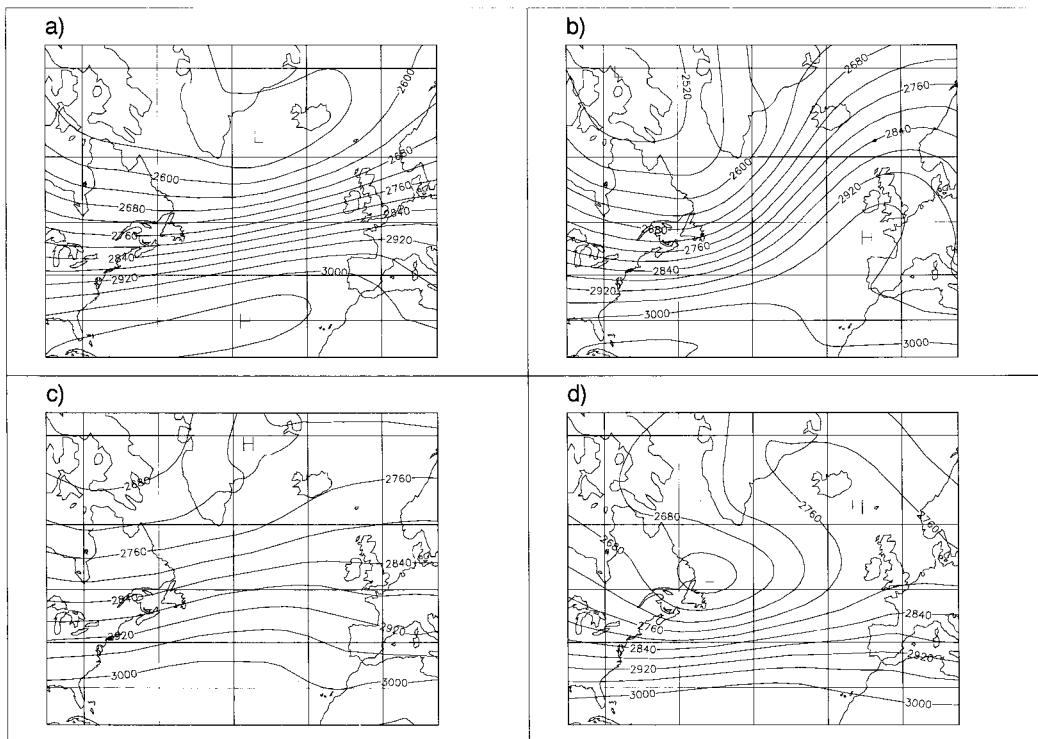


FIG. 1. The North Atlantic flow patterns found by Vautard (1990) and reported on in Ayrault et al. (1995). They have been named (a) zonal, (b) blocking, (c) Atlantic ridge, and (d) Greenland anticyclone. See text for a more detailed description.

ed by two main considerations: the roughly exponential structure of the model error growth with time (Simmons et al. 1995) and the fact that after a few days the model loses memory of the initial conditions and the main error sources are typically the model inaccuracies. The approach used in the definition of this simple ensemble is very close to operational forecasting practice, when consistency of consecutive forecasts is taken as an indication of the predictability for the coming atmospheric flow.

The weights given to each of the three members vary with the lead time. For the initial time steps the most recent run is the most heavily weighted; thereafter, its weight decreases roughly exponentially and those of the other two days increase smoothly so that for the latter lead times they are weighted nearly equally. It must be stressed, though, that the outcomes proved not to be very sensitive to the use of different weights, provided that the more recent forecast was given the highest weight. By construction, the CE forecast can be available only up to 192 h.

### 3. North Atlantic weather regimes

The regimes used in this work (Fig. 1) are described in Ayrault et al. (1995). They are similar to the quasi-stationary North Atlantic ( $21^{\circ}\text{N}$ – $90^{\circ}\text{W}$ ,  $72^{\circ}\text{N}$ – $15^{\circ}\text{E}$ ) patterns of 700-hPa geopotential height found by Vautard

(1990) and then used in several contexts (e.g., Robertson and Metz 1990; Joly et al. 1999).

Vautard (1990) defined these regimes using an algorithm developed by Vautard and Legras (1988). This algorithm is designed to find low-frequency quasi-stationary patterns. In this way the low-frequency patterns were considered stationary while small-scale features, which play a role in setting up and feeding the former (Illari and Marshall 1983; Shutts 1983; Lau and Holopainen 1984; Mullen 1987), were kept active.

