The new 80-km High-Resolution ECMWF EPS

Small errors in initial conditions (initial uncertainties) and the approximate representation of atmospheric processes in numerical models (model uncertainties) are the two main sources of forecast error. These two sources of uncertainty limit the skill of single, deterministic forecasts in a highly flow-dependent way, with days of high quality forecasts followed by days of poor quality forecasts.

A complete description of the weather prediction problem can be stated in terms of the time evolution of an appropriate probability density function (PDF) in the atmosphere's phase space. Ensemble prediction based on a sampling of this PDF by a finite number of deterministic integrations designed to represent both initial and model uncertainties appears to be the only feasible method to predict the PDF beyond the range of linear error growth (Fig. 1).

Routine real-time execution of the ECMWF EPS started in December 1992 with a 33-member T63L19 configuration (spectral triangular truncation T63 and 19 vertical levels, *Palmer et al.* 1993, *Molteni et al.* 1996). This first version of the EPS simulated initial uncertainties due to errors growing in the forecast time (*Buizza & Palmer* 1995) but did not simulate model uncertainty. A major upgrade to a 51-member T_L159L31 system (spectral triangular truncation T159 with linear grid) took place in 1996 (*Buizza et al.* 1998). Two further important modifications were introduced in 1998. In March 1998, initial uncertainties due to

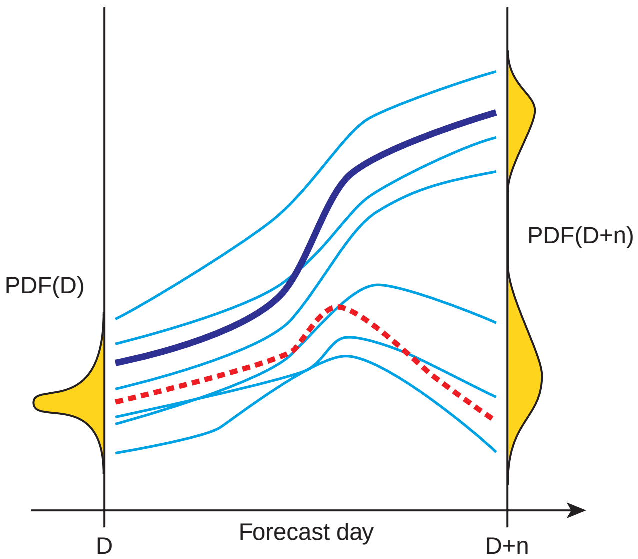

Figure 1 Schematic of ensemble prediction. The initial probability PDF(D) represents the initial uncertainties. From the best estimate of the initial state a single deterministic forecast (bold blue solid curve) is performed. This single deterministic forecast fails to predict correctly the future state (red dotted curve). An ensemble of perturbed forecasts (thin blue solid curves) starting from perturbed initial conditions designed to sample the initial uncertainties can be used to estimate the probability PDF(D+n) at the future time D+n.

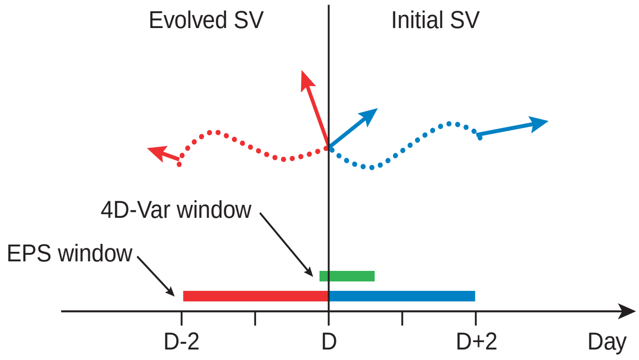

Figure 2 The initial perturbations of the EPS starting at day D are defined to sample perturbations growing in the past, between days (D–2) and D, and in the future, between days D and (D+2). The first set of perturbations (red arrows), are defined by using singular vectors evolved from day (D–2) to day D, while the second set of perturbations (blue arrows) are defined by using initial singular vectors. With this choice, the time-interval spanned by the two sets of singular vectors cover the time-interval spanned by the 4D-Var data-assimilation procedure (green). NH – summer (30 cases) NH – winter (57 cases)

perturbations that had grown during the data assimilation system (evolved singular vectors, *Barkmeijer et al.* (1999)) were added to the initial uncertainties due to errors growing rapidly in the forecast time (initial singular vectors). Figure 2 is a schematic of the way the ensemble initial perturbations have been defined since March 1998. The initial perturbations of the ensemble starting at day D are defined by combining the singular vectors growing between day D-2 and day D, evolved at day D, and the singular vectors growing between day D and D+2, at initial time. In October 1998, a scheme to simulate model uncertainties due to random model error in the parametrized physical processes was introduced (*Buizza et al.* 1999).

The changes introduced in 1996 and 1998 improved the performance of the EPS (*Buizza et al.* 2000). Compared to the initial 33*T63L19 system, the upgraded 51*T_L159L31 system had a better level of spread (both in terms of anomaly correlation and root-mean-square distance), a more skilful ensemble-mean, a higher chance of including the verification analysis inside the forecast range, and more accurate probabilistic predictions (as measured by different skill measures such as the Brier skill score, the area under the relative-operating-characteristic curve, the ranked-probability skill score and the potential economic value.

Long-term goals for the EPS.

The Strategic Plan 1999-2008 (adopted by the ECMWF Council in June 1999, ECMWF (1999)) states the long-term goals for deterministic forecasting and ensemble prediction:

Deterministic Forecasting: A robust measure of the performance of the Centre's deterministic forecasts is the forecast range for which the anomaly correlation of the

Northern Hemisphere 500hPa height field exceeds the 60% level. A realistic 10-year target is to extend this forecast range by one day.

