389

# Benefits of increased resolution in the ECMWF ensemble system and comparison with poor-man's ensembles

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

Research Department

October 2002

Submitted for publication in Q. J. R. Meteorol. Soc.

For additional copies please contact

The Library
ECMWF
Shinfield Park
Reading, Berks RG2 9AX

library@ecmwf.int

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/publications.html

## Abstract

In November 2000 the resolution of the forecast model in the operational European Centre for Medium-range Weather Forecasts (ECMWF) Ensemble Prediction System was increased from a 120-km truncation scale (EPS) to an 80-km truncation scale (High-resolution EPS, HEPS). The HEPS performance is compared with the performance of EPS and of different flavours of poor-man's ensembles. Average results based on Brier skill scores and potential economic value of probabilistic predictions for 57 winter and 30 summer cases indicate that the new HEPS system is about 12-hour more skilful than the old EPS. Average (39-winter cases) results indicate that HEPS forecasts perform better than a five-center ensemble forecasts. Results also show that if forecasts are transformed into parameterised Gaussian distribution functions centred on the bias-corrected ensemble-mean and with re-scaled standard deviation, HEPS-based parameterised forecasts outperform all other configurations. Diagnostics based on parameterised forecast probabilities indicate that the different impact on the probabilistic or deterministic forecast skill is related to the fact that HEPS better represents the daily variation in the uncertainty of the atmosphere and is not simply a reflection of improved mean bias or of a better level of spread.


# 1 Introduction

Errors in atmospheric initial conditions and the approximate representation of atmospheric processes in numerical models are sources of uncertainty which limit forecast skill in a highly flow-dependent way. The variability of forecast error growth is related to the flow-dependent sensitivity of the forecast model to the above sources of uncertainty. This is particularly true for single, deterministic forecasts, with days of high quality followed by days of poor quality predictions. 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 state 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 (*Epstein* 1969, *Fleming* 1971a and 1971b, *Leith* 1974).

The Ensemble Prediction System (EPS) has been part of the operational suite at the European Centre for Medium-Range Weather Forecasts (ECMWF) since December 1992. The first version, a 33-member T63L19 configuration (spectral triangular truncation T63 with 19 vertical levels, Palmer et al. 1993, *Molteni et al.* 1996) simulated the effect of initial uncertainties by starting 32 members from perturbed initial conditions defined by perturbations rapidly-growing during the first 48 hours of the forecast range (*Buizza and Palmer* 1995). In 1996 the system was upgraded to a 51-member TL159L31 system (spectral triangular truncation T159 with linear grid; *Buizza et al*. 1998). In March 1998, initial uncertainties due to perturbations that had grown during the 48 hours previous to the starting time (evolved singular vectors, *Barkmeijer et al* 1999) were included. In October 1998, a scheme to simulate model uncertainties due to random model error in the parameterised physical processes was introduced (*Buizza et al*. 1999). As a result of these changes, the upgraded 51*$T_L$159L31 system had a better level of spread, a more skilful ensemble mean, a higher chance of including the verification analysis inside the forecast distribution and more accurate probabilistic predictions. In October 1999, following the increase of the number of vertical levels of the data-assimilation and high-resolution deterministic model from 31 to 60, the number of vertical levels in the EPS was increased from 31 to 40. In 1999, extensive experimentation started to investigate the potential benefit of further increasing the ensemble resolution from $T_L$159L40 to $T_L$255L40.

In the first part of this paper, results from the set of experiments designed to assess the impact of a further resolution increase of the ensemble system from $T_L$159L40 to $T_L$255L40 is discussed, while in the second part of the paper the performance of the $T_L$255L40 ensemble is compared to alternative ("poor-man's") methods of generating probability forecasts. Section 2 describes the higher-resolution system and the accuracy measures used to assess the ensemble performance. Section 3 documents the positive impact of this resolution increase on the ensemble system. In Section 4, the $T_L$255L40 ensemble system is compared with different variants of the poor-man's ensemble system, based on few high-resolution forecasts run at different centres. This comparison examines the value of a multi-model approach to ensemble prediction, the value of fitting parameterised distributions to ensemble forecast data and the value of applying a bias-correction and a spread re-scaling to the ensemble forecasts. Section 5 investigates possible reasons of the greater improvement in probability than in deterministic scores induced by the resolution increase. This difference implies an improvement in the HEPS representation of the day-to-day variability of the forecast PDF.

# 2 Methodology

## 2.1 The new 80-km High-resolution EPS (HEPS)

*Molteni et al* (1996), *Buizza et al* (1998) and *Buizza et al* (1999) describe the successive versions of the operational Ensemble Prediction System (EPS) used at ECMWF. Until the 20th of November 2000, the EPS was based on 51 10-day integrations performed with a $T_L$159L40 version of the ECMWF model, with unperturbed initial conditions interpolated from the $T_L$319L40 analysis. One forecast, the control, started from the interpolated analysis while the other 50 forecasts started from the analysis perturbed by adding/subtracting a combination of the dynamically fastest-growing perturbations (with total energy used as a measure of growth), scaled to have an amplitude consistent with analysis error estimates. These perturbations, called singular vectors (*Buizza and Palmer* 1995), have been shown to capture growing components of the analysis errors (*Gelaro et al* 1998). On the 21st of November 2000, the resolution of the ECMWF analysis and of the forecasting system were increased to:

- Deterministic model and analysis: from $T_L$319L60 (60km grid-point spacing) to $T_L$511L60 (40km grid-point spacing);
- Ensemble System: from $T_L$159L40 (120 km grid-point spacing) to $T_L$255L40 (80 km grid-point spacing).

Hereafter, HEPS denotes the new 80-km High-resolution EPS (ensemble membership remained the same, i.e. 50 perturbed and 1 unperturbed members).

## 2.2 Performance measures and data used

For each ensemble configuration, the following measures of ensemble performance have been considered for the 850 hPa temperature and the 500 hPa geopotential height:

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

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 degrees, for two regions, Northern Hemisphere (NH) and Europe. Results are shown mostly for NH mainly to maximise statistical significance (and also for reason of space).

The verifying analysis is defined by the operational $T_L$319L60 analysis, from which the HEPS starts, interpolated on the regular latitude-longitude 2.5-degree resolution grid, rather than the $T_L$511L60 analysis. 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 where it slightly favours the EPS (see discussion of Fig. 1 in Section 3.1).

For each area and ensemble configuration, average scores are computed separately for both the summer and the winter periods (confidence intervals have been computed, but they are not shown since otherwise figures become unreadable). The degree of similarity between the distributions of scores of the two ensemble configurations is measured by the Rank-Mann-Wilcoxon (RMW) test (*Wilks* 1995). 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, HEPS and EPS distributions are considered statistically different if RMW $\leq$10, that is if there is a 10% or lower probability that the two distributions of scores comes from the same overall distribution.

### 2.3 Relative Improvement index

To highlight the level of skill gained by the resolution increase, HEPS scores are contrasted with EPS scores and with EPS scores shifted by 1-day (EPS(d-1)), i.e. with the scores of an EPS system characterised by a 1-day gain in skill. More specifically, EPS(d-1) is the EPS forecast of one-day-shorter lead time but verifying on the same day as EPS and HEPS. For any score measure SC, the Relative Improvement index RI(SC) is defined as

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

The RI index is a normalised measure of the gain in skill obtained by configuration HEPS. RI=100% indicates an improvement equivalent to a 1-day gain in skill when measured using the score SC.

## 3 Impact of the resolution increase

### 3.1 Accuracy of the control and the ensemble-mean forecasts, and ensemble spread (850 hPa temperature field)

Figure 1 shows the T850hPa ACC of the EPS and HEPS control forecast, the ensemble spread and the ACC of the ensemble-mean. The ACC of the HEPS control is higher than the ACC of the EPS control, with statistically significant differences (from the RMW test) up to forecast day 6 for summer (Fig. 1a) and up to day 8 for winter (Fig. 1b). 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 using the operational TL319L60 analysis for verification (it should be reminded that the EPS unperturbed analysis is a TL159 interpolation of the operational TL319 analysis, while HEPS starts from a TL255 interpolation of the TL511 analysis). The HEPS spread is larger than the EPS spread especially for the summer period (Fig. 1c-d). The ACC of the HEPS ensemble-mean is higher for all but forecast day 1 (Fig. 1e-f), especially during winter (Fig. 1f). The RMW test shows that the differences are statistically significant for all forecast times for winter (Fig. 1f) and up to forecast day 8 for summer (Fig. 1e).