For practical reasons, the regimes used in this work are those recomputed in Ayrault et al. (1995), but they exhibit negligible differences from those obtained by Vautard (1990). The first (Fig. 1a) is the zonal regime (ZO) and is characterized by a dipole in the anomaly field with the minimum east of Greenland. It is usually associated with rapid and unstable developments. The second pattern (Fig. 1b) represents a blocked flow (BL) with the anomaly's dipole stronger than ZO, but with reversed polarity. It generally brings easterly flows over Europe and prevents synoptic disturbance from the western Atlantic from reaching Europe. The third regime (Fig. 1c) has an Atlantic ridge (AR) as a main feature, while the fourth (Fig. 1d) is characterized by a maximum in the geopotential field around Greenland (GA), with the westerly flow located farther south with respect to the zonal regime.

TABLE 2. EPS forecast probabilities for 23 Jan 1999.

Step	Probability			
	ZO	BL	AR	GA
+72	0.98	0.02	0	0
+96	0.43	0.41	0.16	0
+120	0.14	0.60	0.18	0.08
+144	0.12	0.58	0.18	0.12
+168	0.57	0.43	0	0
+192	0.08	0.51	0.02	0.39
+216	0.38	0.43	0.14	0.24
+240	0.45	0.38	0.12	0.06

#### 4. Probabilistic forecast and verification

##### a. Probabilistic forecast

The data used in this study come from about 400 operational EPS runs, collected from 1 January 1997 to the end of April 1999. Forecasts were examined for extended winter seasons (Oct–Apr) as defined in Ayraut et al. (1995) and Vautard (1990). For each day and each forecast time step every perturbed member of the EPS and the control forecast (CF) was compared with the four regimes, using as a similarity measure the correlation coefficient for the 700-hPa geopotential height anomalies, that is,

$$\rho_{i,k}(d, t) = \frac{\overline{f_i(d, t)r_k} - \overline{f_i(d, t)}\overline{r_k}}{\sigma_{f_i(d, t)}\sigma_{r_k}}, \quad (5)$$

where  $\rho_{i,k}(d, t)$  represents the correlation coefficient between the EPS member  $i$  for the day  $d$  and time step  $t$  [ $f_i(d, t)$ ] and the regime  $k$  ( $r_k$ ), while the overbars refer to an average on all grid points. Then the EPS members are associated to the closest regime and at the end of the process the empirical frequencies for each regime are used to derive the corresponding probabilities.

To accomplish this task two basic assumptions are made: 1) the ensemble perfectly spans the space of possible outcomes; probabilities are assumed to be zero for events either below the minimum or above the maximum of the ensemble; and 2) between these two extremes each member is considered equiprobable.

Probabilities therefore are

$$p(x > \alpha) = \frac{1}{N} \sum_{i=1}^N H(\alpha_i - \alpha), \quad (6)$$

where  $p(x > \alpha)$  is the forecast probability that the verification will lie beyond  $\alpha$ , while  $H(\alpha_i - \alpha)$  is 1 when EPS member  $i$  takes a value beyond  $\alpha$ , 0 otherwise (Heaviside function). As an example the forecast issued on 23 January 1999 is reported in Table 2 for the forecast times ranging from 72 to 240 h with 24-h intervals. In this case the EPS was forecasting a clear transition from zonal to blocking flow.

As far as the CF is concerned the forecast probability could obviously assume only two values for each possible flow regime, 1 or 0. Therefore, for each forecast

TABLE 3. Probabilistic forecast for 23 Jan 1999 obtained by means of the control ensemble.

Step	Probability			
	ZO	BL	AR	GA
+72	1	0	0	0
+96	0.30	0.70	0	0
+120	0.15	0.85	0	0
+144	0	1	0	0
+168	0	0.60	0	0.40
+192	0.65	0	0	0.35

the closest regime has a probability of occurrence equal to 1 and the other three regimes equal to zero.

Finally, for CE, the previously described weights were used in order to set probabilities using the three forecasts. Thus if  $C_k(d, t)$  is the probability (either 0 or 1) obtained by means of the control forecast, the CE estimate will be

$$P_k(d, t) = w_{(0)}C_k(d, t) + w_{(-1)}C_k(d - 1, t + 24) + w_{(-2)}C_k(d - 2, t + 48). \quad (7)$$

As an example in Table 3 the CE outcomes for 23 January 1999 are shown.