Ensemble Prediction: Brier-skill-scores for probabilistic forecasts are the appropriate measure of performance of the EPS. Here the historical record is too short to be helpful in formulating targets. However, in line with the first goal, a reasonable target is a gain of one day at D+6 in the Brier-skill-score for probabilistic forecasts of moderate 850hPa temperature anomalies (4K or larger) in Europe.

This strategic plan defines the goals that any modifications of the EPS should aim to achieve in the 10-year period from 1999. It is the intention of this work to document whether a recent increase of the EPS resolution had made achievement of the long-term strategic targets closer.

The new 80-km High-resolution EPS

Following extensive experimentation, the resolution of the ECMWF analysis and forecasting system was increased on 21 November 2000:

- Deterministic model and analysis: from T_L319L60 (60 km grid-point spacing) to T_L511L60 (40 km grid-point spacing);
- EPS: from T_L159 (120km grid-point spacing) to T_L255 (80km grid-point spacing).

Thus, compared to the EPS the new 80-km High Resolution EPS (HEPS) is run with a finer horizontal resolution and starting from a higher-resolution analysis.

This report documents the impact of the increased resolution on the accuracy of the ECMWF ensemble system. To limit the number of figures and tables, and following the strategic goals, only 850 hPa temperature fields are considered. For each ensemble configuration, the following measures of ensemble performance (hereafter called 'scores') have been considered:

- Skill of the EPS control, measured in terms of anomaly-correlation coefficient (ACC);
- Ensemble spread with respect to the control forecast, measured in terms of ACC;
- Skill of the ensemble-mean, measured in terms of ACC;
- Brier skill score (BSS) of EPS probabilistic predictions of positive and negative anomalies with amplitudes larger than the seasonal variability (defined as the standard deviation of the analysed fields).

The EPS and the HEPS configurations have been compared for 87 cases covering two periods: summer 1999 (30 cases, from 2 to 30 August) and winter 1999-2000 (57 cases, from 26 November to 27 December 1999 and from 22 January to 15 February 2000). All scores have been computed using forecast and analysed fields defined on a regular latitude-longitude grid, with a spacing of 2.5°. Scores have been computed for two regions, the Northern Hemisphere (NH) and Europe. These are shown mostly for NH, mainly for reason of space, but also because differences between the EPS and HEPS score distributions are less statistically significant for Europe.

The verifying analysis is defined by the operational T_L319L60 analysis, and not by the T_L511L60 analysis from which the HEPS starts. This choice has a negligible effect in the forecast range after forecast day 2, but it has a small but detectable impact for earlier forecast ranges when it slightly favours the EPS.

For each area and for each ensemble configuration, mean scores computed for the summer and the winter periods are shown (confidence intervals have been computed, but figures become unreadable if they are added to the average values). The degree of similarity between the distribution of scores of the two ensemble configurations is measured by the Rank-Mann-Wilcoxon (RMW) test. The RMW test estimates the probability that the distribution of scores of the EPS and the HEPS configurations are statistically distinguishable: low/high RMW values indicate that there is a small/large probability that the two distributions are sub-samples of the same overall distribution. For any score, the HEPS and EPS distributions are considered statistically different if RMW≤10, that is if there is a 10% or lower probability that the two distributions of ACC scores comes from the same overall distribution.

Target skill and Relative Improvement index

To highlight the level of skill gained, compared with the long-term strategic target of a one-day improvement, the average HEPS scores are contrasted with the average EPS scores and with the average EPS scores shifted by one day (EPS(d-1)), i.e. with an EPS system characterised by a one-day gain in skill. If the skill of the HEPS lies between the skills of the EPS and the EPS(d-1) it means that HEPS is better than EPS; if the HEPS skill is equal or above the skill of EPS(d-1) it means that HEPS shows a gain in skill of one day or more.

To be able to compare the impact of the change in resolution on different scores, for any score measure, SC, the Relative Improvement (RI) index $RI(SC)$

$$RI(SC) = \frac{SC[HEPS] - SC[EPS]}{SC[EPS(d{-}1) - SC[EPS]}$$

has been computed. The RI index is a normalised measured of the gain in skill obtained by configuration HEPS, with the normalisation coefficient defined by the long-term strategic target of a one-day gain. A 100% positive RI indicates an improvement equivalent to one day. Since, by construction, the EPS has been designed for forecast lengths of two days or longer, the RI has been computed only for forecast days 2 to 10.

The skill of the EPS control, the skill of the ensemble-mean, and the ensemble spread

Figure 3 shows the skill of the control forecast, the skill of the ensemble-mean and the ensemble spread. The top panels of Fig. 3 show that the skill of the HEPS control is slightly higher than the skill of the EPS control, with differences statistically significant up to forecast day 6 for summer (Fig. 3, top-left panel) and 8 for winter (Fig. 3, top-right panel). The comparison between the ACC curves of HEPS and EPS(d-1) indicates that the HEPS gain in skill is about 3-6 hours, definitely less than 12 hours. The fact that at forecast day 1 the ACC of the HEPS control is lower than the ACC of the EPS control is a direct consequence of the choice of using the operational T_L319L60 verifying analysis (see comment above). The middle panels of Fig. 3 shows the ensemble spread of the two systems: the EPS spread is larger than the HEPS spread especially for the summer period (Fig. 3, left-middle panel). The bottom panels of Fig. 3 shows the skill of the ensemble-mean: apart for forecast day 1, the HEPS ensemble-mean has a higher ACC. The gain in skill induced by the increased resolution is larger for winter (Fig. 3, bottom-right panel). The RMW test shows that the differences are statistically significant for all forecast times for winter (Fig. 3, bottom-right panel) and up to forecast day 8 for summer (Fig. 3, bottom-left panel). The comparison of the HEPS and the EPS (d-1) scores indicates that that the HEPS ensemble-mean gain in skill is about 6-12 hours (Fig. 3, bottom panels), which is higher than the gain shown for the control (Fig. 3, top panels).

