Q. J. R. Meteorol. Soc. (2003), **129**, pp. 1269–1288 doi: 10.1256/qj.02.92

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

By R. BUIZZA*, D. S. RICHARDSON and T. N. PALMER
European Centre for Medium-Range Weather Forecasts, UK

(Received 25 March 2002; revised 9 October 2002)

SUMMARY

In November 2000 the resolution of the forecast model in the operational European Centre for Medium-Range Weather Forecasts Ensemble Prediction System was increased from a 120 km truncation scale (EPS) to an 80 km truncation scale (High-resolution EPS or HEPS). The HEPS performance is compared with that of EPS and with different flavours of poor-man's ensembles. Average results based on Brier skill scores and the potential economic value of probabilistic predictions for 57 winter and 30 summer cases indicate that the new HEPS system is about 12 hours more skilful than the old EPS. Averages over 39 winter cases indicate that HEPS forecasts perform better than five-centre ensemble forecasts. Results also show that if forecasts are transformed into parametrized Gaussian distribution functions centred on the bias-corrected ensemble mean and with re-scaled standard deviation, HEPS-based parametrized forecasts outperform all other configurations. Diagnostics based on parametrized 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.

KEYWORDS: Ensemble prediction Poor-man's ensemble Predictability

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,b; 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 $T_L159L31$ 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 parametrized physical processes was introduced (Buizza *et al.* 1999). As a result of these changes, the upgraded 51-member system had a better level of spread, a more skilful ensemble mean, a higher chance of including the verification analysis inside

* Corresponding author: ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, UK. e-mail: buizza@ecmwf.int

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_L159L40 to T_L255L40.

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_L159L40 to T_L255L40 is discussed, while in the second part of the paper the performance of the T_L255L40 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_L255L40 ensemble system is compared with different variants of the poor-man's ensemble system, based on a 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 parametrized 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 for the greater improvement in probability than in deterministic scores induced by the resolution increase. This difference implies an improvement in the high-resolution EPS representation of the day-to-day variability of the forecast PDF.

2. METHODOLOGY

(*a*) *The new 80 km High-resolution EPS (HEPS)*

Molteni *et al.* (1996) and Buizza *et al.* (1998, 1999) describe the successive versions of the operational EPS used at ECMWF. Until 20 November 2000 the EPS was based on 51 10-day integrations performed with a T_L159L40 version of the ECMWF model, with unperturbed initial conditions interpolated from the T_L319L40 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 21 November 2000, the resolutions of the ECMWF analysis and of the forecasting system were changed as follows.

- Deterministic model and analysis: from T_L319L60 (60 km grid point spacing) to T_L511L60 (40 km grid point spacing).
- Ensemble system: from T_L159L40 (120 km grid point spacing) to T_L255L40 (80 km grid point spacing).

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

(*b*) *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 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 and from 22 January to 15 February). 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—the northern hemisphere (NH) and Europe. Results are shown mostly for NH, mainly to maximize statistical significance but also for reasons of space.

The verifying analysis is defined by the operational $T_L319L60$ analysis, from which the HEPS starts, interpolated on the regular latitude–longitude 2.5-degree resolution grid, rather than the $T_L511L60$ 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(a)).

For each area and ensemble configuration, average scores are computed separately for the summer and winter periods (confidence intervals have been computed, but they are not shown since otherwise figures become unreadable). The degree of similarity between the score distributions 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 $\leqslant 10$, i.e. if there is a 10% or lower probability that the two distributions of scores come from the same overall distribution.

(*c*) *Relative improvement index*

To highlight the level of skill gained by the resolution increase, HEPS scores are contrasted with EPS scores and also with EPS scores shifted by 1 day (EPS(d − 1)), i.e. with the scores of an EPS system characterized 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) is defined as:

$$\mathrm{RI}(SC) = \frac{SC(\mathrm{HEPS}) - SC(\mathrm{EPS})}{SC(\mathrm{EPS(d-1)}) - SC(\mathrm{EPS})}. \tag{1}$$

The RI is a normalized 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*.