##### b. Verification scores

The performance of the different forecasts used in this work have been evaluated by means of the Brier score (BS) and the Brier skill score; see, for example, Brier (1950) and Wilks (1995). The BS is defined as

$$BS^k(t) = \frac{1}{N} \sum_{i=1}^N [p_i^k(t) - o_i^k]^2. \quad (8)$$

In this equation  $N$  represents the number of days considered,  $P_i^k(t)$  is the forecast probability for the time step  $t$  and the regime  $k$ , and  $o_i^k$  represents the occurrence or nonoccurrence of the event being forecast. This quantity is equal to 1 in the first case and to 0 in the second. The Brier score is the mean-squared error of a probability forecast and it is negatively oriented: the lower the score the better the forecast. It is also bound to assume values from 0 to 1 for each regime. Other than for each regime, the BS was also calculated for the four classes considered together. In this case, where nonordinal categories are considered, it is simply the sum of the single BSs and must be used carefully (Stanski et al. 1989).

The Brier skill score (BSS) is (see, e.g., Wilks 1995) defined as

$$BSS^k(t) = \frac{BS_{\text{ref}} - BS_k(t)}{BS_{\text{ref}}}. \quad (9)$$

In this case higher values of BSS indicate better performances of the system checked with respect to the one (among many possible) used as a reference.

The properties of the BS can be more easily under-

TABLE 4. Climatological frequencies of each regime as obtained from ECMWF analysis for 1984–94 (see Ayrault et al. 1995).

Regime	Frequency
Zonal	0.34
Blocking	0.22
Atlantic ridge	0.27
Greenland	0.17

stood using the decomposition proposed by Murphy (1973):

$$\text{BS}(t) = \frac{1}{N} \sum_{i=1}^I N_i [p_i(t) - \bar{o}_i]^2 - \frac{1}{N} \sum_{i=1}^I N_i (\bar{o}_i - \bar{o})^2 + \bar{o}(1 - \bar{o}), \quad (10)$$

where  $N_i$  is the number of times  $p_i(t)$  is used,  $N = \sum_{i=1}^I N_i$ , and  $I$  is the number of probability intervals considered. In addition,  $o_i = 1/N_i \sum_{j \in N_i} o_j$  is the *conditional* frequency of occurrence and  $\bar{o} = 1/N \sum_{j=1}^N o_j$  is the sample climatology. Conditional frequency refers to a given range of forecast probabilities. The first term in the equation above is called reliability and a perfect score occurs when it equals 0, that is, when the conditional frequency is equal to the corresponding forecast probability. It accounts for the conditional bias of the forecast distribution. The second term is called resolution and provides information about the forecast capability to resolve different forecast periods characterized by different relative frequencies of the event. The last part in the decomposition is called uncertainty and depends only on the sample climatology. For this reason it cannot be influenced by the forecast technique and its effects can only be compensated by a suitable combination of reliability and resolution. Its maximum value is 0.25 and it is obtained when the event has a frequency equal to 0.5. All the different attributes just described can be obtained by means of the so-called reliability or attribute diagrams [a complete description can be found for instance in Wilks (1995)] that will be used later on in this work.

## 5. Results and discussion

The BSs were calculated for EPS, CE, CF, long-term climate (CLIM), and what was defined as perfect ensemble in Molteni et al. (1996). Basically it represents the score that can be achieved by an ensemble whose forecast probabilities are an exact estimate of the frequency of occurrence of each regime. When computing Brier skill scores, the sample climatology uses the frequencies calculated in Ayrault et al. (1995) and reported in Table 4.

The BSSs were computed for EPS, CE, and CF using different reference forecasts and in particular EPS has been compared to CE, CF, and CLIM; CE has been compared to CF and CLIM; and CF has been compared to CLIM.

In Fig. 2 the relative frequencies of the various forecasts and climatology are reported for each regime. The frequencies of occurrence of ZO and AR regimes in the period studied appear to be quite different from those obtained for the period 1984–94 by Ayrault et al. (1995). However the forecast frequencies are comparable to the actual occurrences. Although differences are small, there is a general underforecast of ZO and BL regimes and a slight over forecast of AR flow.