Brier skill score of temperature-anomaly predictions

The following two events have been considered: 'positive 850hPa temperature anomalies larger than one standard deviation' and 'negative 850hPa temperature anomalies larger than one standard deviation'. The accuracy of any probabilistic prediction of these two events have been assessed using the Brier skill score, the rank probability skill score and measures related to the Relative Operating Characteristic curve. For reasons of space, only BSS are shown, but similar conclusions could have been drawn by considering the other scores.

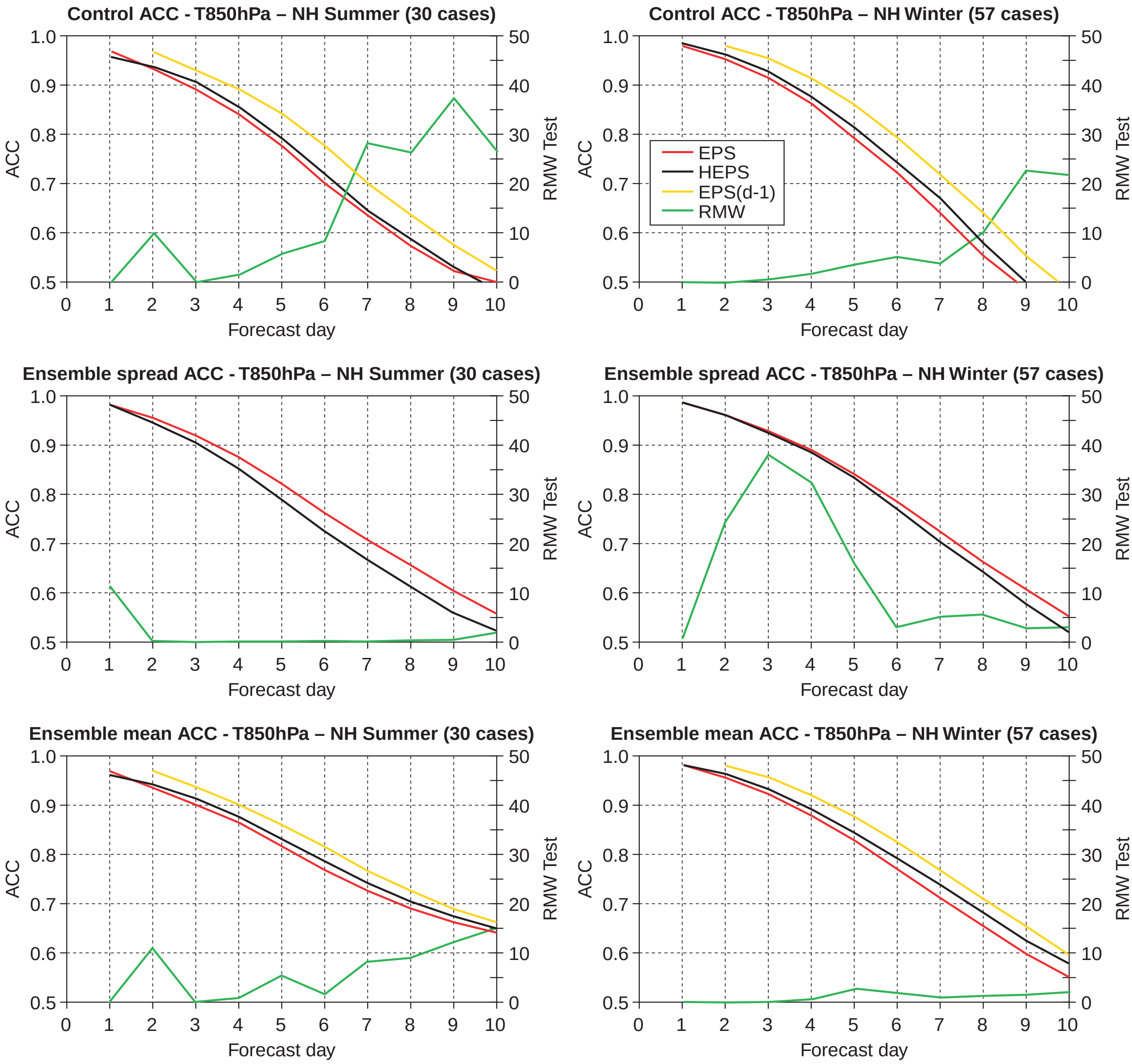

Figure 3 Top panels: mean ACC skill scores (left vertical axis) of the EPS forecasts (black curve), the HEPS forecasts (red curve) and the EPS(d-1) forecasts (yellow curve), and the Rank-Mann-Wilcoxon test value (green curve, right vertical axis) over the NH for summer (left) and winter (right). Middle panels: the ensemble spread. Bottom panels: as top panel but for the ACC skill score of the ensemble-mean. NH – summer (30 cases) NH – winter (57 cases)