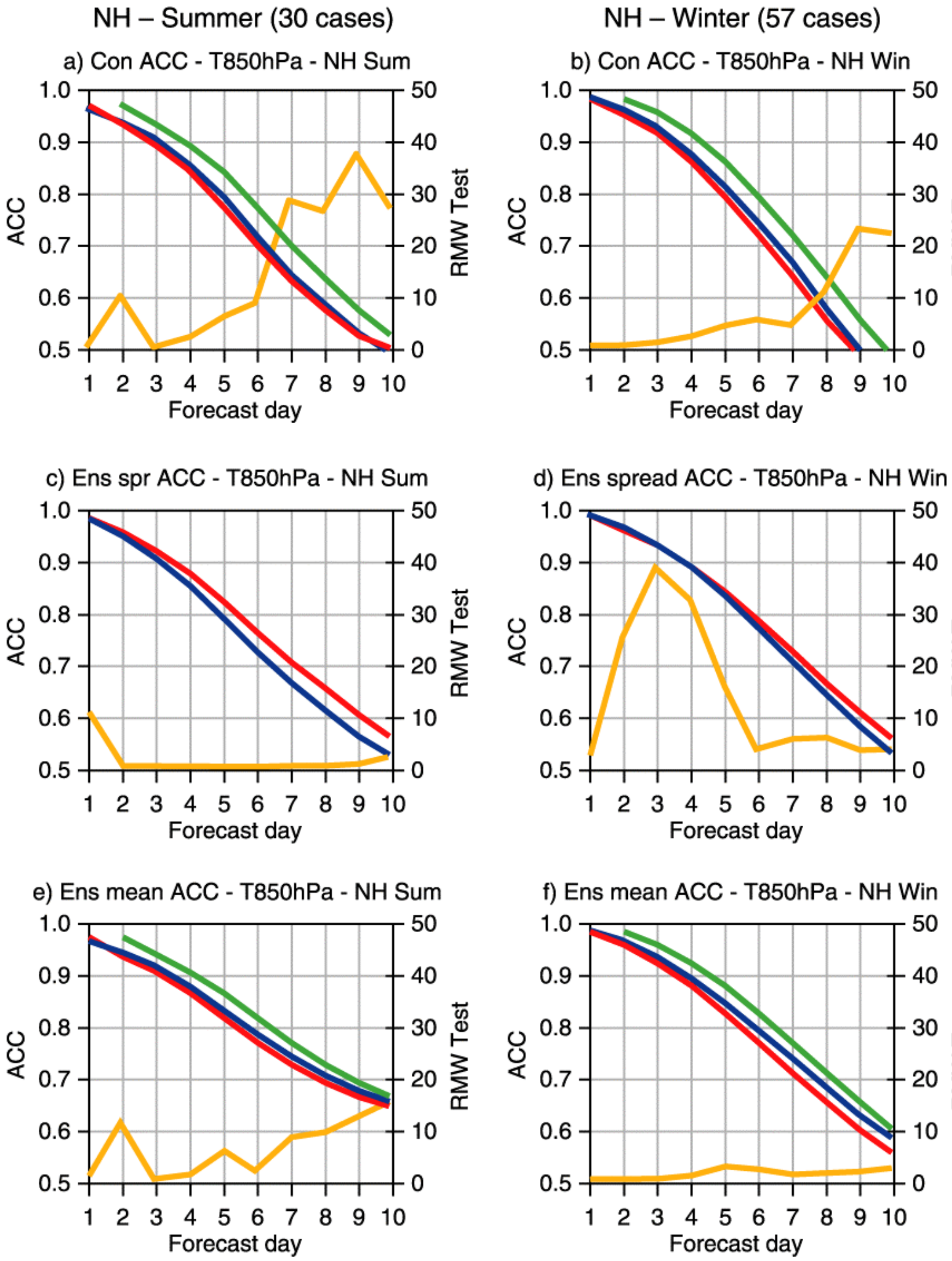


*Figure 1 850 hPa temperature. (a): mean ACC of the EPS (red), HEPS (blue) and EPS(d-1) (green) control forecasts (left vertical axis) and Rank-Mann-Wilcoxon test value (orange, right vertical axis) over the NH for summer. (b) as (a) but for winter. (c-d) as (a-b) but for the ensemble spread. (e-f) as (a-b) but for the ACC of the ensemble-mean.*

### 3.2 Brier skill score of 850 hPa temperature-anomaly predictions

The following two events have been considered: '850hPa temperature positive anomalies larger than one standard deviation' and '850hPa temperature negative anomalies larger than one standard deviation'. The accuracy of any probabilistic prediction of these two events has been assessed using the Brier skill score (BSS), the rank probability skill score and measures related to the Relative Operating Characteristic curve (see *Mason* 1982, *Stanski et al* 1989 and *Wilks* 1995 for a description of these measures). For reasons of space, only BSS are shown, but similar conclusions could have been drawn by considering the other scores.

Figure 2 shows the BSS for the three ensemble configurations, EPS, HEPS and EPS(d-1), with BSSs computed using a climatological forecast as reference. During summer (Fig. 2a, c), results indicate that the HEPS performs better, with significant differences between the EPS and the HEPS shown for all forecast ranges other than day 2 and 10. Similar results are shown for winter (Fig. 2b, d), with slightly larger positive differences significant for all forecast ranges but day 10.

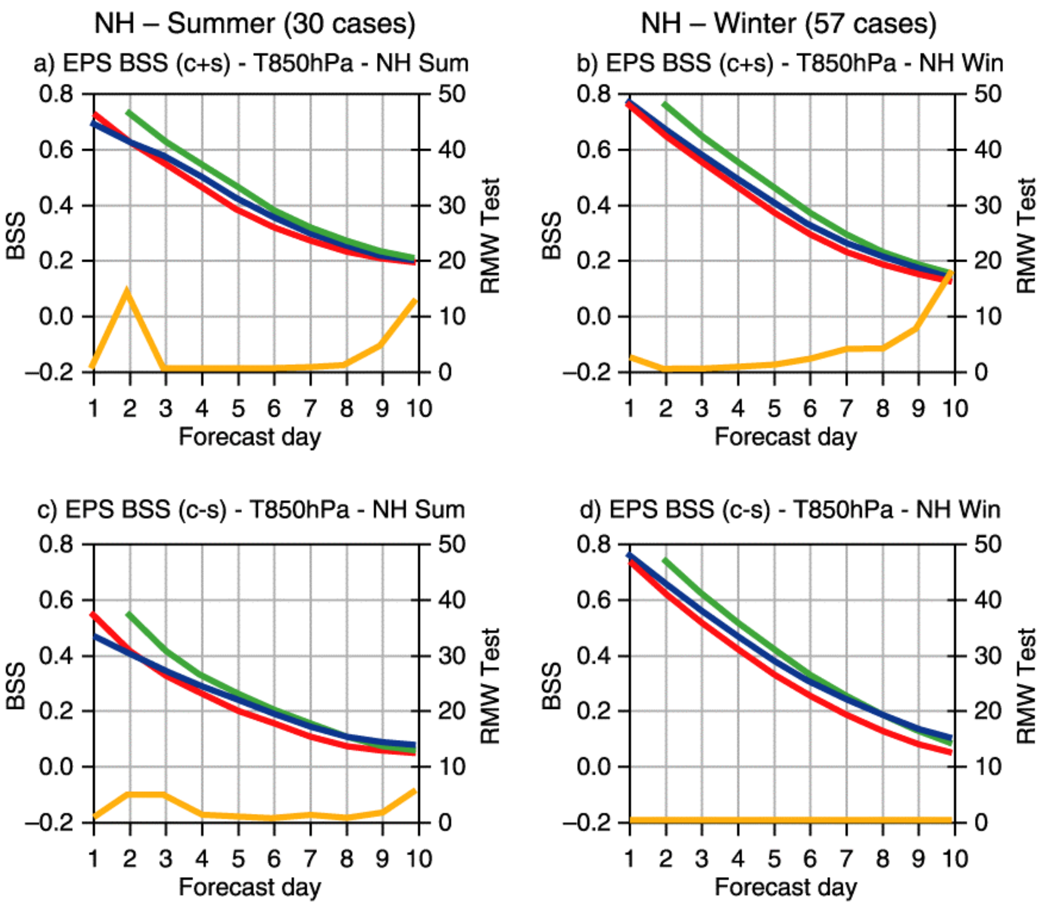


*Figure 2 (a) BSS of the EPS (red), HEPS (blue) and EPS(d-1) (green) probabilistic prediction of '850hPa temperature positive anomalies larger than one standard deviation' (left vertical axis) and Rank-Mann-Wilcoxon test value (orange, right vertical axis) over the NH for summer. (b) as (a) but for winter. (c-d) as (a-b) but for the event '850hPa temperature negative anomalies larger than one standard deviation'.*

Figure 3 is similar to Fig. 2 but shows the BSSs for Europe. Compared to the NH (Fig. 2), the RMW test values for Europe indicate that the distributions of EPS and HEPS scores are less significantly different.