In Fig. 3, the EPS Brier scores for the different regimes are reported. In the case of CF and CE the results, not shown, are similar. The discrepancies between the flow patterns are basically due to the different sample population, a characteristic to which BS is very sensitive (e.g., Stanski et al. 1989). Both EPS and CE have low BSs up to the late medium range. Not as low are the BSs for CF, especially after 120 h. In Fig. 4 a comparison of BSs for all systems, including CLIM and perfect ensemble (PE), is presented but considering all regimes together. Note that the ordinate axes scale is different from the previous plot because all the categories have been summed (see, e.g., Stanski et al. 1989) and the CE forecast is, for obvious reasons, limited to 192 h. It is worth noting how close the EPS Brier score is to the PE one, even for the last time step. This denotes an improvement with respect to the results presented in Molteni et al. (1996, see for instance their Fig. 10). Other notable features include the high values of the Brier score for CF, which actually exceed those for CLIM from 168 h.

Based on the EPS BSSs with respect to CLIM for the various regimes, EPS is skillful in all cases and for all lead times (Fig. 5). Figure 5 also shows that even a probabilistic forecast for 240 h is more accurate than climatology when the EPS is used. Another noticeable feature is how the skill is flow dependent: for instance blocking situations seem more difficult to forecast many days in advance than the zonal ones.

The EPS performs better than both CF and CE when the regimes are considered (Fig. 6). In the first case the most interesting property is the increase of the EPS skill with respect to the forecast time. The same results, not shown here, have been obtained also using the *high-resolution* operational model (spectral triangular truncation T319 and 50 vertical levels) and this seems to confirm that after 120–144 h a probabilistic forecast is a more appropriate and effective approach than a *deterministic* one. Similar results to those for CF can be seen when CE is used as a reference forecast but in this case the EPS skill is less evident.

Figure 7 shows that CE has positive skills with respect both to CLIM and CF at least for the time range considered. By contrast, after 120 h CF turns out to be less skillful than forecasts made using the climatological frequencies.

So far nothing has been said about different characteristics of the forecast such as reliability or resolution. In order to give an example in Figs. 8 and 9 the

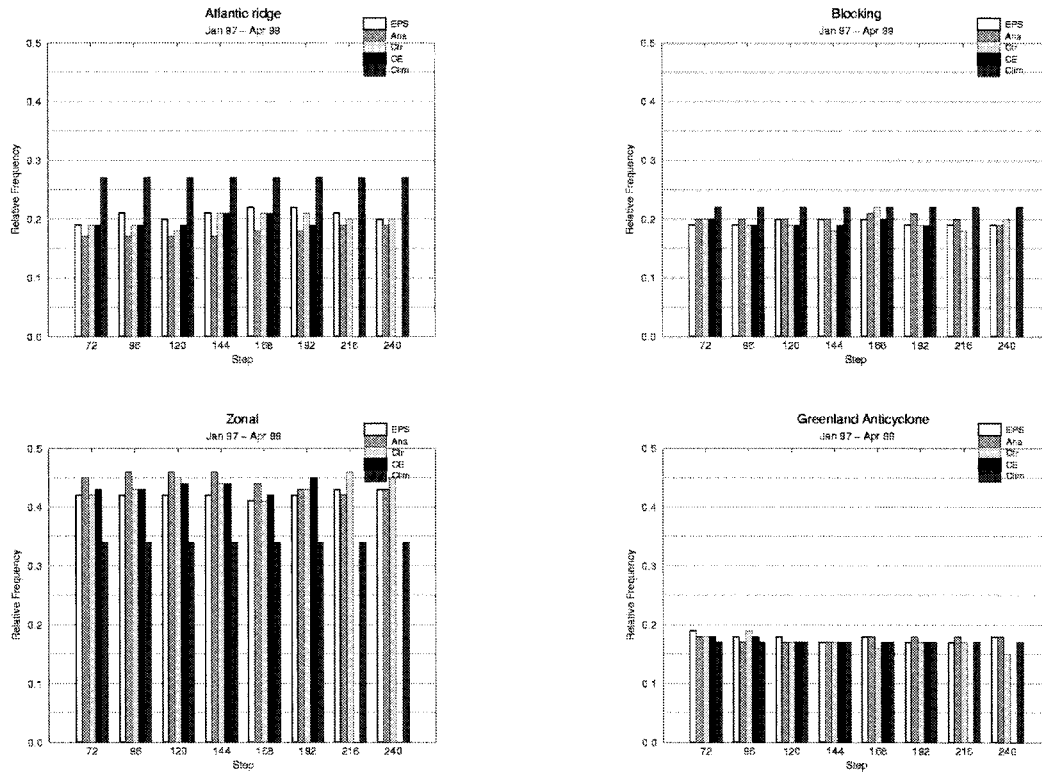


FIG. 2. Relative frequencies of all forecasts used in the analysis for the period considered in the study and the climatological frequencies (long-term climatology) as reported in Ayrault et al. (1995). Note that CE is not available after 192 h.