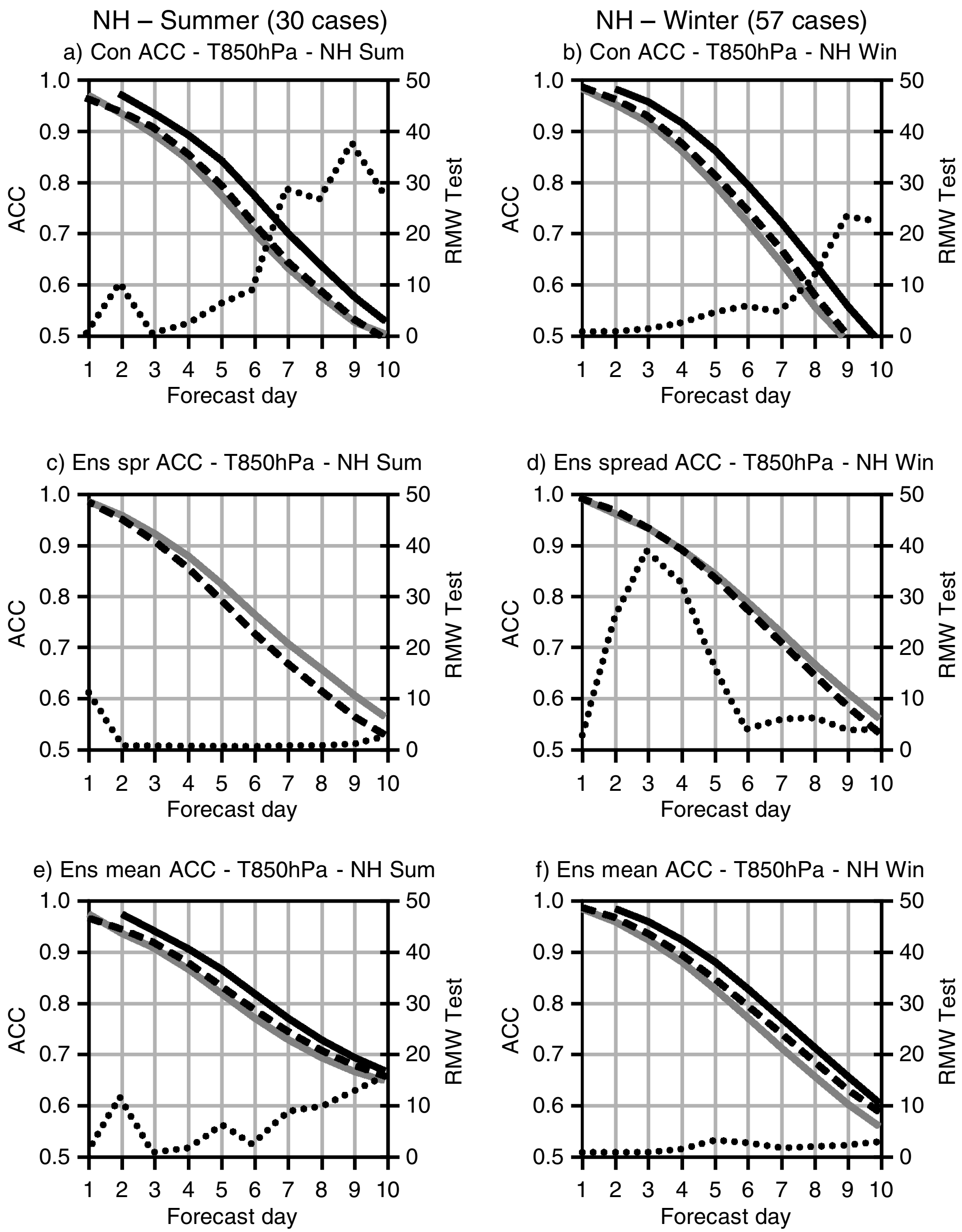

Figure 1. Mean anomaly correlation coefficients (ACCs) of 850 hPa temperatures over the northern hemisphere for: (a) the EPS (grey solid line), HEPS (dashed line) and EPS(d − 1) (black solid line) control forecasts (Con, left vertical axis) and Rank–Mann–Wilcoxon (RMW) test value (dotted line, right vertical axis) for summer; (b) as (a) but for winter; (c) and (d), as (a) and (b) but for the ensemble spread; (e) and (f), as (a) and (b) but for the ensemble-mean. See text for details.

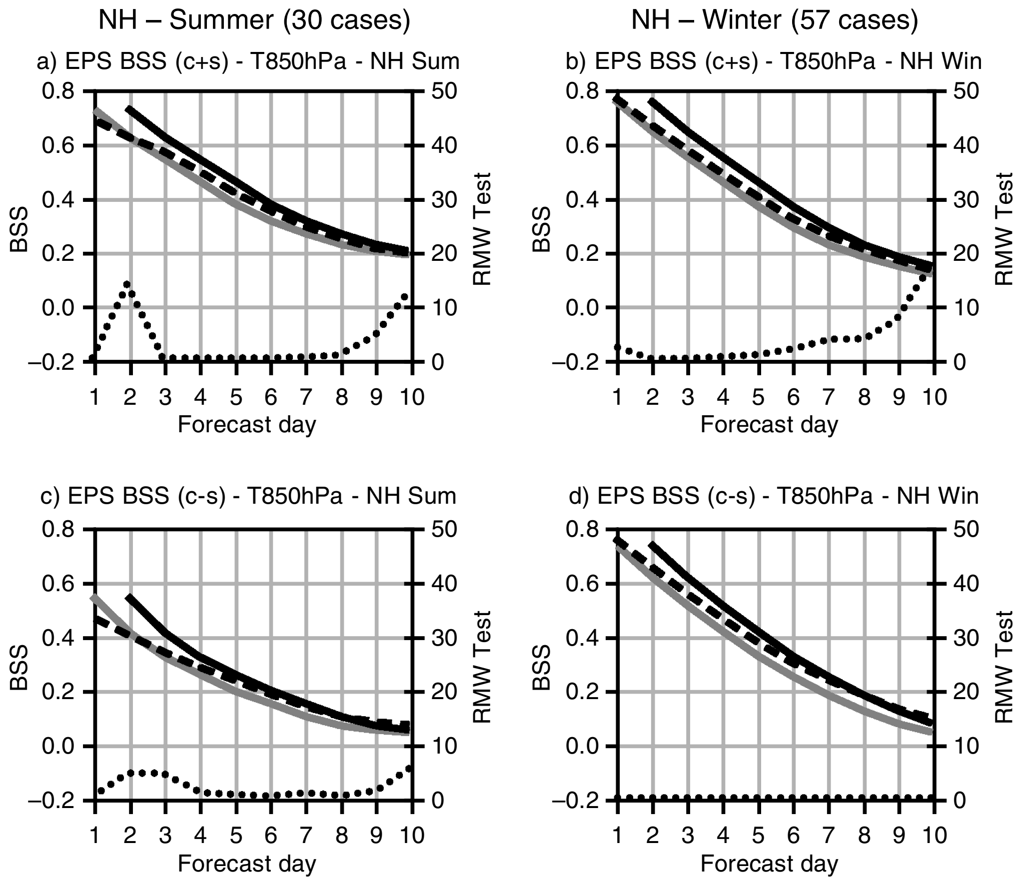

Figure 2. Briers Skill Scores (BSSs) over the northern hemisphere for: (a) the EPS (grey solid line), HEPS (dashed line) and EPS(d − 1) (black solid line) probabilistic predictions of the event '850 hPa temperature positive anomalies larger than one standard deviation' (left vertical axis) and Rank–Mann–Wilcoxon test value (dotted line, right vertical axis) for summer; (b) as (a) but for winter; (c) and (d) as (a) and (b) but for the event '850 hPa temperature negative anomalies larger than one standard deviation'. See text for details.

3. Impact of the resolution increase

(*a*) *Accuracy of control and ensemble-mean 850 hPa temperature forecasts, and ensemble spread*

Figure 1 shows the ACCs of the 850 hPa temperature for the EPS and HEPS control forecasts, the ensemble spread and 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. 1(a)) and up to day 8 for winter (Fig. 1(b)). 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 $T_L319L60$ analysis for verification (it should be remembered that the EPS unperturbed analysis is a T_L159 interpolation of the operational T_L319 analysis, while HEPS starts from a T_L255 interpolation of the T_L511 analysis). The HEPS spread is larger than the EPS spread especially for the summer period (Figs. 1(c) and (d)). The ACC of the HEPS ensemble-mean is higher for all but forecast day 1 (Figs. 1(e) and (f)), especially during winter.