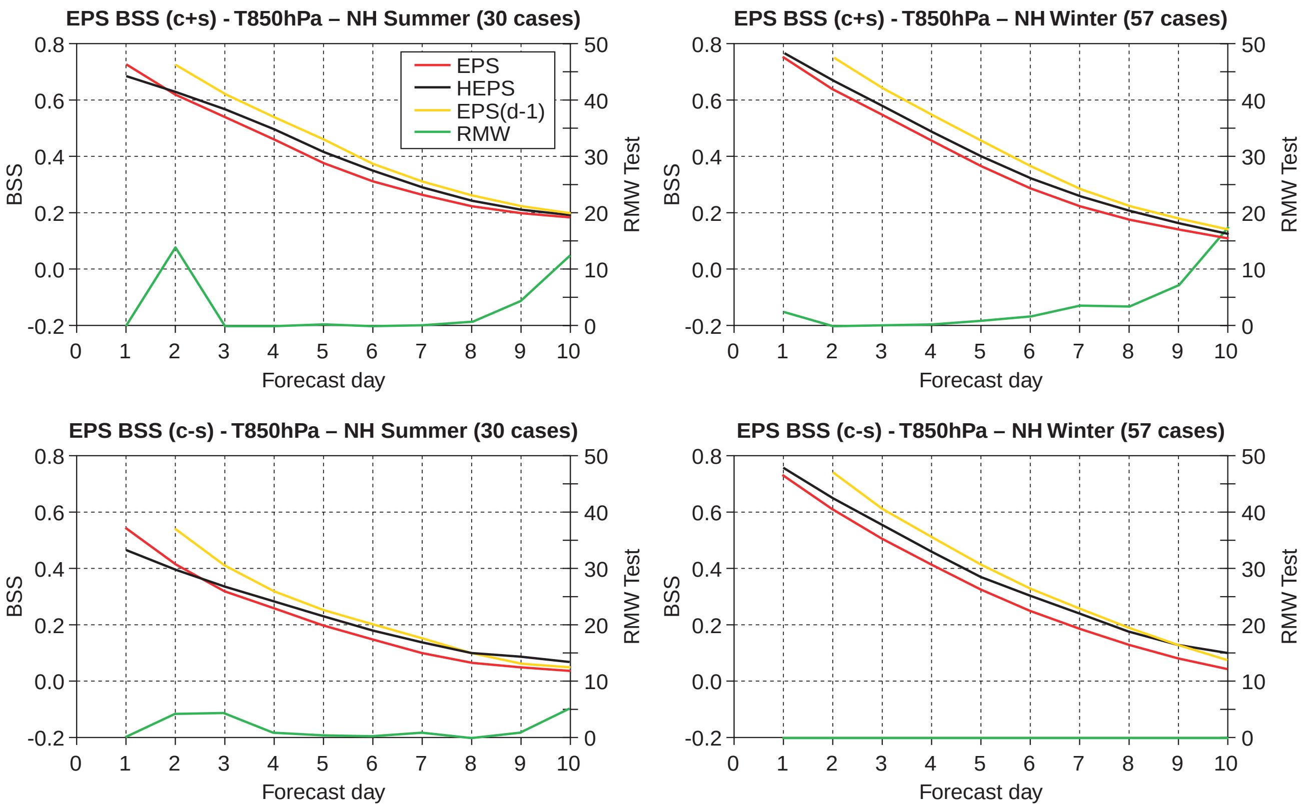

Figure 4 Top panels: The Brier skill score of the probabilistic predictions of 'positive 850hPa temperature anomalies larger than one standard deviation' (left vertical axis) for the EPS forecasts (black curve), the HEPS forecasts (red curve) and the EPS(d-1) forecasts (yellow curve), and the Rank-Mann-Wilcoxon test value (green curve, right vertical axis) over the NH for summer (left) and winter (right). Bottom panels: as top panels but for the event 'negative 850hPa temperature anomalies larger than one standard deviation'. Europe - summer (30 cases) Europe - winter (57 cases)

Figure 4 show the BSS for the three ensemble configurations, EPS, HEPS and EPS (d-1). During summer (top panels), results indicate that the HEPS performs better, with gains in predictability of 12 hours or more for forecast ranges beyond day 4 with peaks of more than 1 day at forecast days 9 and 10 for negative anomalies (Fig. 4, top-right panel). Differences between the EPS and the HEPS are significant for all forecast ranges, apart for forecast day 10 during summer (Fig. 4, top-left panel). Similar results are shown for winter (Fig. 4, bottom panels), where differences are slightly larger.

Figure 5 is similar to Fig. 4 but shows the BSSs for Europe. Compared to the NH (Fig. 4), the RMW test values for Europe indicate that the distributions of EPS and HEPS scores are less significantly different. However, the results confirm the positive impact of the resolution increase, with gains in predictability even larger than the one shown for the NH for the last forecast days.

Relative Improvement Index

Figure 6 shows the relative improvement index, RI, computed over the NH for five skill measures: the control ACC and BSS, the ensemble-mean ACC and BSS, and the EPS BSS. The results indicate that, for winter (Fig. 6, bottom panel), all the RIs are positive while, for summer (Fig. 6, top panel), the RIs are positive for all forecast days apart for some at day 2 and 10. The very small day-2 differences are due to the fact that the T_L319L60 analysis is used for verification (see comment above). For summer at forecast day 10, the negative RIs are the ones computed for the control forecast.

The long-term strategic goal sets a target of a one-day gain in predictability at forecast day 6 for EPS probabilistic predictions of moderate 850hPa temperature anomalies, using the BSS as a measure of skill. The results show that the RIs for EPS probabilistic predictions (Fig. 6, red bars) are between 55% and 7% for summer (Fig. 6, top) and between 45% and 66% for winter (Fig. 6, bottom) at forecast days 5 to 7.

It is interesting to compare the RIs computed for the BSS of the control (red), the ensemble-mean (cyan) and the EPS (burgundy). For all forecast steps, the RIs computed for the EPS are the largest. Differences are particularly large between the RIs computed for the EPS and the control forecast, especially at the end of the ten-day forecast range. These differences among the RIs indicate that, due to the increased resolution, the ensemble probability forecasts have improved more than categorical forecasts given by the control or the ensemble-mean forecast (this result is valid when using the BSS as accuracy measure).

Since the distributions of the scores for the EPS and HEPS are less statistically significant for Europe, relative improvement indices have not been computed for Europe. Despite this, an indication of the level of gains in predictability obtained over Europe can be gathered by looking at Fig. 5.