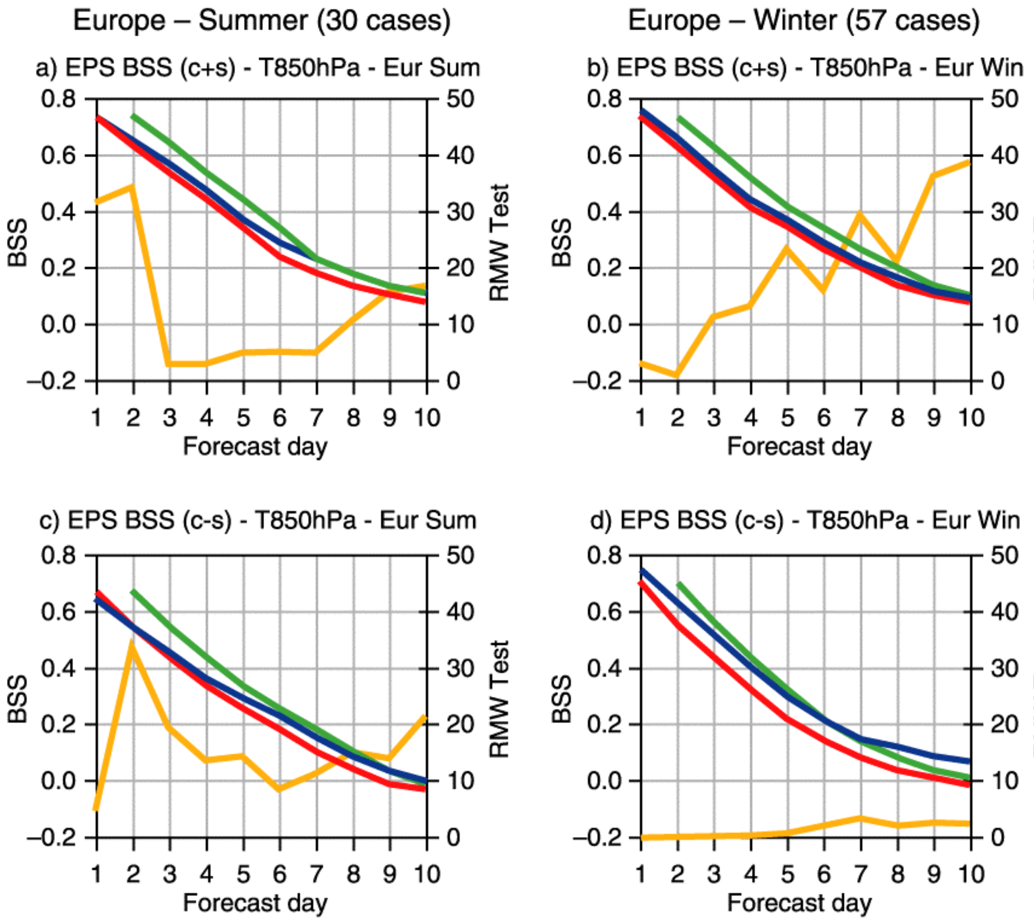


*Figure 3 (a) BSS of the EPS (red), HEPS (blue) and EPS(d-1) (green) probabilistic prediction of '850hPa temperature positive anomalies larger than one standard deviation' (left vertical axis) and Rank-Mann-Wilcoxon test value (orange, right vertical axis) over Europe for summer. (b) as (a) but for winter. (c-d) as (a-b) but for the event '850hPa temperature negative anomalies larger than one standard deviation'.*

### 3.3 Relative Improvement Index for 850 hPa temperature

Figure 4 shows the relative improvement index, RI, computed over the NH for 5 accuracy measures: control ACC and BSS, ensemble-mean ACC and BSS, and ensemble BSS. Results indicate that for summer (Fig. 4a) RIs are positive for all but forecast days 2 and 10, while for winter (Fig. 4b) all RIs are positive. The day-2 negative RI shown for the control and the ensemble-mean are due to the fact that the TL319L60 analysis is used as verification. Note that only the control forecasts (but not the ensemble-mean or EPS forecasts) show a negative RI at day-10. Considering, for example, the day 5 to 7 forecast range, RI results show that the summer HEPS probabilistic predictions are 55-70% better than the EPS (Fig. 4a) and that the winter HEPS are 45-66% better than the EPS (Fig. 4b).

Comparing the RIs computed for the BSS of the control, the ensemble-mean and the EPS it can be seen that for all forecast steps, the largest RIs are those for the EPS. In particular, the EPS RIs are always larger than the control RIs, especially at the end of the forecast period. Considering for example d+5 forecasts during summer (Fig. 4, top panel), results indicate RI=20% for the control forecast, RI=30% for the ensemble-mean and RI=55% for the BSS (in other words, a gain in skill of about 5, 7.5 and 12 hours for the three different forecast products). This indicates that the upgrade from EPS to HEPS has induced a larger relative impact on the ensemble probability forecasts than on the deterministic forecasts given by the control or the ensemble-mean forecast.

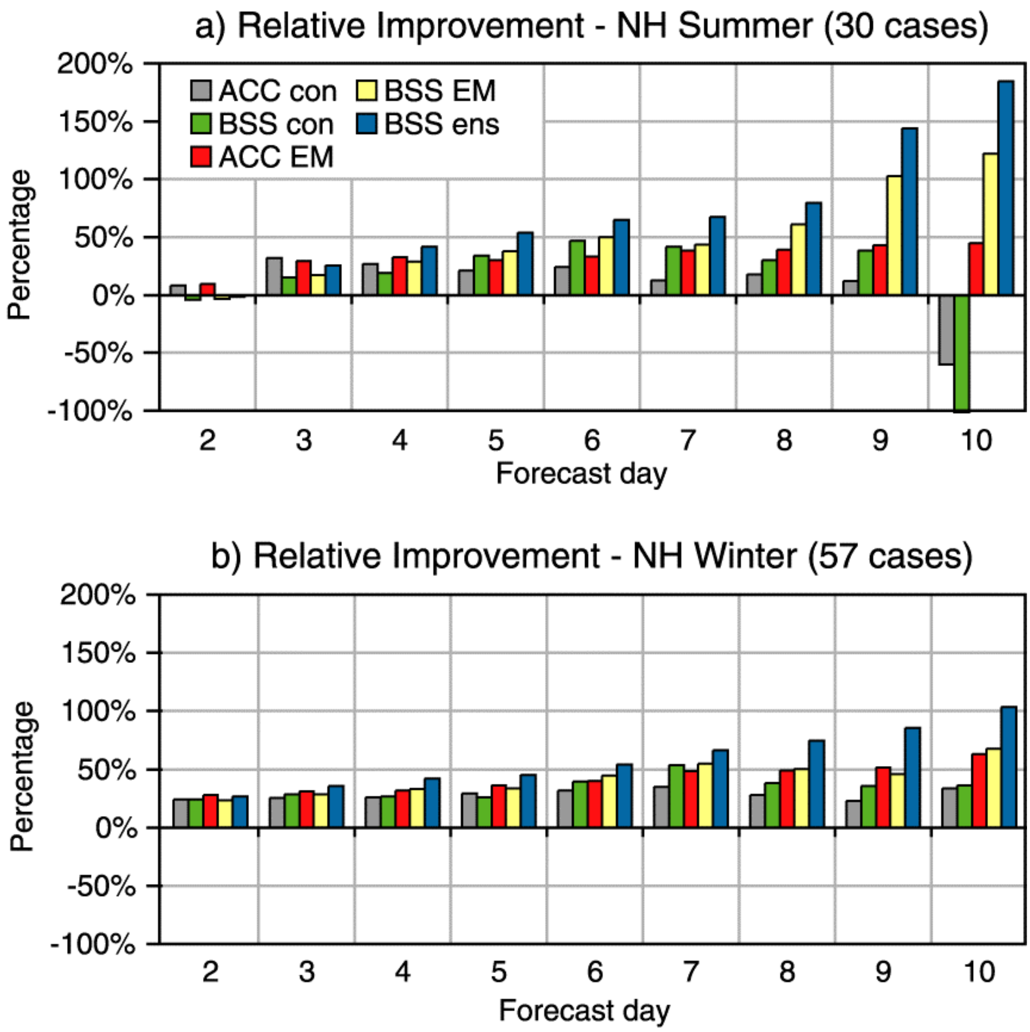


*Figure 4 (a) Relative improvement index for 850 hPa temperature summer computed over the Northern Hemisphere: control ACC (grey) the control BSS (green), the ensemble-mean ACC (red), the ensemble-mean BSS (yellow) and the EPS BSS (blue). A Relative Improvement of 100% indicate a gain in predictability of 1-day (see text for details). (b) as (a) but for winter.*

### 3.4 Potential economic value of 850 hPa temperature forecasts

The user-dependant benefit of a forecast system can be quantified using the value diagnostic (V) derived from a simple decision-making model, the cost-loss model (*Liljas and Murphy* 1994, *Murphy* 1977, Richardson 2000). According to this model, a user can decide to spend an amount C to protect himself against a possible loss L, and thus depending on whether the event occurs or not the user incurs an expense of either C or L (Table 1). The value V is a relative measure of the savings made by a forecast user in such a decision process with a cost-loss ratio C/L: maximum value, V=1, will be obtained if one has perfect knowledge of future weather, while V=0 indicates that the forecasts have no value over climatological information. Each user has a different sensitivity to a particular weather event, and this is represented by considering different cost-loss ratios (C/L) ranging from 0 to 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 low C/L (*Roebber and Bosart* 1996).

| | Event occurs | Event does not occur |
|---|---|---|
| User protect | C | C |
| User does not protect | L | 0 |

*Table 1 Cost loss decision model that describes the expenses that a user incurs if he decides to spend C to protect against a possible loss L.*

Figure 5 shows V at day 6 for the two events '850hPa temperature positive anomalies larger than one standard deviation' and '850hPa temperature negative anomalies larger than one standard deviation'. HEPS is consistently better than EPS for all users, with greatest benefit for those users with low cost-loss ratios.