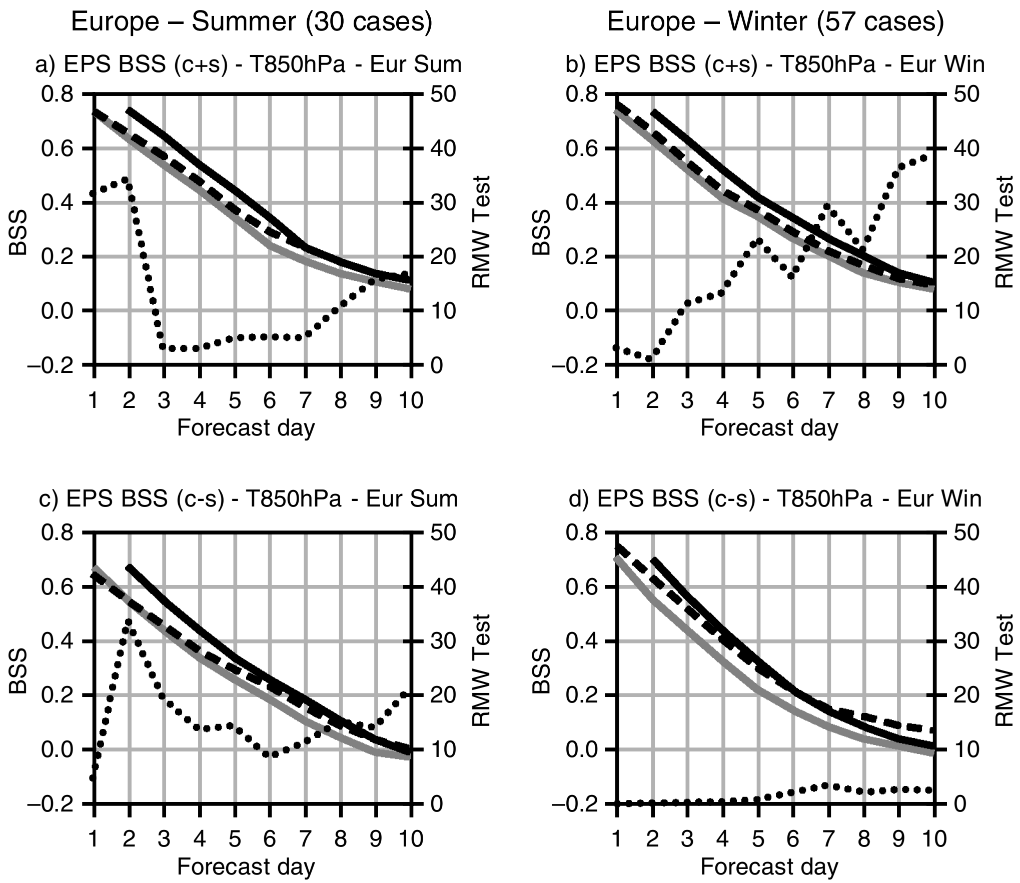

Figure 3. Same as Fig. 2 but BSSs over Europe.

The RMW test shows that the differences are statistically significant for all forecast times for winter (Fig. 1(f)) and up to forecast day 8 for summer (Fig. 1(e)).

(*b*) *Brier skill score (BSS) of 850 hPa temperature anomaly predictions*

The following two events have been considered: '850 hPa temperature positive anomalies larger than one standard deviation' and '850 hPa temperature negative anomalies larger than one standard deviation'. The accuracy of any probabilistic prediction of these two events has been assessed using the 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 descriptions of these measures). For reasons of space, only BSSs 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 (Figs. 2(a) and (c)) results indicate that the HEPS performs better, with significant differences between the EPS and the HEPS for all forecast ranges other than days 2 and 10. Similar results are shown for winter (Figs. 2(b) and (d)), with slightly larger positive differences significant for all forecast ranges but day 10.

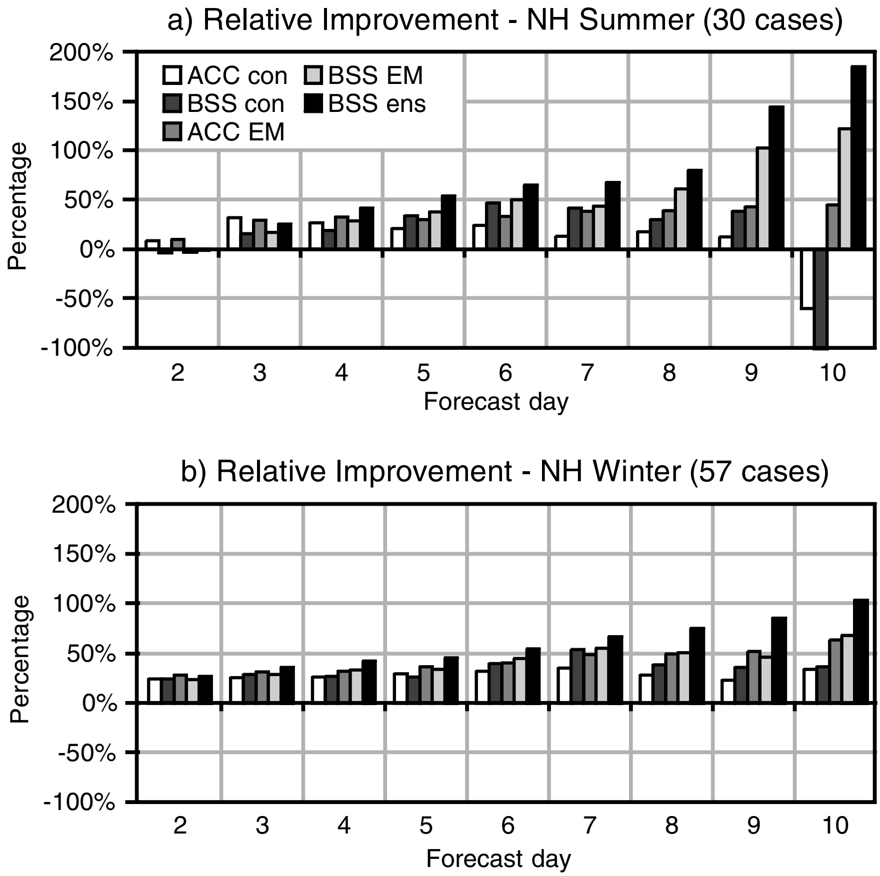

Figure 4. Relative improvement index (RI) for 850 hPa temperature computed over the northern hemisphere for: (a) control anomaly correlation coefficient (ACC, first bar, white), the control Brier skill score (BSS, second bar, grey), the ensemble-mean ACC (third bar, dotted), the ensemble-mean BSS (fourth bar, grey) and the EPS BSS (fifth bar, black), for summer. A RI of 100% indicates a gain in predictability of one day; (b) as (a) but for winter. See text for details.