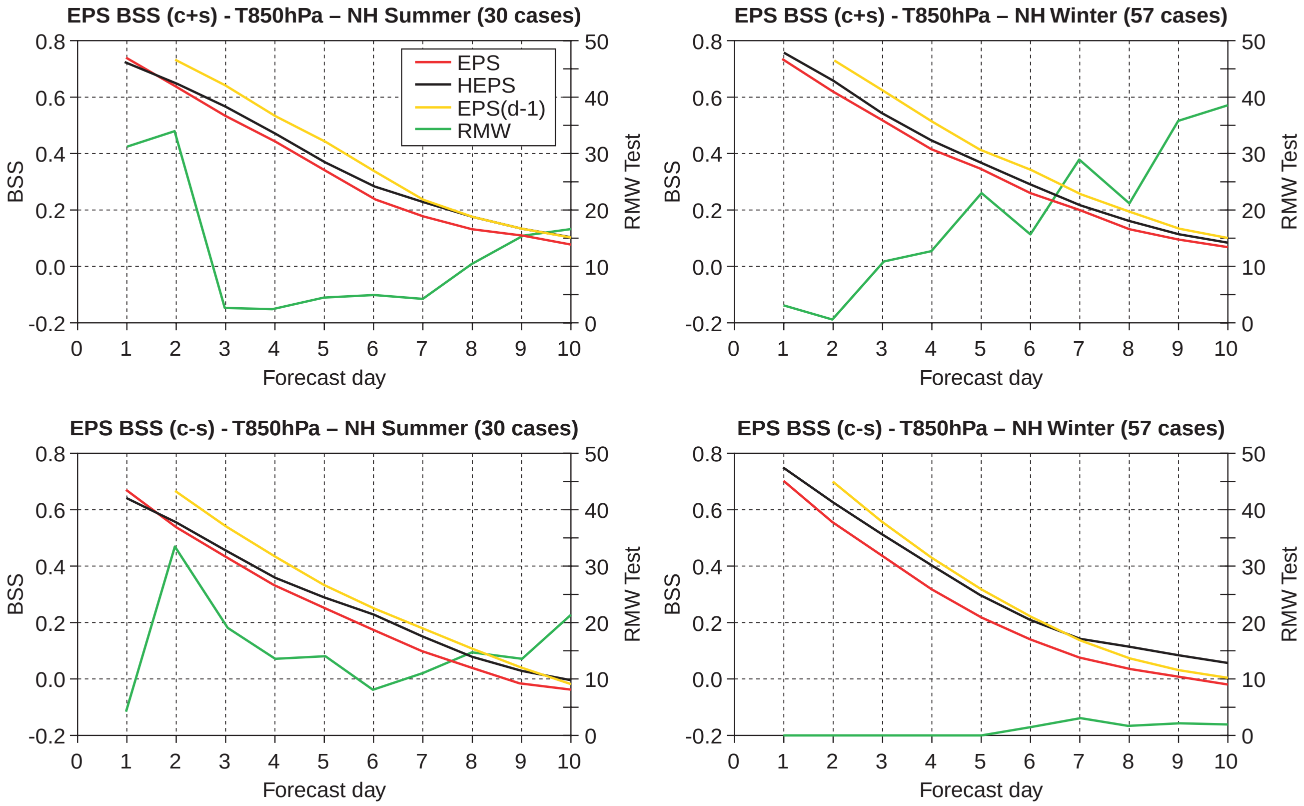

Figure 5 Top panels: The Brier skill score of the probabilistic prediction of 'positive 850hPa temperature anomalies larger than one standard deviation' (left vertical axis) for the EPS forecasts (black curve), the HEPS forecasts (red curve) and the EPS(d-1) forecasts (yellow curve), and the Rank-Mann-Wilcoxon test value (green curve, right vertical axis) over Europe for summer (left) and winter (right). Bottom panels: as top panels but for the event 'negative 850hPa temperature anomalies larger than one standard deviation'.

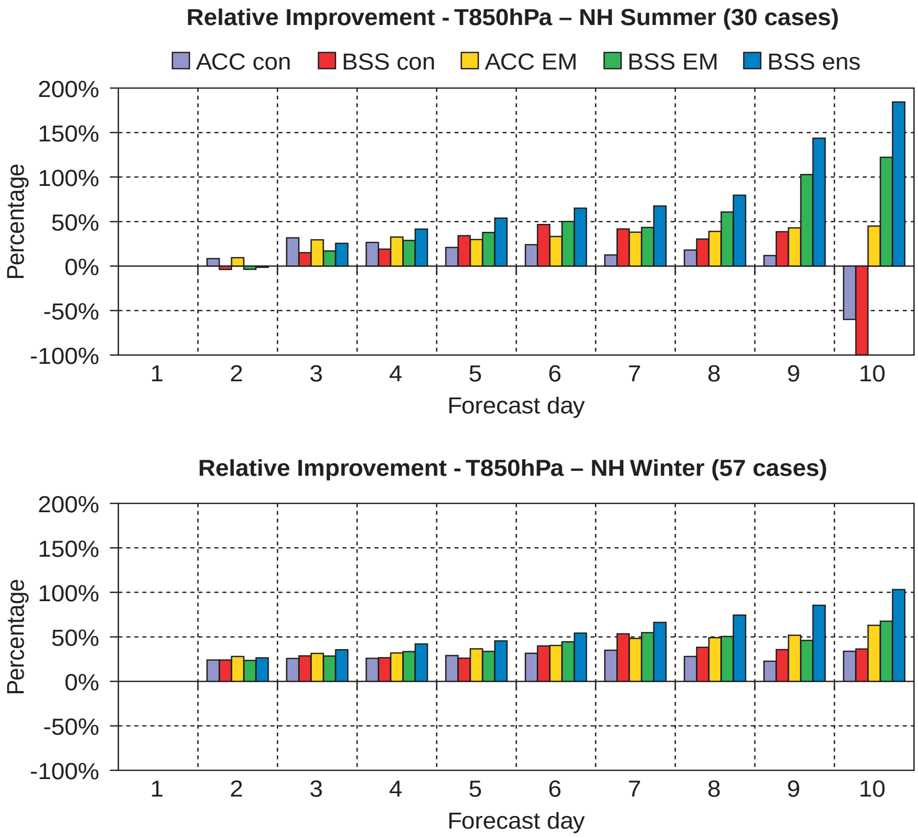

Figure 6 The Relative Improvement index (RI) for summer (top) and winter (bottom) computed over the northern hemisphere: the control ACC skill score (violet), the control Brier skill score (red), the ensemble-mean ACC skill score (yellow), the ensemble-mean Brier skill score (green) and the EPS Brier skill score (blue). A Relative Improvement of 100% indicates a gain in predictability of one day (see text for details).