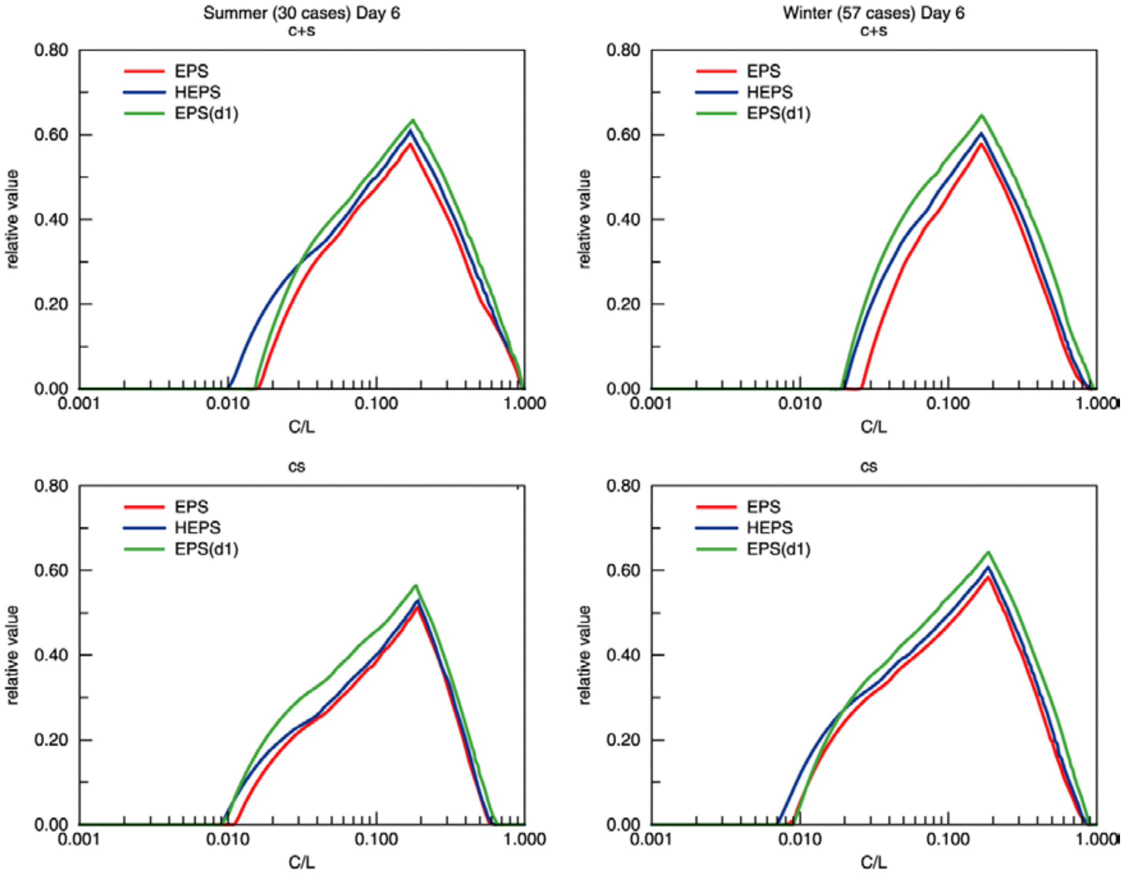


*Figure 5 Value V of the EPS (red), HEPS (blue) and EPS(d-1) (green) ensemble configurations for summer (left panels) and winter (right panels) for the prediction of '850hPa temperature anomalies larger than one standard deviation (top panels) and '850hPa temperature negative anomalies larger than one standard deviation (bottom panels).*

Figure 6 shows the Relative Improvement index for V, RI(V), calculated for a selection of cost-loss ratios (0.02, 0.05, 0.10 and 0.25). The variation in benefit with different users seen in Fig. 5 is seen at all forecast times. The RI for the lowest resolvable cost-loss (C/L=0.02) is close to or exceeds 100% for forecast days 4-10. For larger cost-loss ratios, the RI is generally closer to 40%, similar to the RIs for the BSS (Fig. 4).

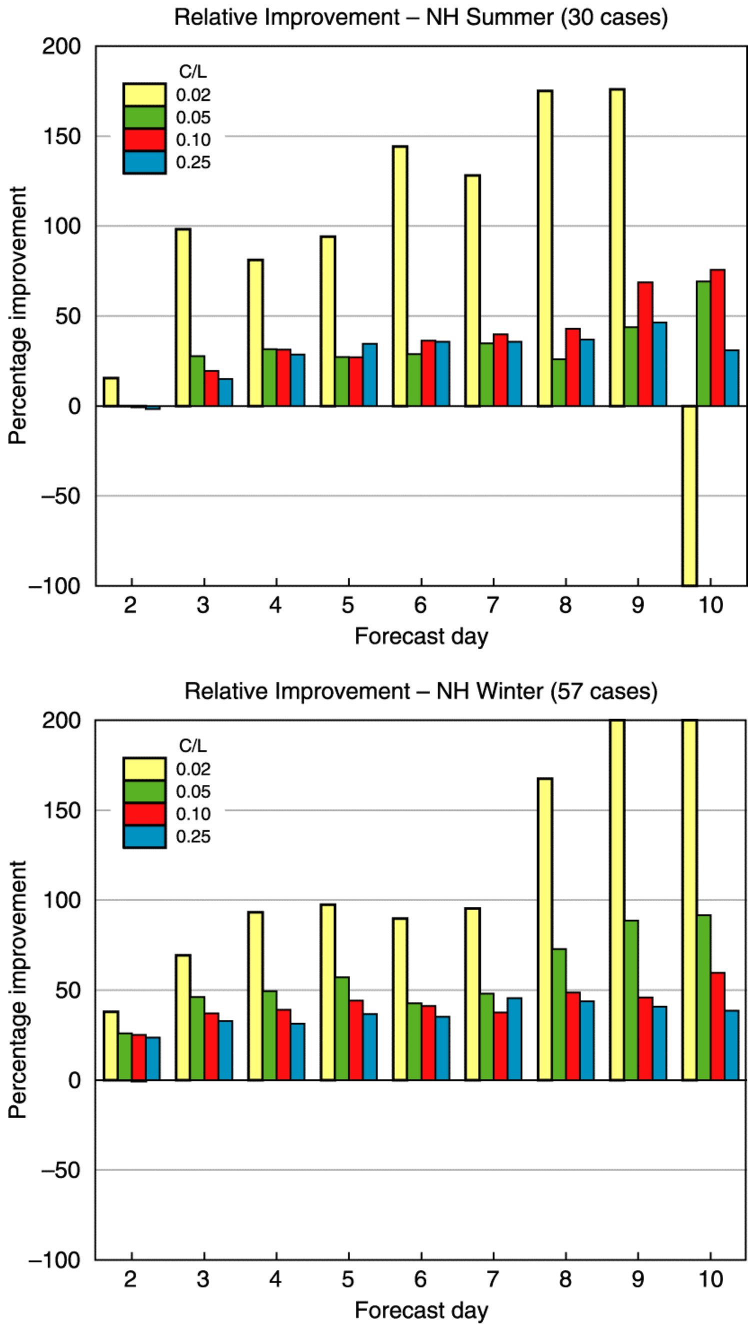


*Figure 6 Value Relative Improvement index, RI(V), for summer (top) and winter (bottom) for selected cost/loss ratios: C/L=0.02 (white columns), C/L=0.05 (light grey columns), C/L=0.10 (dark grey columns) and C/L=0.25 (black columns).*

# 4 Comparison of HEPS with a 5-member multi-centre's ensemble (MCEPS)

In the previous sections it was shown that the greatest improvement in the new HEPS system is for probability forecasts. HEPS shows a significant gain in predictability of 12-24 hours over EPS. In this section the HEPS probability forecasts are compared with a number of alternative probability forecasts systems (Table 2). These alternatives are examples of systems often referred to as "poor man's ensembles" because they are less expensive to produce than a full EPS. In this section the HEPS is compared with a poor-man's ensemble defined following *Ziehmann* (2000) approach, and in the next section the HEPS is compared with poor-man's ensembles defined following *Atger* (1999).

