TECHNICAL MEMORANDUM

Scale-dependent verification of ensemble forecasts

Thomas Jung and Martin Leutbecher

Research Department

ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

Abstract

A scale-dependent verification of the ECMWF ensemble prediction system (EPS) in the Northern Hemisphere is presented. The relationship between spread and skill is investigated alongside probabilistic forecast skill for planetary, synoptic and sub-synoptic spectral bands. The filtering is accomplished using total and zonal wavenumber filters. Diagnosed overdispersiveness of ECMWF EPS in the short-range is primarily due to excessive amounts of spread on synoptic scales. Diagnosed underdispersiveness of the ensemble beyond day 5 of the forecast, can be explained by too little spread on both synoptic and planetary scales.

1 Introduction

Forecast verification is the process of evaluating the quality of forecasts (Wilks 2006). Traditionally, most of the verification has been carried out for *unfiltered* forecast fields of meteorological relevance such as geopotential height. The fact that mostly unfiltered fields are used for verification implies that all scales of atmospheric motion are simultaneously taken into account. A root mean square error (RMSE) computed for unfiltered geopotential height fields, for example, contains contributions from planetary, synoptic and sub-synoptic scales. Given that there is a scale-dependence of the relative importance of various physical processes (e.g., Charney 1948) it seems promising to carry out forecast verification as a function of spatial scale.

The relatively small number of existing studies dealing with a scale-dependent forecast verification are primarily based on single *deterministic* forecasts. A spectral analysis of deterministic predictability and error of forecasts from the Canadian Meteorological Centre was presented by Boer (1984), who showed that high wavenumbers are dominated by transient error components; at lower wavenumbers, however, the stationary error component becomes much more prominent (see also Mureau 1990). Furthermore, Boer (2003) showed the 'classical predictablity behaviour' with forecast skill going to zero immediately at small scales and with loss of predictive skill penetrating to larger spatial scales with forecast time. The way how the dependence of forecast error on horizontal scale of deterministic forecasts with the ECMWF model have changed from 1980 to 2004 has been described by Simmons (2006). Simmons shows, for example, that 500 hPa geopotential height (Z500, hereafter) short-range forecast error has been substantially reduced over time, particularly at the large synoptic scales (band of wavenumbers centred around total wavenumber nine). As a result, short-range forecast error is more evenly distributed across the spectrum for more recent years.

While there has been certainly some work on scale-dependent verification of *deterministic* forecasts (see previous paragraph), to our knowledge little work has been done on scale-dependent verification of ensemble forecasts. By applying the methodology described in this paper, Jung and Leutbecher (2007) pointed out that there is substantial *probabilistic* forecast skill well into the medium-range of synoptic and sub-synoptic Z500 features over the Arctic; this is particularly true after November 2000 when the high-resolution ECMWF EPS was introduced. In another study, Casati and Wilson (2007) have introduced a spatial-scale decomposition of the Brier score using a wavelet approach in order to allow for a scale-dependent, diagnostic verification of probabilistic forecasts. Casati and Wilson illustrate their method using lightning probability forecasts over North America, which they show to be skilful only on scales larger than 700 km (e.g., fronts).

The aim of this study is two-fold: Firstly, a method is proposed which can be used to carry out a scale-dependent verification of ensemble forecast. The method basically works by spectrally filtering the data such as to retain only a specific range of spatial scales; skill scores and other diagnostics are then computed for the spectrally filtered fields. Secondly, the new method is applied to assess the performance of the ECMWF EPS in a scale-dependent framework.

An overview of the ECMWF EPS is briefly given in section 2. This is followed by a description of the filtering technique along with the scale-dependent climatology used for probabilistic forecast verification in section 3.

The results are presented in section 4. Firstly, the scale-dependent relationship between spread and skill is studied. This is followed by a discussion of the probabilistic forecast skill of the EPS on planetary, synoptic and sub-synoptic scales, both for the recent winter of 2006/07 and in terms of the period 1994–2007. Finally, the main conclusions of this study are discussed in section 5.

2   Data

The scale-dependent verification is applied to the operational ECMWF EPS forecasts of the 500 hPa geopotential and the 850 hPa meridional wind component for the winter season December 2006 to February 2007, referred to as DJF07. The operational forecasts are issued twice daily at 00 UTC and 12 UTC. The results presented here use forecasts started at 12 UTC. Due to the temporal correlation of forecast errors, it seems appropriate to base results on the 12 UTC runs only.

The EPS configuration that is currently operational and was operational in DJF07 consists of 50 perturbed forecasts and one unperturbed forecast. The horizontal resolution is variable (Buizza et al. 2007). Up to a lead time of 10 days, the model is triangularly truncated at wavenumber 399; this corresponds to a grid spacing of about 50 km. For lead times greater 10 days up to the forecast range of 15 days, the truncation is at wavenumber 255 (equivalent to a grid spacing of 80 km). In the vertical, the model is discretized with 62 levels from the surface to the model top at 5 hPa.