Brier score decomposition is shown for both EPS and CF in the case of zonal regime. The EPS reveals a good reliability and a fairly good resolution, while for the control forecast the cooperative effect of the two quantities is not enough to compensate the uncertainty after 144 h. As a matter of fact while the EPS resolution is only marginally better, most of the difference is in the reliability, which implies a better calibration of the EPS.

As already mentioned, a more complete picture of the

statistical problem behind this kind of verifications can be obtained using attributes diagrams (Wilks 1995; Stan-ski et al. 1989). In Fig. 10 the reliability diagram for the EPS forecast is presented. For clarity the regimes are all plotted together and only forecasts for 72, 120, 168, and 216 h are considered. Also the so called no-resolution line for  $y = \bar{\sigma}$  (for clarity the no-resolution

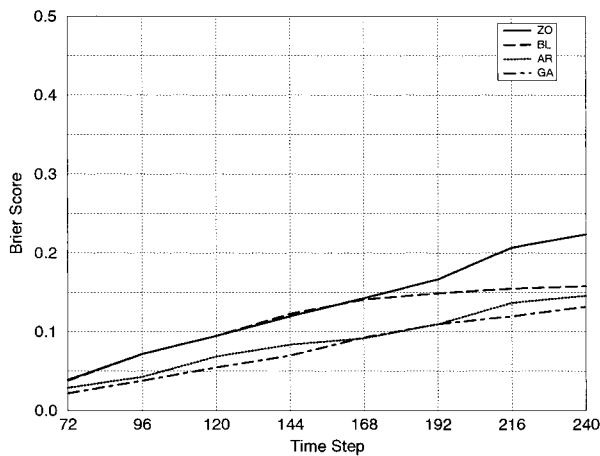


FIG. 3. EPS Brier score for each weather regime.

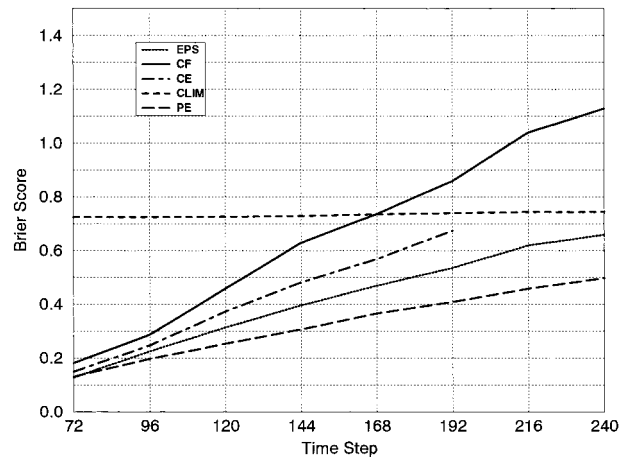


FIG. 4. Brier scores for EPS, control forecast (CF), control ensemble (CE), long-term climatology (CLIM), and perfect ensemble (PE). In this case the Brier score is calculated for all regimes together.

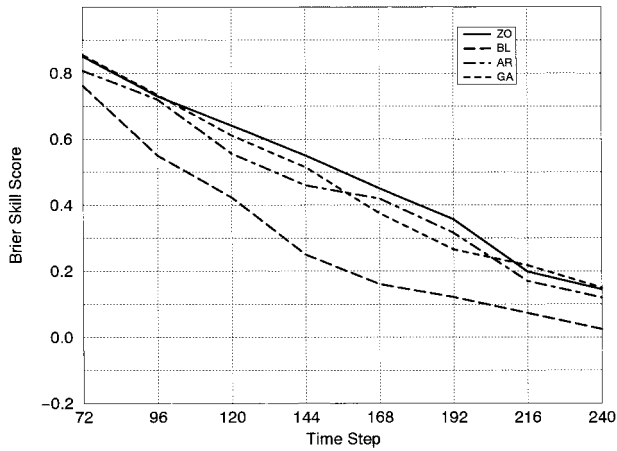


FIG. 5. EPS Brier skill score for each weather regime with CLIM as reference forecast.