*Ziehmann* (2000) poor-man's ensemble is based on independent deterministic forecasts from different sources combined to generate a multi-centre ensemble. More specifically, the single deterministic high-resolution forecasts of ECMWF (~60km resolution), the UK Meteorological Office (Met Office, ~60km resolution) and Deutscher Wetterdienst (DWD, ~60km resolution) are used together with the lower resolution control forecasts available for ECMWF (~120km resolution) and the Met Office (~90km resolution) to construct a 5-member ensemble of independent forecasts. This will be referred to as the multi-centre (MC) EPS. Forecasts from the MCEPS are available only for 39 of the 57 winter cases, for forecast days 1 to 6, and for the 500 hPa geopotential height field. Thus, in this section only predictions of 500 hPa height for 39 winter cases are considered (850 hPa temperature fields were not available).

| Prediction System | Horizontal Resolution | Members | Models | Parameterised PDF | | Bias corrected |
|---|---|---|---|---|---|---|
| | | | | PDF-Mean | PDF-std | |
| EPS | 120 km | 51 | 1 | - | - | No |
| HEPS | 80 km | 51 | 1 | - | - | No |
| MCEPS | 60-120 km | 5 | 5 | - | - | No |
| HCG | 80 km | 1 | 1 | $f_c - \mu_c$ | $\sigma_c$ | Yes |
| HEPS-bias | 80 km | 51 | 1 | - | - | Yes |
| EPSG | 120 km | 51 | 1 | $f_m - \mu_m$ | $\sigma_f$ | Yes |
| HEPSG | 80 km | 51 | 1 | $f_{mH} - \mu_{mH}$ | $\sigma_H$ | Yes |
| MCEPSG | 60-120 km | 5 | 5 | $f_{mMC-\mu_{mMC}}$ | $\sigma_{MC}$ | Yes |

*Table 2 List of ensemble configurations. Symbols denote: fC the HEPS control forecast, mC the mean error of the HEPS control forecast, and sC the error variance of the HEPS control forecast; fm the EPS ensemble-mean forecast, mm the mean error of the EPS ensemble-mean forecast, and sf the re-scaled error variance of the EPS ensemble-mean forecast; fmH the HEPS ensemble-mean forecast, mmH the mean error of the HEPS ensemble-mean forecast, and sH the re-scaled error variance of the HEPS ensemble-mean forecast; fmMC the MCEPS ensemble-mean forecast, mmMC the mean error of the MCEPS ensemble-mean forecast, and sMC the error variance of the MCEPS ensemble-mean forecast.*

Figure 7 shows the relative improvement for HEPS and MCEPS relative to EPS for 500 hPa geopotential height forecasts. The improvement for HEPS for 500 hPa height is similar to that already seen for 850 hPa temperature (Fig. 4), indicating that the same conclusions on the impact of the resolution increase are valid for the two parameters. MCEPS outperforms EPS up to forecast day 5 in terms of BSS and potential economic value for C/L=0.25, but MCEPS has less potential value than EPS for smaller C/L, not surprisingly given the small size of the multi-centre ensemble. With only 5 members, it is not possible to distinguish between low probabilities that are important for users with small C/L; for these users the large size of EPS is important. One way of addressing the problem of small ensemble size is considered in the next section.

Considering the three systems (EPS, HEPS and MCEPS), Figure 7 shows that HEPS has the highest BSS and provides the greatest value for all users.

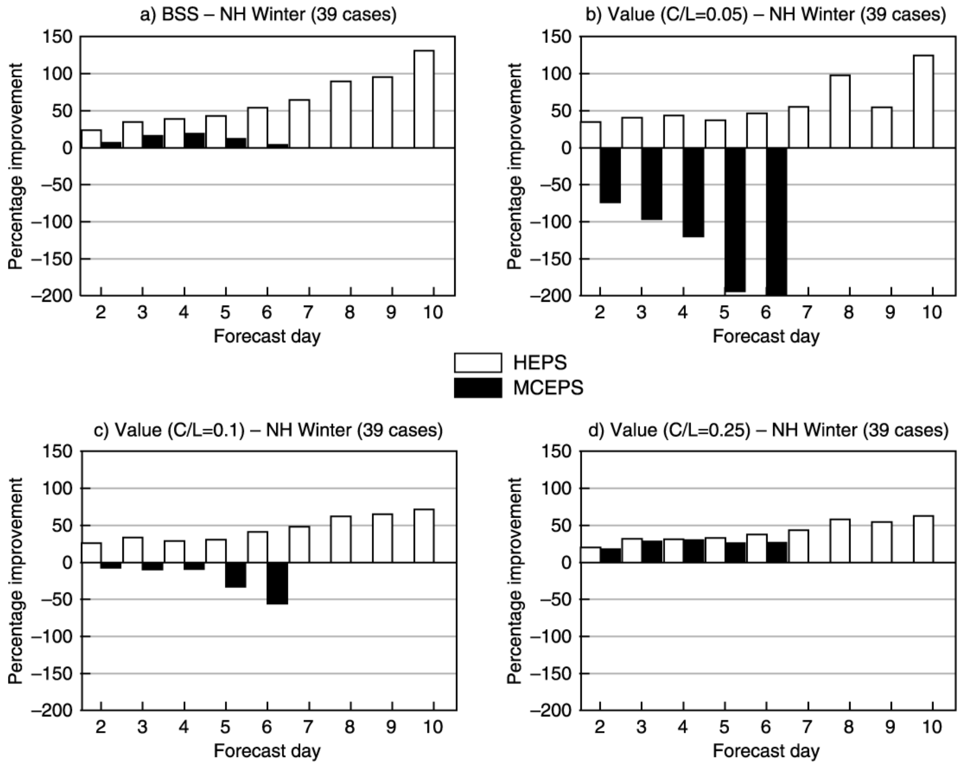


*Figure 7 Relative improvement index (RI) for 500 hPa height over the Northern Hemisphere for 39 winter cases. Bars show RI for HEPS (unfilled) and MCEPS (shaded), both relative to EPS, for BSS (a), and Value for C/L=0.05 (b), C/L=0.1 (c) and C/L=0.25 (d).*

# 5 Comparison of HEPS with parameterised poor-man's ensembles (HCG, MCEPSG)

*Atger* (1999) poor-man's ensemble is based on probability forecasts defined by a single deterministic control forecast and the distribution of errors for the control forecast. A probabilistic prediction system defined following this approach will be referred to as the high-resolution-control Gaussian (HCG).

In Section 3 the control forecast was treated as a deterministic forecast: forecast probabilities generated from the control were taken as delta functions centred at 1 or 0 depending on whether the event was predicted or not. A smoother probabilistic forecast can be generated from the control forecasts by using information on the expected error statistics of the control forecast. In theory, the control error statistics should be taken from independent data, but unfortunately for the new HEPS and also for the MCEPS, such independent error statistics are not available. Thus, the mean and variance of the forecast error are calculated at each grid-point using the set of 39 cases. Hence the potential benefits of error correction are likely to be upper bounds for what could be achieved in practice.

Consider the deterministic control forecast with mean error $\mu_c$ and error variance $\sigma_c^2$ and assume that the distribution of forecast errors is Gaussian. Then, if the 500 hPa height predicted by the control forecast is $f_c$, the PDF for the actual value will be Gaussian with mean $f_c\mu_c$ and variance $\sigma_c^2$. The probability that the actual value will be above a given threshold T can be calculated by integrating the forecast PDF:

$$P(a>T)=\frac{1}{(2\pi\sigma_c^2)^{1/2}}\int_T^{\infty}\exp\{-\frac{1}{2}[\frac{(x-(f_c-\mu_c)}{\sigma_c}]^2\}dx \qquad (2)$$

A probability forecast system based on the HEPS control forecast and using Gaussian error statistics will be referred to as HCG (High-resolution Control Gaussian). Note that in the definition of this system, the mean error mc is subtracted from the control forecast; in other words, the forecast is corrected for mean bias. In comparing the HCG probability forecasts with the HEPS, the effect of bias correction alone on the HEPS forecasts is also considered.