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

(*c*) *Relative improvement index (RI) for 850 hPa temperature*

Figure 4 shows the RI computed over NH for five accuracy measures: control ACC and BSS, ensemble-mean ACC and BSS, and ensemble BSS. Results indicate that for summer (Fig. 4(a)) RIs are positive for all but forecast days 2 and 10, while for winter (Fig. 4(b)) all RIs are positive. The day 2 negative RIs shown for the control and the ensemble-mean are due to the fact that the $T_L319L60$ 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 days 5 to 7 forecast range, RI results show that the summer HEPS probabilistic predictions are 55–70% better than the EPS (Fig. 4(a)) and that the winter HEPS are 45–66% better than the EPS (Fig. 4(b)).

Comparing the RIs computed for the BSSs 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

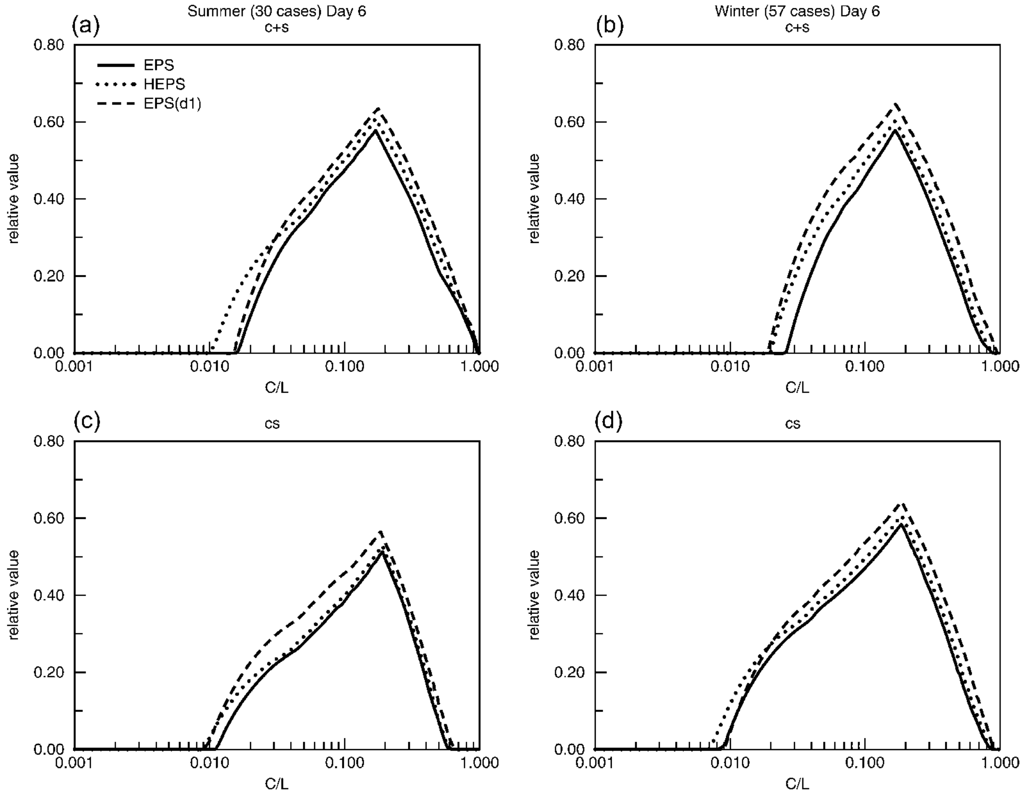

Figure 5. Value, V, of the EPS, HEPS and EPS(d − 1) ensemble configurations for: (a) the prediction of event '850 hPa temperature positive anomalies larger than one standard deviation' in summer; (b) as (a) but in winter; (c) and (d), as (a) and (b) but for the prediction of event '850 hPa temperature negative anomalies larger than one standard deviation'. See text for details.

particular, the EPS RIs are always larger than the control RIs, especially at the end of the forecast period. Considering, for example, 5-day forecasts during summer (Fig. 4(a)), 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.

(*d*) *Potential economic value of 850 hPa temperature forecasts*

The user-dependent benefit of a forecast system can be quantified using the value diagnostic (V) derived from a simple decision-making model, the cost–loss model (Murphy 1977; Liljas and Murphy 1994; 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