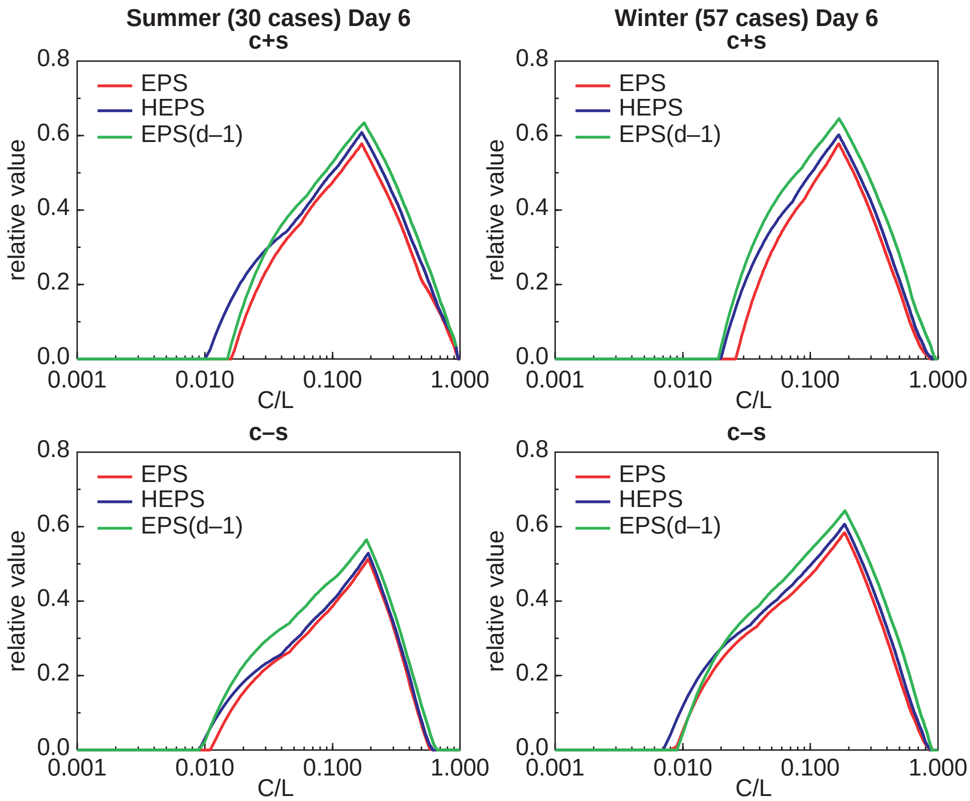

Figure 7 Value V of the EPS (red), HEPS (blue) and EPS(d–1) (green) ensemble configurations for summer (left panels) and winter (right panels) for the prediction of '850hPa temperature anomalies larger than one standard deviation' (top panels) and '850hPa temperature negative anomalies larger than one standard deviation' (bottom panels).