Figure 8 shows the relative improvement index for the HCG probability forecasts and for the bias-corrected HEPS (HEPS-bias), computed using HEPS as reference. The inclusion of error information in the HCG has some benefit over HEPS for the first two or three days for both BSS (Fig. 8a) and Value (Fig. 8b-d). But this benefit is almost completely removed if the systematic error (bias) is removed from the HEPS forecasts. The improvement due to bias-correction is substantial and increases throughout the forecast. The HCG forecasts only outscore the bias-corrected HEPS at day two and at day three for the smaller values of C/L. In this early forecast range the control forecast is generally more skilful than the ensemble members and the smoother

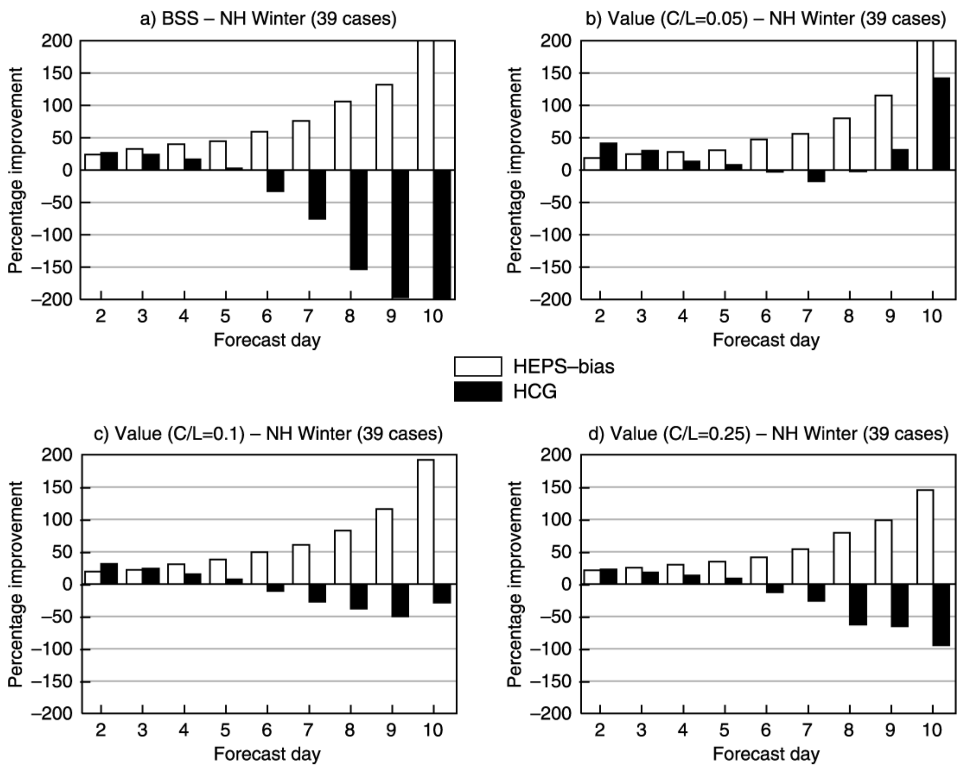


*Figure 8 Relative improvement index (RI) for 500 hPa height over the Northern Hemisphere for 39 winter cases, computed using HEPS as reference. Bars show RI for HEPS-bias (unfilled) and HCG (shaded) for BSS (a), and Value for C/L=0.05 (b), C/L=0.1 (c) and C/L=0.25 (d).*

probabilities from the parameterised PDF of the HCG may have some advantage over the raw values from HEPS for the tails of the PDF (low probability thresholds). Beyond this early range, the performance of HCG rapidly becomes worse than the performance of the bias-corrected HEPS.

A similar approach could be followed but based on the ensemble-mean and with a PDF parameterised using error statistics of the EPS (see below for details). This may provide more reliable estimates of the tails of the forecast PDF that may not be well sampled by the original EPS members. It also allows the probabilities to be corrected for under-estimation of ensemble spread. The greatest benefits of this approach may be expected for the prediction of low probabilities and for small ensembles where a limited number of members may give a poor approximation of the distribution. Hence the MCEPS may be expected to benefit more from the approach than the larger HEPS. However, the success of this approach depends on the suitability of the parameterised PDF (here the ensemble members are assumed to be distributed normally about the ensemble mean). Fitting a parameterised PDF to the ensemble members may actually remove information from the forecast if the ensemble distribution does not fit the assumed form.

Consider an EPS forecast with ensemble mean $f_m$ and spread (variance about the ensemble mean) $s^2$. Let the mean error and error variance of the ensemble mean be $\mu_m$ and $\sigma_m^2$ respectively and let the average spread over a large number of cases be $<s^2>$. Then a parameterised forecast PDF can be constructed as Gaussian with mean $f_m - \mu_m$ and variance $\sigma_f^2$ where

$$\sigma_f^2 = \frac{s^2}{<s^2>}\sigma \tag{3}$$

This ensures that on average the forecast variance matches the ensemble mean error variance, while allowing the forecast variance to vary from case to case depending on the ensemble spread.

Parameterised probability forecasts based on HEPS and MCEPS systems will be referred to as HEPSG and MCEPSG. Since these forecasts are both corrected for model bias, they are compared to the bias-corrected HEPS probabilities to explore the additional benefit to be obtained from the parameterisation.

Figure 9 shows the relative improvement for MCEPSG and HEPSG computed using HEPS-bias as reference. The effect of the parameterisation on the HEPS forecasts is positive throughout the forecast, although substantial improvements are found mostly later in the forecast and for smaller cost-loss ratios. Although the parameterisation of the PDFs reduces the explicit dependence of the forecasts on ensemble size, the multi-centre MCEPSG is still not as skilful as HEPSG. The greater number of members in HEPSG is still beneficial in providing better estimates of the parameters of the Gaussian PDF.

It should be emphasised that the parameterisation approach followed here 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 ensembles is that probabilities for any multivariate combination of (time-lagged and spatially-separated) weather parameters can easily be extracted. Equally easily, data for each ECMWF ensemble 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 ECMWF ensemble systems are ship routing (*Hoffschildt et al*. 1999), ice prediction (*Mureau et al.* 1997) and electricity demand (*Taylor and Buizza* 2002).

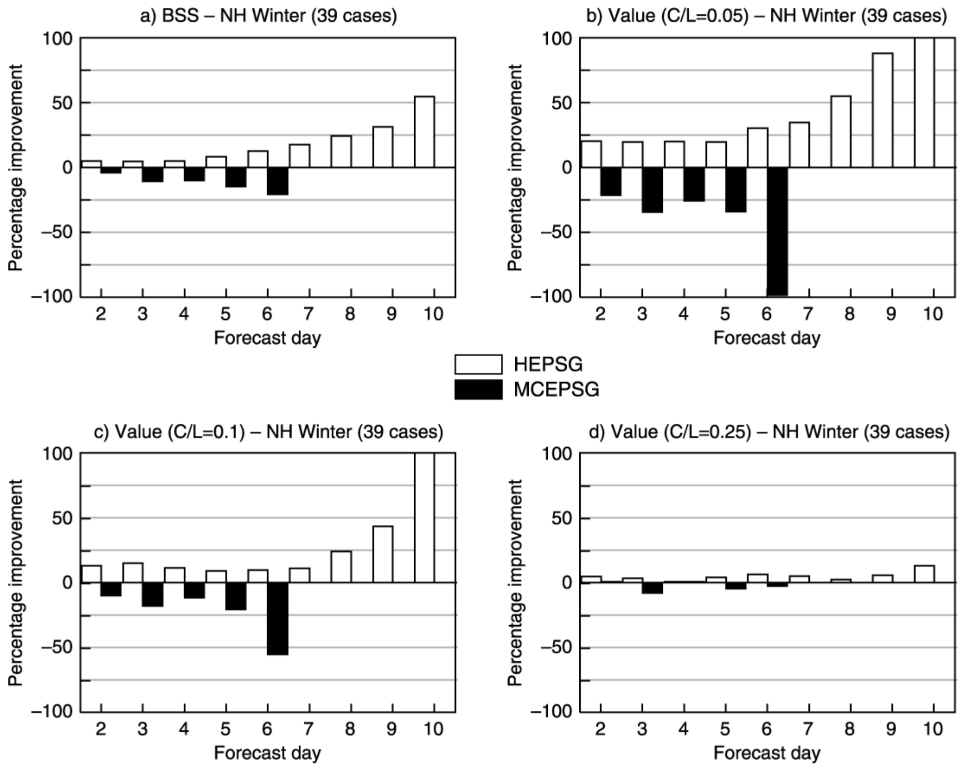


*Figure 9 Relative improvement index (RI) for 500 hPa height over the Northern Hemisphere for 39 winter cases, computed using HEPS-bias (the bias-corrected HEPS) as reference. Bars show RI for parameterized ensembles HEPSG (unfilled) and MCEPSG (shaded) for BSS (a), and Value for C/L=0.05 (b), C/L=0.1 (c) and C/L=0.25 (d).*

# 6 The impact of model resolution on deterministic and probabilistic scores

Comparison of the EPS and HEPS systems showed a greater relative improvement in probability scores (between 12 and 24 hours) than in the deterministic scores (between 3 and 12 hours). The difference between the control and ensemble RI increases throughout the forecast range. The parametric approach of the previous section is used to investigate these differences, including the effect of model bias and spread under-estimation.