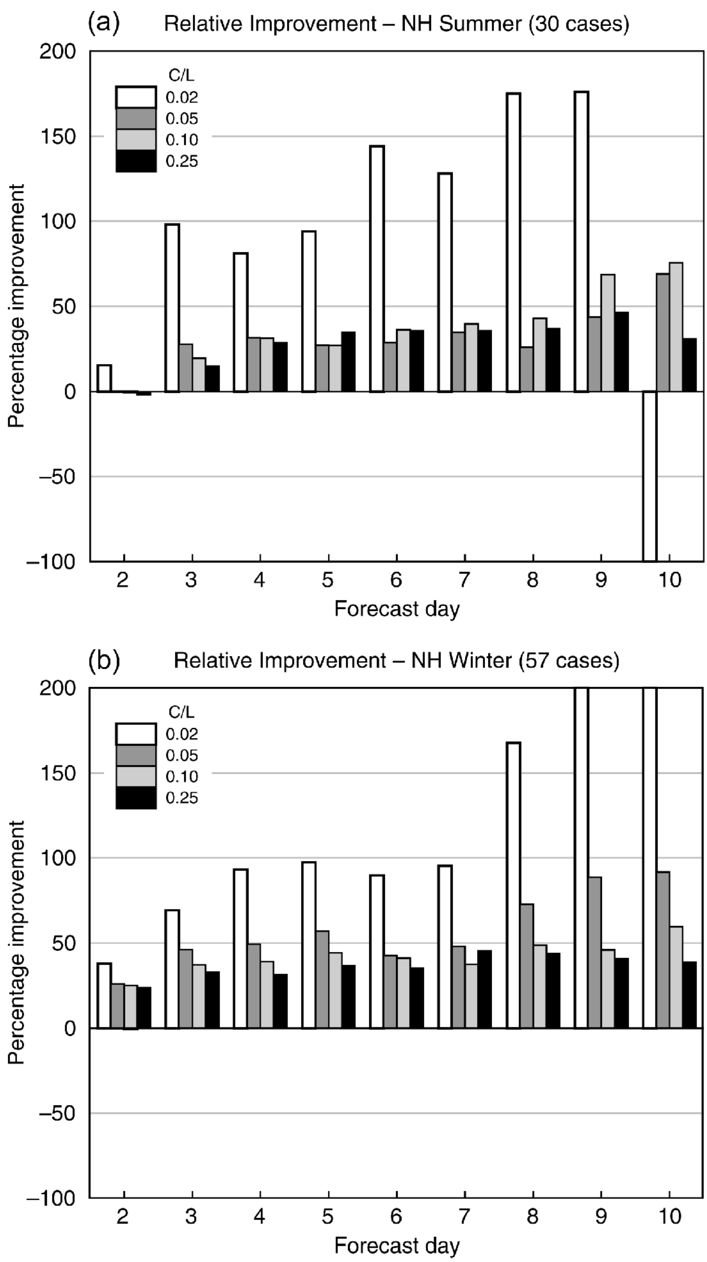

Figure 6. Value relative improvement index, RI(V), for the northern hemisphere in (a) summer and (b) winter, for selected cost–loss ratios: $C/L = 0.02$ (white), $C/L = 0.05$ (light grey), $C/L = 0.10$ (dark grey) and $C/L = 0.25$ (black).

TABLE 1. COST/LOSS DECISION MODEL

	Event occurs	Event does not occur
User protects	C	C
User does not protect	L	0

This shows the expense a user incurs if he spends C to protect against a possible loss L.

TABLE 2. ENSEMBLE CONFIGURATIONS

Horizontal prediction system	Resolution (km)	Number of members	Models	Parametrized PDF		Bias corrected
				PDF-Mean	PDF-std	
EPS	120	51	1	–	–	No
HEPS	80	51	1	–	–	No
MCEPS	60–120	5	5	–	–	No
HCG	80	1	1	$f_C - \mu_C$	σ_C	Yes
HEPS-bias	80	51	1	–	–	Yes
EPSG	120	51	1	$f_m - \mu_m$	σ_f	Yes
HEPSG	80	51	1	$f_{mH} - \mu_{mH}$	σ_H	Yes
MCEPSG	60–120	5	5	$f_{mMC} - \mu_{mMC}$	σ_{MC}	Yes

Symbols are as follows: f_C the HEPS control forecast with mean error μ_C and error variance σ_C; f_m the EPS ensemble-mean forecast with mean error μ_m and re-scaled error variance σ_f; f_{mH} the HEPS ensemble-mean forecast with mean error μ_{mH} and re-scaled error variance σ_H; f_{mMC} the MCEPS ensemble-mean forecast with mean error μ_{mMC} and error variance σ_{MC}.

C/Ls 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/Ls is not well known, but is likely to be concentrated towards low C/Ls (Roebber and Bosart 1996).

Figure 5 shows V at day 6 for the two events '850 hPa temperature positive anomalies larger than one standard deviation' and '850 hPa 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 C/L.

Figure 6 shows the RI for V, RI(V), calculated for a selection of C/Ls (0.02, 0.05, 0.10 and 0.25). The variation in benefit with different users seen in Fig. 5 applies 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 C/Ls, the RI is generally closer to 40%, similar to the RIs for the BSS (Fig. 4).

4. COMPARISON OF HEPS WITH A FIVE-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 forecasting 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. Here the HEPS is compared with a poor-man's ensemble defined following Ziehmann's (2000) approach, and in the next section the HEPS is compared with poor-man's ensembles defined following Atger (1999).

The poor-man's ensemble of Ziehmann (2000) is based on independent deterministic forecasts from different sources combined to generate a multi-centre ensemble. More