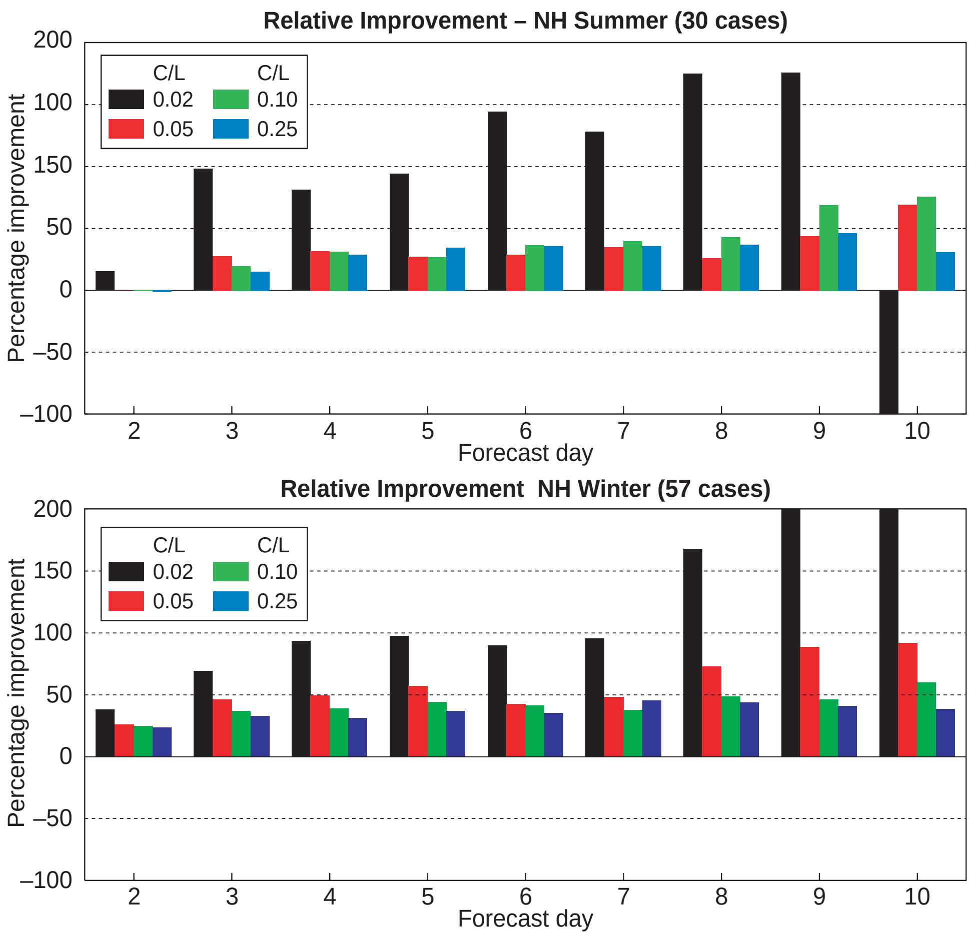

Figure 8 Relative Improvement index (RI) for the Value shown in Fig. 7 for summer (top) and winter (bottom) calculated for selected cost/loss ratios: C/L=0.02 (black), C/L=0.05 (red), C/L=0.10 (green) and C/L=0.25 (blue).

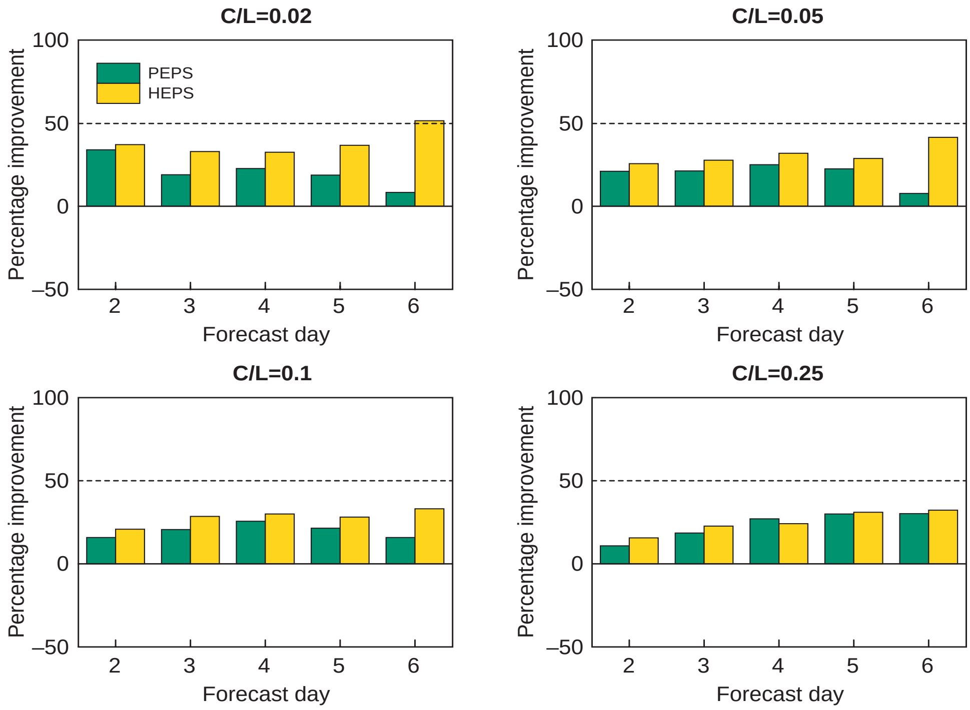

Figure 9 Relative Improvement index (RI) for the "Poor man's" ensemble PEPS (green) and the HEPS ensemble for winter (yellow) for selected cost/loss ratios: C/L=0.02 (top-left), C/L=0.05 (top-right), C/L=0.10 (bottom-left) and C/L=0.25 (bottom-right). These results have been computed for the 39 cases (out of the 57 considered so far) for which 500 hPa geopotential height fields for the other Centres were available (data for the 850hPa temperature field were not available, see text).

Potential economic value

The benefit of a forecast system will be different for different users. We illustrate this using the value diagnostic (V) derived from a simple decision-making model, the cost-loss model (*Richardson* 1998, 2000). V is a relative measure of the savings made by a forecast user: maximum value, V=1, will be obtained only with perfect knowledge of future weather, while V=0 indicates that the forecasts have no benefit over climatological information. Each user will have a different sensitivity to a particular weather event. This is represented in the model by the cost-loss ratio (C/L) which ranges between 0 and 1. Low values of C/L represent users with high sensitivity to adverse weather: the potential economic loss is high compared to the cost of taking protective action. The distribution of users' C/L is not well known, but is likely to be concentrated towards lower C/L.

Figure 7 shows V at day 6 for the two events. HEPS is consistently better than EPS for all users, with greatest benefit for those users with low cost-loss ratios. It can also be seen that the improvement relative to EPS(d-1) also varies with the user: the high-resolution benefits low cost-loss ratio users by one day or more.

Figure 8 shows the relative improvement index for V, calculated for a selection of cost-loss ratios. The variation in benefit with different users seen in Fig 7 is repeated at all forecast times. The RI for low cost-loss (0.02) is close to or exceeds 100% for forecast days 4-10. For larger cost-loss ratios, the RI is generally closer to 50%, similar to the RIs for the Brier Skill Score.

HEPS compared to "Poor man's" ensemble

An alternative approach to ensemble generation is provided by the so-called "poor man's" ensemble. One system proposed recently is the use of a small number of operational forecasts from different centres (*Ziehmann* 2000). Probabilities are obtained as in the EPS using the fraction of ensemble members that predict an event. Another option is to use a single deterministic integration and to generate probabilities using known error statistics of the deterministic model (*Atger* 1999). Here we consider a combination of the two methods. We use the operational high-resolution forecasts of ECMWF, the Met Office and DWD together with the lower resolution control forecasts available for ECMWF and the Met Office to construct a five-member ensemble of forecasts. Probabilities are generated by adding a Gaussian PDF to the ensemble-mean; the variance of the PDF is fixed, using the errors of the contributing models. This poor man's system will be referred to as PEPS. For comparison, the probabilities from the EPS and HEPS ensembles are similarly constructed, using the respective ensemble mean and a Gaussian PDF; however, for these systems, the variance of the PDF changes depending on the ensemble spread.