The HEPSG probabilities of the previous section were constructed using Gaussian PDFs centred on the bias-corrected ensemble mean and with variance based on the ensemble spread but corrected to match, on average, the ensemble mean error variance (Eq. 3). Equivalent probability forecasts, EPSG, can be constructed from the low resolution EPS, using the appropriate bias and corrected spread. Comparison of the HEPSG and EPSG (Fig. 10) shows the relative improvement once the mean model bias and spread have been corrected. Remaining differences are then due to day-to-day variations in spread and the ensemble mean. Figure 10 shows the relative improvement of HEPSG relative to EPSG for the set of 39 winter cases discussed in the previous section. The consistent improvement, increasing with forecast day, is apparent for both BSS and value and compares well with the corresponding improvements seen for the uncorrected model output (Fig 7; note that the vertical scales in Fig. 7 and 10 are different).

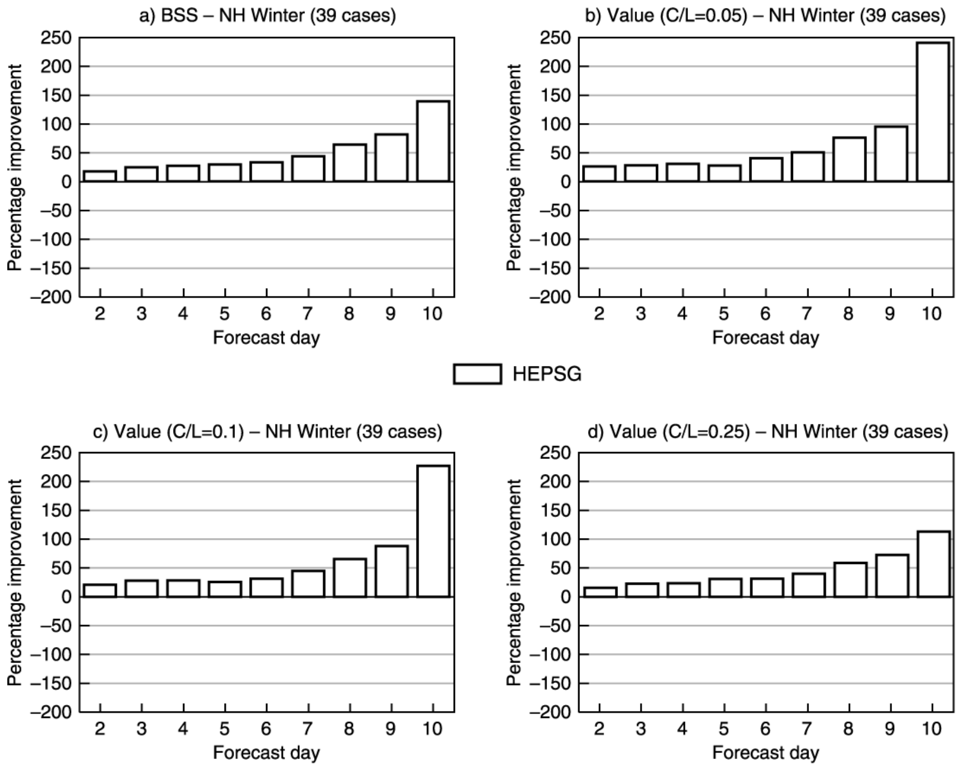


*Figure 10 Relative improvement index (RI) for 500 hPa height over the Northern Hemisphere for 39 winter cases, relative to EPSG (the parameterised EPS). Bars show RI for parameterised ensembles HEPSG for BSS (a), and Value for C/L=0.05 (b), C/L=0.1 (c) and C/L=0.25 (d).*

The substantial improvement of the ensemble probability forecasts in the HEPS configuration is not simply a reflection of improved mean bias or a better average level of spread. Rather, it represents an improvement of the capability of the ensemble to represent the day-to-day variability of the (unknown) underlying PDF of uncertainty. The BSS can be used to measure this improvement.

An idealised, perfectly specified EPS would consist of an effectively infinite number of forecasts, all equally likely and together representing the full uncertainty of both analysis and model errors. The EPS control forecast can be considered as a single representative member of such an ideal ensemble, and the BSS of the control, BSSc, can be used to estimate the BSS of this hypothetical perfect ensemble and of a finite-sized, M-member ensemble drawn from this perfect distribution (*Richardson* 2001):

$$BSS_M^{\mathrm{perf}} = \frac{(M+1)BSS_C + M - 1}{2M} \tag{4}$$

This estimate can be compared with the actual EPS BSS to give a measure of how well the EPS meets the expectation of a perfectly representative ensemble system. The results for the 39 winter cases are shown in Figure 11. This shows the actual BSS for the EPS as a fraction of the hypothetical $BSS_M^{\mathrm{perf}}$. The improvement for the new HEPS is consistent and increases throughout the forecast. This matches the difference in RI between the control and the EPS. Figure 11 shows that the HEPS performance is substantially closer to the ideal level than the previous EPS.

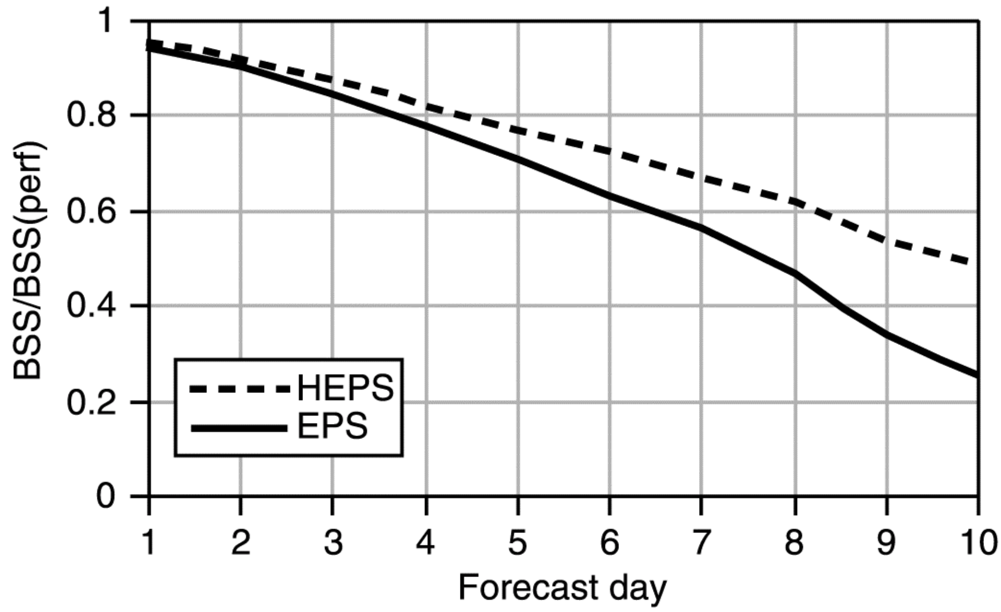


*Figure 11 BSS as a fraction of the expected score for a perfect ensemble for 500 hPa height over the Northern Hemisphere for 39 winter cases (see text for details) for EPS (red) and HEPS (blue).*

Equation 4 can also be used to quantify the expected gain in the BSS for the ensemble probability forecasts, for a given improvement of the deterministic forecast. So, given BSSC for EPS and HEPS, the expected improvement for ensemble probability forecasts is [$BSS_M^{perf}$(HEPS)- $BSS_M^{perf}$(EPS)]. Figure 12 compares this expected gain with the actual improvement in BSS for the ensemble using the ratio *I*

$$I = \frac{BSS_M\,(\mathrm{HEPS}) - BSS_M\,(\mathrm{EPS})}{BSS_M^{\mathrm{perf}}\,(\mathrm{HEPS}) - BSS_M^{\mathrm{perf}}\,(\mathrm{EPS})} \tag{5}$$

The improvement *I* found for HEPS is in general two to three times greater than could be expected as a direct result of the improvement of the deterministic model. As previously shown in Fig 10, this improvement remains once effects of bias and spread have been removed, and it is presumed that the benefit is due to the new HEPS configuration being able to capture in a better way the daily variation of the forecast uncertainty.