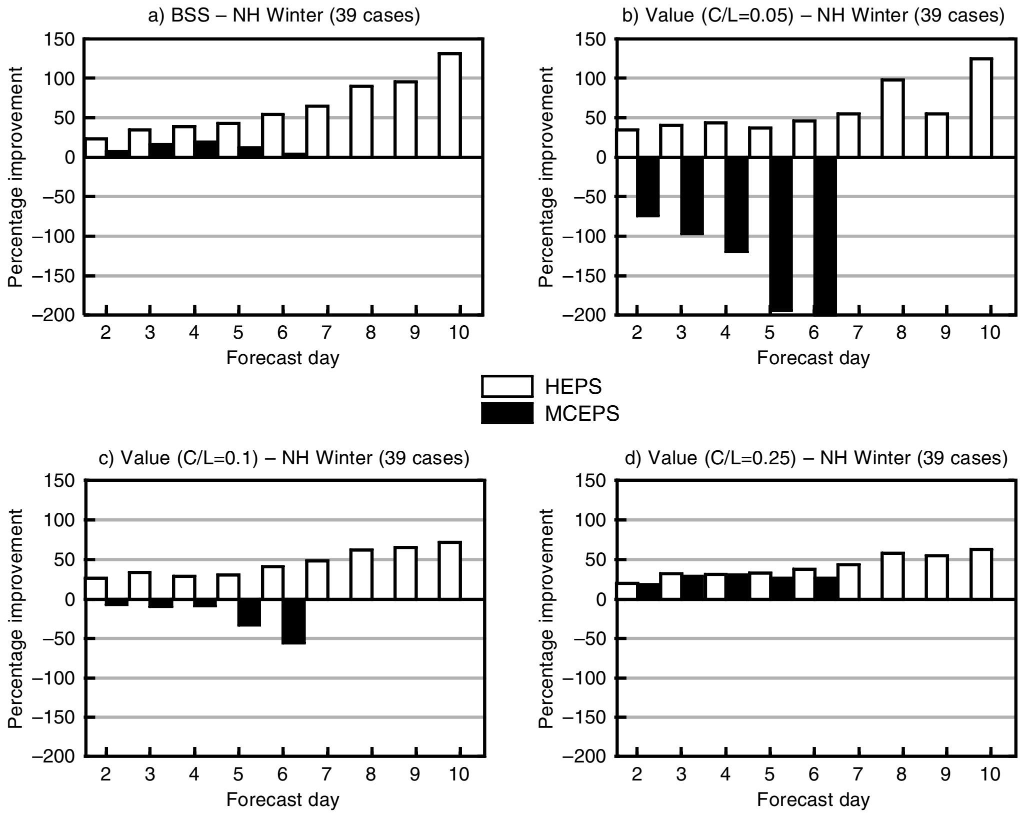

Figure 7. Relative improvement index (RI) for 500 hPa height over the northern hemisphere for winter cases. Bars show RI for HEPS (unfilled) and MCEPS (shaded), both relative to EPS, for (a) BSS, and value, V, for cost–loss ratios (b) $C/L = 0.05$, (c) $C/L = 0.1$ and (d) $C/L = 0.25$. See text for details.

specifically, the single deterministic high-resolution forecasts of ECMWF ($\sim$60 km resolution), the UK Meteorological Office (Met Office, $\sim$60 km resolution) and Deutscher Wetterdienst (DWD, $\sim$60 km resolution) are used, together with the lower-resolution control forecasts available for ECMWF ($\sim$120 km resolution) and the Met Office ($\sim$90 km resolution), to construct a five-member ensemble of independent forecasts. This will be referred to as the multi-centre EPS or MCEPS. Forecasts from MCEPS are available for only 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).

Figure 7 shows the RI 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 five 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, Fig. 7 shows that HEPS has the highest BSS and provides the greatest value for all users.

5. COMPARISON OF HEPS WITH PARAMETRIZED POOR-MAN'S ENSEMBLES

Atger's (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 μ_c and error variance σ_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 σ_c^2. The probability that the actual value will be above a given threshold T can be calculated by integrating the forecast PDF:

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

A probability forecast system based on the HEPS control forecast and using Gaussian error statistics will be referred to as HCG. Note that in the definition of this system, the mean error μ_c 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 RI 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. 8(a)) and V (Figs. 8(b)–(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 HEPS-bias at days two and 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 probabilities from the parametrized 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 HEPS-bias.

A similar approach could be followed but based on the ensemble mean and with a PDF parametrized 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 underestimation of ensemble spread. The greatest benefits of this approach may

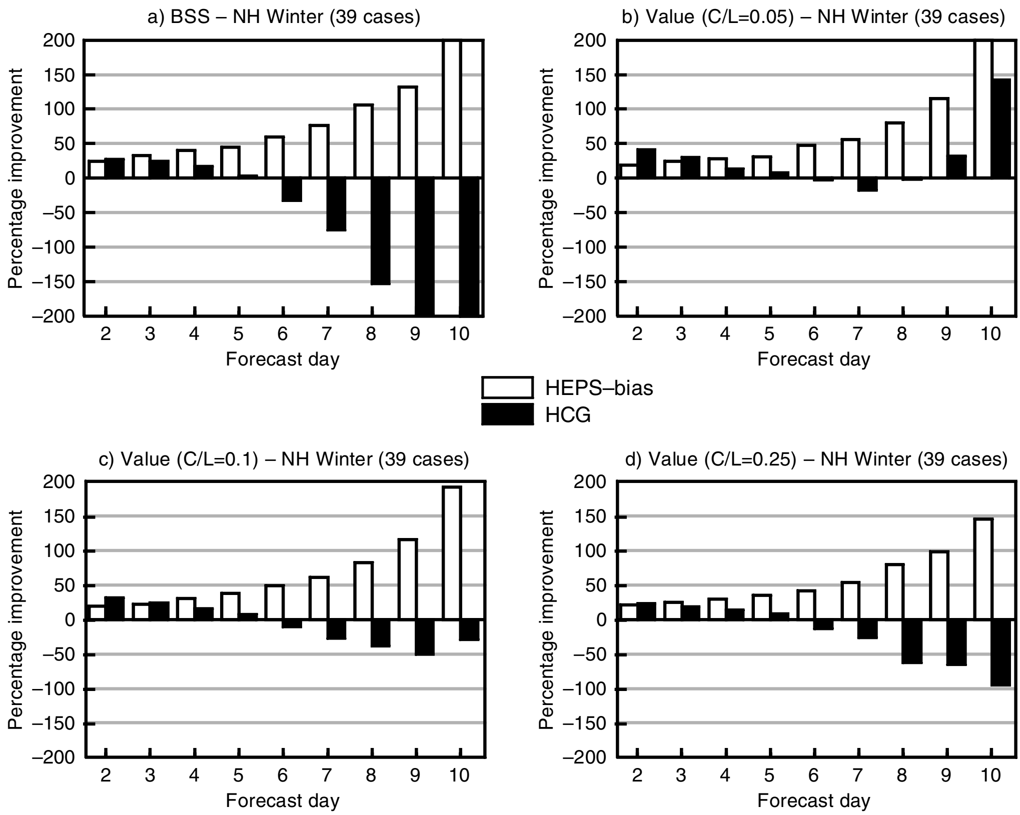

Figure 8. Relative improvement index (RI) for 500 hPa height over the northern hemisphere for winter cases, computed using HEPS as reference. Bars show RI for HEPS-bias (unfilled) and HCG (shaded), for (a) BSS, and value, V, for cost–loss ratios (b) $C/L = 0.05$, (c) $C/L = 0.1$ and (d) $C/L = 0.25$. See text for details.

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 parametrized PDF (here the ensemble members are assumed to be distributed normally about the ensemble mean). Fitting a parametrized 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 μ_{m} and σ_{m}^2 respectively and let the average spread over a large number of cases be $\langle s^2 \rangle$. Then a parametrized forecast PDF can be constructed as Gaussian with mean $f_{\mathrm{m}} - \mu_{\mathrm{m}}$ and variance σ_{f}^2 where

$$\sigma_f^2 = \frac{s^2}{\langle s^2 \rangle} \sigma_{\mathrm{m}}^2. \tag{3}$$

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.