Because of the availability of data, results are only available for the 500hPa height and for just 39 of the 57 winter cases. Again we consider positive and negative anomalies of more than one standard deviation.

Figure 9 shows the relative improvement in V for PEPS and HEPS for the same cost-loss ratios as used in Fig 8. PEPS shows a consistent improvement over EPS. Further results (not shown) indicate that almost all of this benefit is a result of the higher resolution of the operational models used in PEPS. The increase in resolution of HEPS, although not as great as that of the operational models, is substantial enough to provide additional benefit over PEPS. The value of PEPS and HEPS is similar for larger values of C/L, but for smaller C/L the extra benefits of HEPS can be substantial.

It should be emphasised that the PEPS is only appropriate for basic single-parameter events where the forecast probabilities can be represented by a simple PDF. A major advantage of the ECMWF EPS is that probabilities for any combination of weather parameters can easily be extracted.

Equally easily, data for each EPS member can be input directly into a user's application model to provide a PDF of a user-specific parameter. Examples of such use of the EPS are ship routing (*Janssen* 2000), ice prediction (*Mureau et al.* 1997) and electricity demand (*Taylor and Buizza* 2000).

Summary

This work has shown that the new 80-km High-Resolution Ensemble Prediction System improves the estimate of the probability distribution function of forecast states. Results based on the Brier skill score of probabilistic predictions of moderate 850hPa temperature anomalies for the NH have indicated that the operational implementation of the new HEPS system has induced a gain in predictability of about 12 hours. Consideration of economic value confirms this overall level of improvement and also indicates substantially larger benefits for users with low cost-loss ratios. The new HEPS system outperforms a "poor man's" EPS containing higher resolution operational models.

The reader is referred to *Buizza & Hollingsworth* (2001a) and (2001b) for a discussion of the impact of the resolution increase on the prediction of severed storms. ❑

References

Atger, **F.**, 1999. The skill of ensemble prediction systems. *Mon. Wea. Rev.*, **127**, 1941-1953.

Barkmeijer, **J.**, **Buizza**, **R.**, & **Palmer**, **T.N.**, 1999: 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System. *Q.J.R. Meteor. Soc.*, **125**, 2333-2351.

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

Buizza, **R.** & **Hollingsworth**, **A.**, 2001a: Severe Weather Prediction using the ECMWF EPS: The European Storms of December 1999. *ECMWF Newsletter No. 90 – Spring 2001.*

Buizza, **R.** & **Hollingsworth**, **A.**, 2001b: Storm prediction over Europe using the ECMWF Ensemble Prediction System. *Mon. Wea. Rev. submitted.*

Buizza, **R.**, **Petroliagis**, **T.**, **Palmer**, **T.N.**, **Barkmeijer**, **J.**, **Hamrud**, **M.**, **Hollingsworth**, **A.**, **Simmons**, **A.**, & **Wedi**, **N.**, 1998: Impact of model resolution and ensemble size on the performance of an ensemble prediction system. *Q.J.R. Meteorol. Soc.*, **124**, 1935-1960.

Buizza, **R.**, **Miller**, **M.**, & **Palmer**, **T.N.**, 1999: Stochastic simulation of model uncertainties. *Q.J.R. Meteorol. Soc.*, **125**, 2887-2908.

Buizza, **R.**, **Barkmeijer**, **J.**, **Palmer**, **T.N.**, & **Richardson**, **D.S.**, 2000: Current status and future developments of the ECMWF Ensemble Prediction System. *Meteorol. Appl.*, **7**, 163-175.

ECMWF, 1999: A Strategy for ECMWF 1999-2008. ECMWF, Shinfield Park, Reading, RG6-9AX, UK, pp. 16.

Janssen, **P.**, 2000: Potential benefits of ensemble prediction of waves. *ECMWF Newsletter No 86 – Winter 1999-2000.*

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

Mureau, **R.**, **Wessels**, **H.**, & **van Dorp**, **H.**, 1997: Skating on EPS. *ECMWF Newsletter No 74 – Winter 1996-97.*

Palmer, **T.N.**, **Molteni**, **F.**, **Mureau**, **R.**, **Buizza**, **R.**, **Chapelet**, **P.**, & **Tribbia**, **J.**, 1993: Ensemble prediction. Proceedings of the ECMWF Seminar on Validation of models over Europe: vol. I, ECMWF, Shinfield Park, Reading, RG2 9AX, UK.

Richardson, **D.S.**, 1998: Obtaining economic value for the EPS. ECMWF *Newsletter No 80 – Summer 1998.*

Buizza, **R.**, 2000: Skill and economic value of the ECMWF Ensemble Prediction System. *Q.J.R. Meteorol. Soc.*, **126**, 649-668.

Ziehmann, **C.**, 2000. Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models. *Tellus,* **52A**, 280-298.

R. Buizza, D.S. Richardson & T.N. Palmer

MARS on the Web: a virtual tour

MARS is ECMWF's archival system. It contains decades of observations and billions of meteorological fields. A year ago we endeavoured to give an easy access to all our users to this large amount of data using a standard web browser. This article describes the status of this project.

Writing an article about a web site is a quite difficult. It is not easy to describe with words something that is already a mixture of textual information, graphics and navigation hyperlinks. The difficulty comes from trying to describe a system that is inherently self-describing.

So the best way to learn about MARS on the Web is to actually use it! However, not every reader of this newsletter may be able to visit us. Therefore, the structure of this article tries to emulate navigation through the key components of the site[1].

But if you have a web browser handy and access to our Member States website meet us at http://wms.ecmwf.int/services/mars.

1 The Web is based on the paradigm of real sites containing virtual pages. You are about to visit a virtual site using a real page.