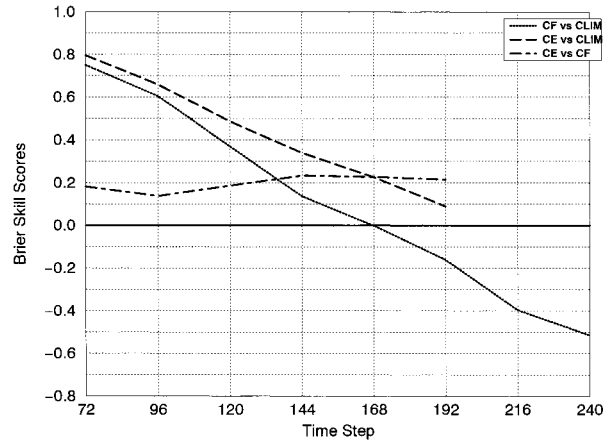


FIG. 7. Same as Fig. 6 but for control and control ensemble forecasts.

line  $x = \bar{o}$  has been omitted) and the no-skill line, half-way between the no-resolution and the perfect reliability lines, are shown in the diagram to facilitate the interpretation of the scores value. Also to this end is a smaller plot, which accompanies the main diagram and takes into account the forecast sharpness, that is, the system ability to issue forecasts generally far apart from climatological frequencies and possibly close to 1 or 0. Only the case for the 72-h step is reported as an example since the plots are similar for all lead times. The bar plot shows that EPS has some sharpness but it is still insufficient for high probabilities. The overall picture is that the EPS is fairly reliable and has a good resolution for probabilities below 0.5, although with a tendency to overforecasting that increases with the time step. For probabilities greater than 0.5 there is a progressive lack of resolution and then an overall lack of forecast skill (typical of rare events) as can also be seen writing by the BS as (resolution - reliability)/uncertainty (Wilks 1995, p. 265).

6. Conclusions

This work is a preliminary evaluation of the skill of the ECWMF Ensemble Prediction System for probabilistic forecasts of large-scale or synoptic flow patterns.

The main question addressed in this paper is: What is the limit of forecasting range for EPS to provide a skillful forecast relative to simple single forecasts as the control run or a simple ensemble such as the one presented here? As only large scales can be expected to be predictable in the late medium range, a simple partition of the outcomes in a limited number of weather regimes is appropriate.

The Ensemble Prediction System performance has been analyzed using the Brier score (and the related skill score) applied to a probabilistic forecast of four regimes defined in Ayrault et al. (1995) for the 700-hPa geopotential height. The area used is the target area of FASTEX (Joly et al. 1999), roughly the North Atlantic Ocean. More than 400 daily forecasts have been ana-

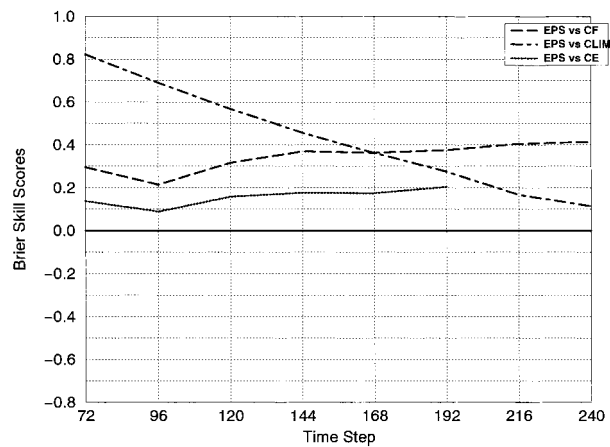


FIG. 6. EPS Brier skill scores with respect to different reference forecasts (indicated in the legend) when the regimes are considered all together.

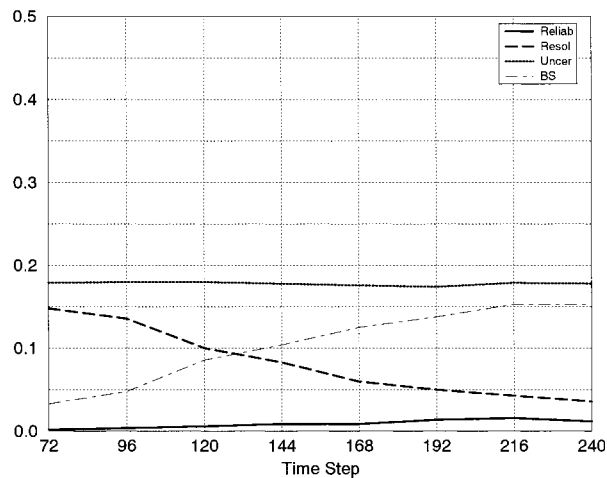


FIG. 8. EPS BS decomposition for the probabilistic forecasts of the zonal regime.