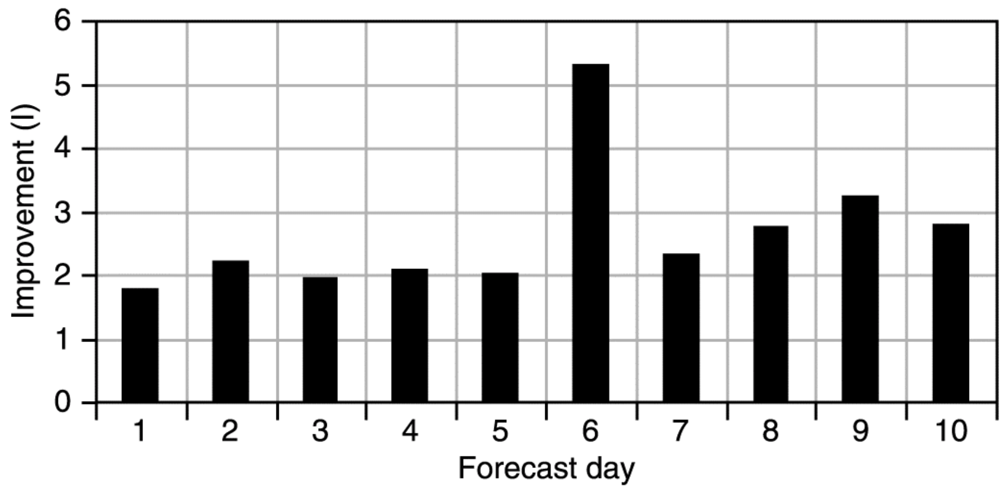


*Figure 12 Actual increase in BSS for HEPS over EPS, expressed as a fraction of the increase expected from the deterministic improvement (see text for details). Results for 500 hPa height over the Northern Hemisphere for 39 winter cases.*

Another possible explanation proposed by *Toth et al* (2002) is that, at increased resolution, the forecast model is resolving details that may deteriorate the skill of a deterministic forecast but improve the skill of a probabilistic forecast. *Toth et al* (2002) found that for single deterministic forecasts using a higher resolution can reduce the forecast error during the first few days, because of a better description of both large and small scales, but it has a detrimental effect afterwards. They argued that this is due to a combination of a progressive loss of correspondence between predicted and observed small scales, and the fact that the small scales can act as a source of random noise that affects the accuracy of the large-scale features. They pointed out that this happens despite the fact that a low resolution forecast gives a less realistic view of reality. By contrast, using a higher resolution can lead to better skill for an ensemble system because each single member of the ensemble gives a more realistic representation of reality. In other words, the fact that a higher-resolution model gives a more realistic representation of reality guarantees a forecast improvement when used in an ensemble configuration.

In contrast with *Toth et al* (2002), the results discussed in this paper indicate that using a higher resolution improves the skill of both single deterministic and ensemble-based probabilistic forecasts for the whole 10-day forecast range, the difference being that the improvement is more substantial for ensemble-based, probabilistic forecasts. It is not clear at this stage whether *Toth et al* (2002)'s argument can be used to explain such a difference.

## 7 Conclusions

The 80-km High-Resolution Ensemble Prediction System (HEPS) gives a better estimate of the probability distribution function of 850 hPa temperature and 500 hPa geopotential height forecast states than the EPS. Average (57-winter and 30-summer cases) results based on BSS of probabilistic predictions of moderate 850hPa temperature anomalies for the NH have indicated that the operational implementation of the new HEPS system has resulted in gain in predictability of about 12 hours. Consideration of economic value supports this overall level of improvement and also indicates substantially larger benefits for users with low cost-loss ratios. This positive impact of the resolution increase on single cases of extreme weather prediction has been documented by *Buizza and Hollingsworth* (2002).

The performance of the HEPS has been compared with the performance of different variants of poor-man's ensemble systems (Table 2) based on a small number of forecasts from different centres (ECMWF, UKMO and DWD). Following Ziehmann (2000) a 5-member multi-center poor-man's ensemble has been considered (MCEPS). Average (39-winter cases) results based on potential economic value have indicated that raw HEPS forecasts perform better than the MCEPS. The larger HEPS membership (51 versus 5) is one of the reasons of the better performance. Then, following Atger (1999), an ensemble based on a parameterised distribution function centred on the ECMWF high-resolution forecast with standard deviation defined by the control error standard deviation, has been considered (High-resolution Control Gaussian ensemble, HCG). HEPS has been shown to perform worse than HCG for forecast steps up to day 5 and better thereafter. This result has been related to the fact that the HCG probability density function has been bias-corrected, while the HEPS has not. Results have shown that a bias-corrected HEPS (HEPS-bias) outperforms HCG for all forecast steps.

Finally, the raw EPS, HEPS and MCEPS forecasts have been transformed into parameterised Gaussian distribution functions centred on the bias-corrected ensemble-mean and with re-scaled standard deviation, specifically, into EPSG, HEPSG and MCEPSG. Results of this latest comparison have shown that HEPSG outperforms all other configurations for every forecast step.

One of the most striking results from the comparison of EPS and HEPS has been that the accuracy of probabilistic forecasts has been improved more than the accuracy of deterministic forecasts. Parameterised probability forecasts have been used to identify potential reasons for this different impact of resolution increase. Results suggest that the different impact on the skill is related to the fact that the HEPS represents in a better way the daily variation of forecast uncertainty, and it is not a simple reflection of improved mean bias or of a better level of spread. This may also be related to the fact that, at increased resolution, the forecast model is resolving details that may deteriorate the skill of a deterministic forecast but improve the skill of a probabilistic forecast (Toth et al 2002).

On the 21st of November 2000, HEPS became the ECMWF operational ensemble configuration.

**Acknowledgement**

Graham Holt is acknowledged for his essential technical support and optimal scheduling of the heavy ensemble experimentation that made this study possible, Tony Hollingsworth for his very valuable comments to a first version of this manuscript, and Els Kooij-Connally and Rob Hine for their editorial and graphical contributions. Two anonymous referees are thanked for their insightful comments that helped improve the quality of this manuscript.

## References

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

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

Buizza, R. and Hollingsworth, A., 2002: Storm prediction over Europe using the ECMWF Ensemble Prediction System. *Meteorol. Appl.*, 9, 289-305.

Buizza, R., Petroliagis, T., Palmer, T. N., Barkmeijer, J., Hamrud, M., Hollingsworth, A., Simmons, A., and 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., and Palmer, T. N., 1999: Stochastic simulation of model uncertainties. *Q. J. R. Meteorol. Soc.*, **125**, 2887-2908.

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

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

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

Gelaro, R., Buizza, R., Palmer, T. N., and Klinker, E., 1998: Sensitivity analysis of forecast errors and the construction of optimal perturbations using singular vectors. *J. Atmos. Sci.*, **55**, 6, 1012-1037.

Hoffschildt, M., Bidlot, J.-R., Hansen, B., and Janssen, P.A.E.M., 1999: Potential benefit of ensemble forecasts for ship. *ECMWF Research Department Technical Memorandum No. 287*, ECMWF, Shinfield Park, Reading, RG6-9AX, UK.

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

Liljas, E., and Murphy, A. H., 1994: Anders Angstrom and his early papers on probability forecasting and the use/value of weather forecasts. *Bull. Amer. Met. Soc.*, **75**, 1227-1236.

Mason, I., 1982. A model for assessment of weather forecasts. *Aust. Met. Mag.*, **30**, 291-303.

Stanski, H. R., Wilson, L. J., and Burrows, W. R., 1989. Survey of common verification methods in meteorology. *World Weather Watch Technical Report No.* 8, WMO/TD. No. 358, World Meteorological Organization. 114 pp.

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

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

Murphy, A. H., 1977: The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. *Mon. Wea. Rev.*, **105**, 803-816.

Palmer, T. N., Molteni, F., Mureau, R., Buizza, R., Chapelet, P., and 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., 2000: Skill and economic value of the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, **126**, 649-668.

Richardson, D. S., 2001: Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Q. J. R. Meteorol. Soc.*, **127**, 2473-2489.

Roebber, P. J., and Bosart, L. F., 1996: The complex relationship between forecast skill and forecast value: a real-world analysis. *Weather and Forecasting*, **11**, 544-559.

Taylor, J., and Buizza, R., 2002: Energy demand prediction using the ECMWF Ensemble Prediction System. Intern. *J. Forecast.*, in press.

Toth, Z., Zhu, Y., Szunyogh, I., Iredell, M., and Wobus, R., 2002: Does increased model resolution enhance predictability? AMS Symposium on Observations, Data Assimilation and Probabilistic Prediction, 13-17 January 2002, Orlando, Florida (available from http://sgi62.wwb.noaa.gov:8080/ens/ens_info.html ), paper J1.8, pp 7.

Wilks, D. S., 1995: *Statistical methods in the atmospheric sciences*. Academic Press, Inc., San Diego, pp. 467 (ISBN 0-12-751965-3).

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