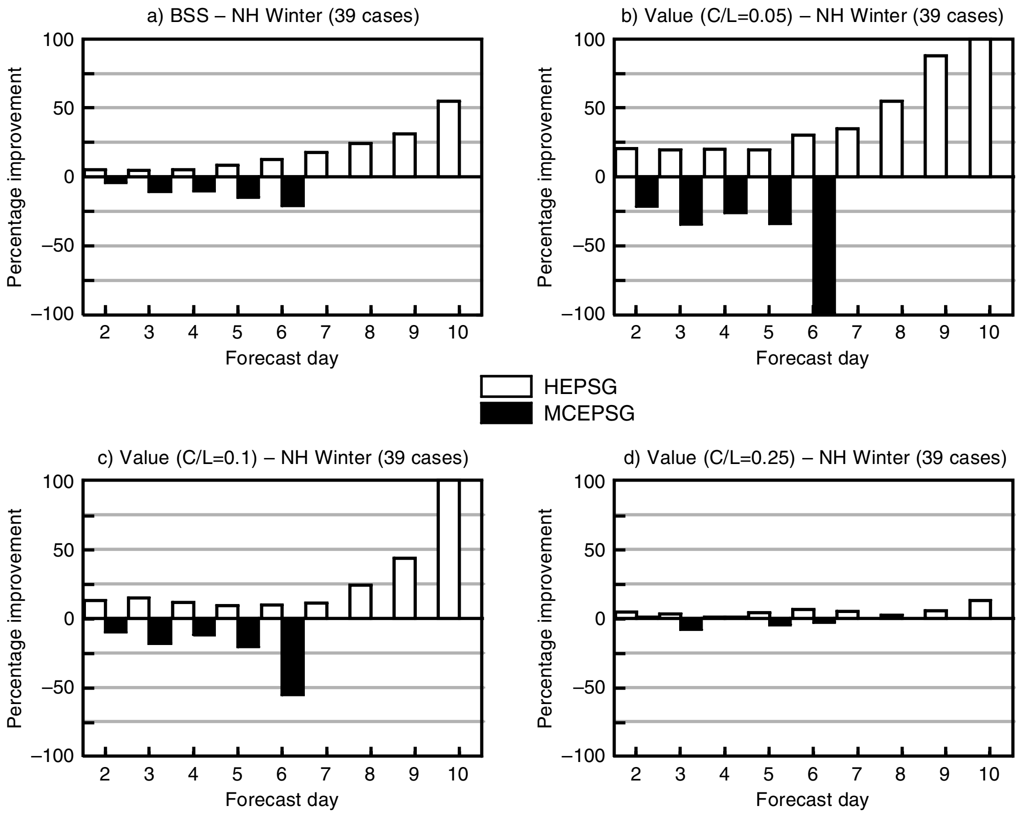

Figure 9. Relative improvement index (RI) for 500 hPa height over the northern hemisphere for winter cases, computed using HEPS-bias (the bias-corrected HEPS) as reference. Bars show RI for parametrized ensembles HEPSG (unfilled) and MCEPSG (shaded), for (a) BSS, and value (V) for cost–loss ratios (b) $C/L = 0.05$, (c) $C/L = 0.1$ and (d) $C/L = 0.25$. See text for details.

Parametrized 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 HEPS-bias probabilities to explore the additional benefit to be obtained from the parametrization.

Figure 9 shows the relative improvement for MCEPSG and HEPSG computed using HEPS-bias as reference. The effect of the parametrization on the HEPS forecasts is positive throughout the forecast, although substantial improvements are found mostly later in the forecast and for smaller C/Ls. Although the parametrization 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 emphasized that the parametrization 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 2003).

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 through the forecast range. The parametric approach of the previous section is used to investigate these differences, including the effects of model bias and spread underestimation.

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 (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 V 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).

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 idealized, perfectly specified EPS would consist of an effectively infinite number of forecasts, all equally likely and together representing the full uncertainty of 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, BSS_{C}, 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):

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

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 Fig. 11. This shows the actual BSS for the EPS as a fraction of BSS_M^{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.

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 BSS_{C} for EPS and HEPS, the expected improvement for ensemble probability forecasts is $\{BSS_M^{\mathrm{perf}}(\mathrm{HEPS}) - BSS_M^{\mathrm{perf}}(\mathrm{EPS})\}$. Figure 12 compares this expected gain