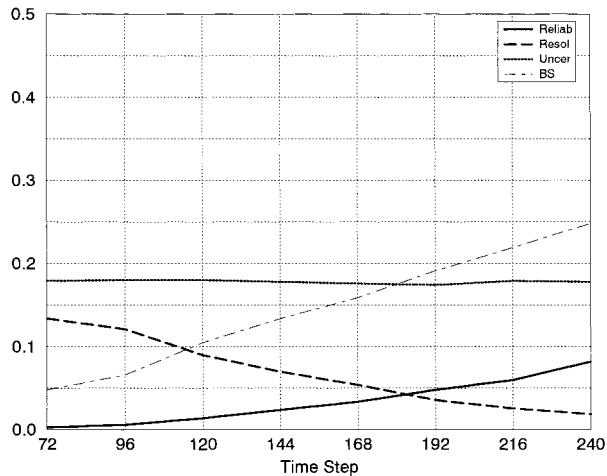


FIG. 9. Same as in Fig. 8 but for the control forecast.

lyzed for forecast lead times from 72 to 240 h. Other than the EPS, the control forecast and a simple ensemble system, here called control ensemble, have been evaluated.

The results are encouraging and show a clear superiority of the EPS with respect to all other systems even in the latter part of the forecast time range. The EPS's positive skill suggests that there might be room for improvements, even for forecasts longer than 240 h. The control ensemble shows positive skill with respect to the climatology as well as to the control, which was not able to supply useful information after 144 h.

A limitation in this study was the use of the 700-hPa geopotential height. The fact that data for this level are not exchanged in the Global Telecommunication System precludes the possibility of comparing ensemble forecasts obtained using different models too, as done in Atger (1999) and Ziehmann (2000).

Apart from providing a basic evaluation of the forecast skill for the large scale, the use of a probabilistic forecast for predefined flow patterns can prove to be a valuable way to complement the forecast of weather parameters. In fact, it is a common practice in meteorological services to find, both on a subjective and an objective basis, a statistical link between recurrent large-scale or synoptic patterns and surface parameters distributions. Hence, the possibility of classifying different flow patterns and assigning to them appropriate probabilities can provide useful information to forecasters.

**Acknowledgments.** The authors would like to thank Roberto Buizza and Horst Boettger for their useful comments to an early version of this manuscript.

#### REFERENCES

Atger, F., 1999: The skill of ensemble prediction systems. *Mon. Wea. Rev.*, **127**, 1941–1953.  
 Ayrault, F., F. Lalaurette, A. Joly, and C. Loo, 1995: North Atlantic ultra high frequency variability. *Tellus*, **47A**, 971–996.

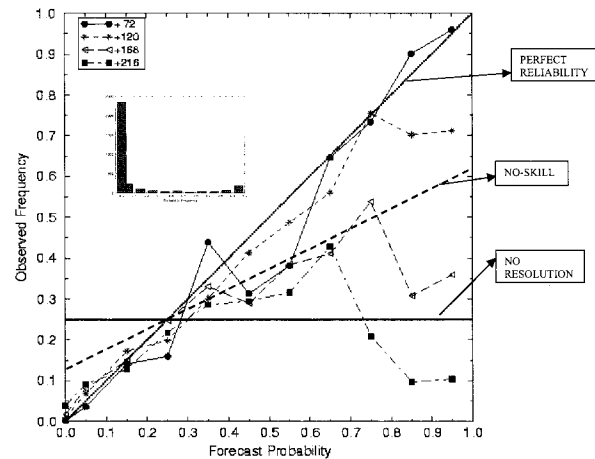


FIG. 10. Reliability diagram and sharpness (small bar plot) for  $T + 72$ ,  $T + 120$ ,  $T + 168$ , and  $T + 216$  EPS forecasts when regimes are considered all together.

Barkmeijer, J., R. Buizza, and T. N. Palmer, 1999: 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2333–2351.  
 Benzi, R., R. Deidda, and M. Marrocu, 1997: Characterization of temperature and precipitation fields over Sardinia with principal component analysis and singular spectrum analysis. *J. Climatol.*, **17**, 1231–1262.  
 Brier, G. W., 1950: Verification of forecasts expressed in term of probability. *Mon. Wea. Rev.*, **78**, 1–3.  
 Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.*, **125**, 99–119.  
 —, and T. N. Palmer, 1998: Impact of the ensemble size on ensemble prediction. *Mon. Wea. Rev.*, **126**, 2503–2518.  
 —, J. Barkmeijer, T. N. Palmer, and D. Richardson, 1999a: Current status and future developments of the ECMWF Ensemble Prediction System. *Meteor. Appl.*, **6**, 1–14.  
 —, A. Hollingsworth, F. Lalaurette, and A. Ghelli, 1999b: Probabilistic predictions of precipitation using the ECMWF Ensemble Prediction System. *Wea. Forecasting*, **14**, 168–189.  
 —, M. Miller, and T. N. Palmer, 1999c: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908.  
 Cacciamani, C., S. Nanni, and S. Tibaldi, 1994: Mesoclimatology of winter temperature and precipitation in the Po Valley of northern Italy. *J. Climatol.*, **14**, 777–814.  
 Chessa, P. A., D. Cesari, and A. M. Delitala, 1999: Mesoscale precipitation and temperature regimes in Sardinia (Italy) and the related synoptic circulations. *Theor. Appl. Climatol.*, **63**, 195–222.  
 Eckert, P., D. Cattani, and J. Ambhul, 1996: Classification of ensemble forecasts by means of an artificial neural network. *Meteor. Appl.*, **3**, 169–178.  
 Illari, L., and C. J. Marshall, 1983: On the interpretation of eddy fluxes during a blocking episode. *J. Atmos. Sci.*, **40**, 2232–2242.  
 Joly, A., 1995: The stability of steady fronts and the adjoint method: Nonmodal frontal waves. *J. Atmos. Sci.*, **52**, 3082–3108.  
 —, and Coauthors, 1999: Overview of the field phase of the Fronts and Atlantic Storm-Track Experiment (FASTEX) project. *Quart. J. Roy. Meteor. Soc.*, **125**, 1–34.  
 Lau, N. C., and E. O. Holopainen, 1984: Transient eddy forcing of the time-mean flow as identified by geopotential tendencies. *J. Atmos. Sci.*, **41**, 313–328.  
 Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliaigis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.



- Mullen, S. L., 1987: Transient eddy forcing of blocking flows. *J. Atmos. Sci.*, **44**, 3–22.
- Murphy, A., 1973: A new vector partition of the probability score. *J. Appl. Meteor.*, **12**, 595–600.
- Murphy, J. M., 1988: The impact of ensemble forecasts on predictability. *Quart. J. Roy. Meteor. Soc.*, **114**, 463–493.
- Palmer, T. N., F. Molteni, R. Mureau, R. Buizza, P. Chapelet, and J. Tribbia, 1992: Ensemble prediction. *Proc. ECMWF Seminar on Validation of Models over Europe*; Vol. 1, Reading, United Kingdom, European Centre for Medium-Range Weather Forecasts, 21–66.
- Richardson, D. S., 2000: Skill and economic value of the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **126**, 649–668.
- Robertson, A. W., and W. Metz, 1990: Transient-eddy feedbacks derived from linear theory and observation. *J. Atmos. Sci.*, **47**, 2743–2764.
- Shutts, H. G. J., 1983: The propagation of eddies in diffluent jet streams: Eddy vorticity forcing of “blocking” flow fields. *Quart. J. Roy. Meteor. Soc.*, **109**, 737–761.
- , R. Mureau, and T. Petroligis, 1995: Error growth and estimates of predictability from ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.*, **121**, 1739–1771.
- Stanski, H. R., L. J. Wilson, and W. R. Burrows, 1989: Survey of common verification methods in meteorology. WMO/WWW Tech. Rep. **8**, 114 pp.
- Strang, G., 1986: *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, 758 pp.
- Strauss, G., and A. Lanzinger, 1995: Validation of the ECMWF Ensemble Prediction System. *Proc. ECMWF Seminar on Predictability*, Vol. I, Reading, United Kingdom, European Centre for Medium-Range Weather Forecasts, 157–166.
- Vautard, R., 1990: Multiple weather regimes over the North Atlantic: Analysis of precursors and successors. *Mon. Wea. Rev.*, **118**, 2056–2081.
- , and B. Legras, 1988: On the source of midlatitude low-frequency variability. Part II: Nonlinear equilibration of weather regimes. *J. Atmos. Sci.*, **45**, 2845–2867.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.
- Wilson, L. J., 1995: Verification of weather element forecasts from an ensemble prediction system. *Proc. Fifth Workshop on Meteorological Operational Systems*, Reading, United Kingdom, European Centre for Medium-Range Weather Forecasts, 114–126.
- Ziehmann, C., 2000: Comparison of a single-model EPS with a multimodel ensemble consisting of a few operational models. *Tellus*, **52A**, 280–299.