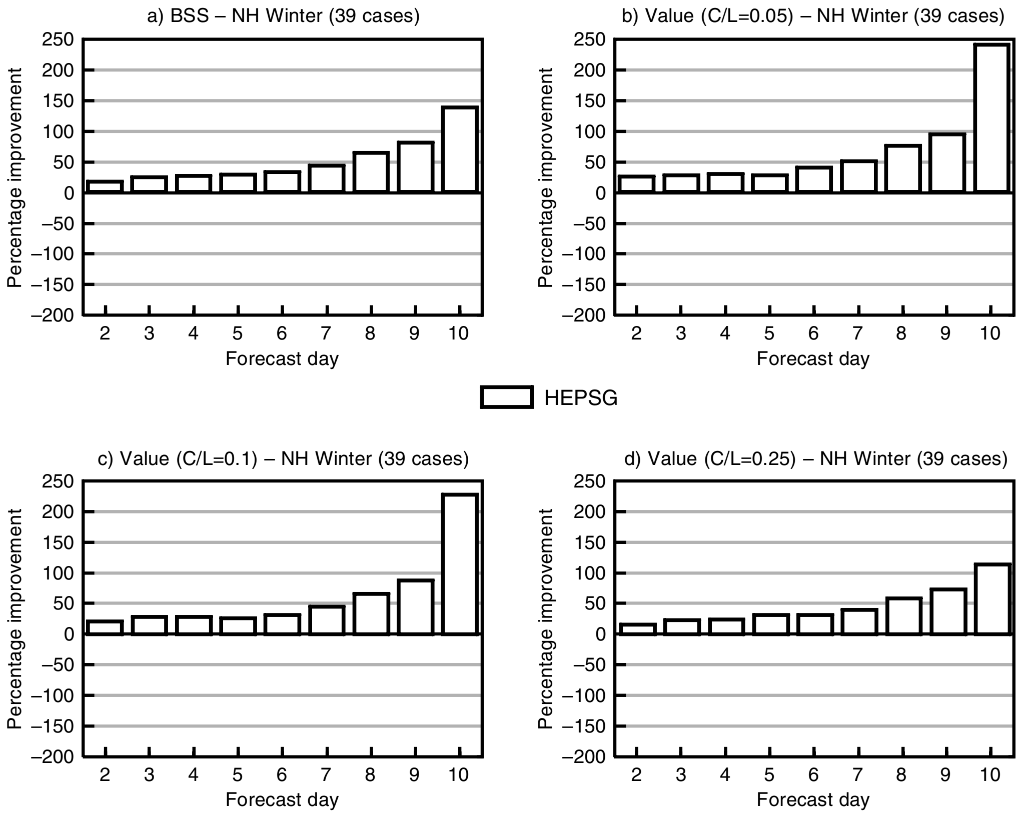

Figure 10. Relative improvement index (RI) for 500 hPa height over the northern hemisphere for winter cases, relative to EPSG (the parametrized EPS). Bars show RI for parametrized ensembles HEPSG, for (a) BSS, and value (V) for cost–loss ratios (b) $C/L = 0.05$, (c) $C/L = 0.1$ and (d) $C/L = 0.25$. See text for details.

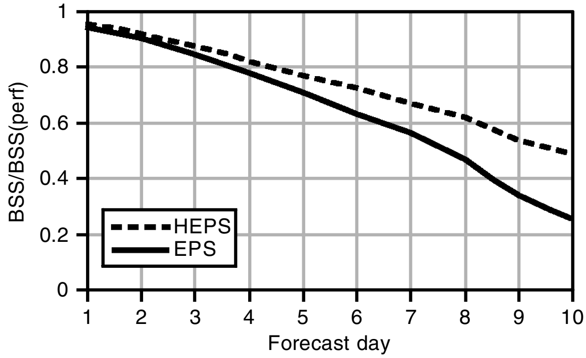

Figure 11. BSS as a fraction of the expected score for a perfect ensemble for 500 hPa height over the northern hemisphere for 39 winter cases. Solid line: EPS; dashed line: HEPS. See text for details.

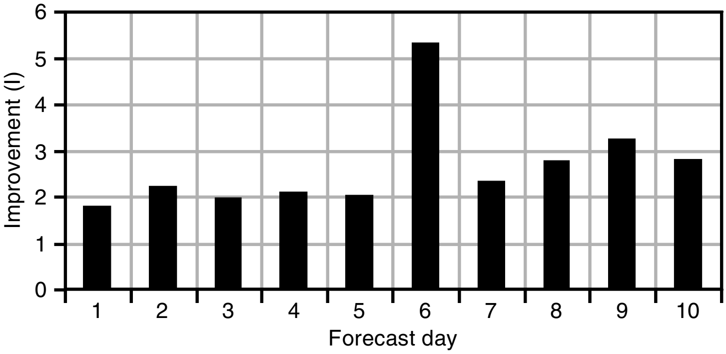

Figure 12. Actual increase in BSS for HEPS over EPS, expressed as a fraction of the increase expected from the deterministic improvement. Results for 500 hPa height over the northern hemisphere for 39 winter cases. See text for details.

with the actual improvement in BSS for the ensemble using the ratio I where:

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

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.

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 it gives 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 the argument of Toth *et al.* (2002) 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 PDF of 850 hPa temperature and 500 hPa geopotential height forecast states than the EPS. Average results (over 57 winter and 30 summer cases) based on BSS of probabilistic predictions of moderate 850hPa temperature anomalies for NH have indicated that the operational implementation of the new HEPS system has resulted in gains 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 C/Ls. 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, Met Office and DWD). Following Ziehmann (2000) a five-member multi-centre poor-man's ensemble has been considered (MCEPS). Average results (over 39 winter cases) 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 for the better performance. Then, following Atger (1999), an ensemble based on a parametrized distribution function centred on the ECMWF high-resolution forecast, with standard deviation defined by the control error standard deviation, has been considered (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 PDF 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 parametrized 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 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. Parametrized 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 21 November 2000, HEPS became the ECMWF operational ensemble configuration.

ACKNOWLEDGEMENTS

Graham Holt is acknowledged for his essential technical support and optimal scheduling of the considerable ensemble experimentation that made this study possible, Tony Hollingsworth for his very valuable comments to a first version of this manuscript, and Rob Hine and Els Kooij-Connally 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. Weather 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. Meteorol. Soc.*, **125,** 2333–2351

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

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

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. Weather Rev.*, **99,** 851–872

1971b On stochastic dynamic prediction. II: Predictability and utility. *Mon. Weather 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,** 1012–1037

Hoffschildt, M., Bidlot, J.-R., Hansen, B. and Janssen, P. A. E. M. 1999 'Potential benefit of ensemble forecasts for ships'. 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. Weather 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. Am. Meteorol. Soc.*, **75,** 1227–1236

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

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'. ECMWF Newsletter No 74–Winter 1996–97. ECMWF, Shinfield Park, Reading, RG6 9AX, UK

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

Palmer, T. N., Molteni, F., Mureau, R., Buizza, R., Chapelet, P. and Tribbia, J. 1993 'Ensemble prediction'. In 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

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

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, Geneva, Switzerland

Taylor, J. and Buizza, R. 2003 Energy demand prediction using the ECMWF Ensemble Prediction System. *Int. J. Forecasting,* in press

Toth, Z., Zhu, Y., Szunyogh, I., Iredell, M. and Wobus, R. 2002 'Does increased model resolution enhance predictability?'. Paper J1.8 in Proceedings of the AMS symposium on observations, data assimilation and probabilistic prediction, 13–17 January 2002, Orlando, Florida. American Meteorological Society, Boston, USA (available from http://sgi62.wwb.noaa.gov:8080/e ns/ens_info.html)

Wilks, D. S. 1995 *Statistical methods in the atmospheric sciences*. Academic Press, San Diego, USA

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