Initial uncertainty is represented using the leading singular vectors of the forecast model's tangent-linear propagator. The resolution for the singular vectors is T42 and the perturbations are optimised over a 48-hour interval. The extra-tropical perturbations in each hemisphere are based on the leading 50 singular vectors computed with an adiabatic version of the tangent-linear model (apart from vertical mixing and surface friction). In addition, the initial conditions are perturbed in the vicinity of tropical cyclones using the leading five singular vectors computed with a diabatic version of the tangent-linear model (Puri et al. 2001). The initial condition perturbations are obtained from the singular vectors through a Gaussian sampling methodology, i.e. each initial perturbation is a linear combination of the singular vectors with coefficients drawn from a Gaussian distribution. Further details of the methodology to represent initial uncertainty are described by Leutbecher and Palmer (2007).

Model uncertainty is represented in the EPS with a scheme which perturbs the tendencies due to parameterized processes ("stochastic physics", Buizza et al. 1999). The net tendencies are multiplied with a random number drawn from a uniform distribution between 0.5 and 1.5. The random numbers are kept constant for a tile of $10° \times 10°$ and over 6 model time steps to introduce some temporal and spatial coherence in the forcing.

In addition to the detailed analysis of the DJF07 winter, the ensemble verification statistics has been computed for the different wavebands since daily operational runs began in May 1994. The major changes of the EPS are listed here for reference:

May94: Operational daily runs of the EPS start, T63 resolution, 32 perturbed forecasts using 16 extra-tropical singular vectors.
Mar96: Southern Hemisphere singular vectors
Dec96: T_L159 resolution, 31 levels, 50 perturbed forecasts using 25 extra-tropical singular vectors
Mar98: Evolved singular vectors
Oct98: Stochastic physics
Oct99: 40 levels
Nov00: T_L255 resolution
Jan02: singular vectors for tropical cyclones

Jun05: Use of 50 extra-tropical singular vectors
Feb06: $T_L 399$ resolution, 62 levels
Nov06: Variable resolution ($T_L 255$ from 10 to 15 days)
Nov07: Revised model physics and initial perturbation amplitude.

The latest change, is examined in more detail in section 5 by looking at a clean comparison of the operational EPS and the experimental EPS version for 69 cases in the period June–September 2007.

3 Methodology

3.1 Spectral filtering

As mentioned in the Introduction, ensemble forecasts will be evaluated separately for three different spectral bands: (i) planetary, (ii) synoptic and (iii) sub-synoptic waves. The breakdown into planetary and synoptic-scale waves has a long tradition in the atmospheric sciences (e.g, Blackmon 1976, Blackmon et al. 1977, Gulev et al. 2002). Most previous studies have used temporal filters[1] for the breakdown (e.g, Blackmon et al. 1977, Gulev et al. 2002). In this study, two different kinds of *spatial* filters are used. The advantage of spatial filters in the context of this study is two-fold: firstly, the ECMWF model is a spectral model and therefore the spectral (spatial) filtering is very straightforward; secondly, the time series are relatively short (10–15 day forecasts), which makes it difficult to use effective temporal filters which require relative large number of weights and hence relatively long time series (see von Storch and Zwiers 1999).

In this study two different spectral filters are used, that is, total (N) and zonal (M) wavenumber filters. For the former (latter) the actual filtering is accomplished by retaining only total (zonal) wavenumber

- N=0–7 (M=0–3) for planetary-scale,

- N=8–21 (M=4–14) for synoptic-scale, and

- $N \geq 22$ ($M \geq 15$) for sub-synoptic-scale phenomena.

Of course, the actual choice of the wavenumbers used for the breakdown is somewhat arbitrary given that there is no clear gap in the spectrum of meso-scale to planetary-scale atmospheric motions in the extratropics. Previous studies have shown, however, that generally the results (i.e. the statistics of the filtered fields) are not very sensitive to the exact choice of the wavenumber ranges (Blackmon 1976).

The total wavenumber filter, which is isotropic, selects the same spatial scales whatever the latitude. This has the advantage that the same filters can be used to characterize, for example, synoptic varations in mid-latitudes and polar regions (keeping in mind, though, that the Rossby radius of deformation decreases towards higher latitudes). The isotropy of the total wavenumber filter means, however, that fields that are filtered to retain planetary-scale variations may also contain contributions from remote regions such as the tropics. Therefore, the fact that the ECMWF EPS is underdispersive in the tropics (on all scales) affects the ensemble spread on 'planetary scales' in the extratropics when a total wavenumber filter is employed. To allow for a more meaningful investigation of the Northern Hemisphere mid-latitudes only, particularly on large spatial scales, thus, a set of zonal wavenumber filters has also been employed (see above). This approach allows a precise separation of latitudinal variations for each wavenumber band as the spherical harmonics are a product of an

---

[1]This might partly be due to the fact that temporal filters are easier to implement.
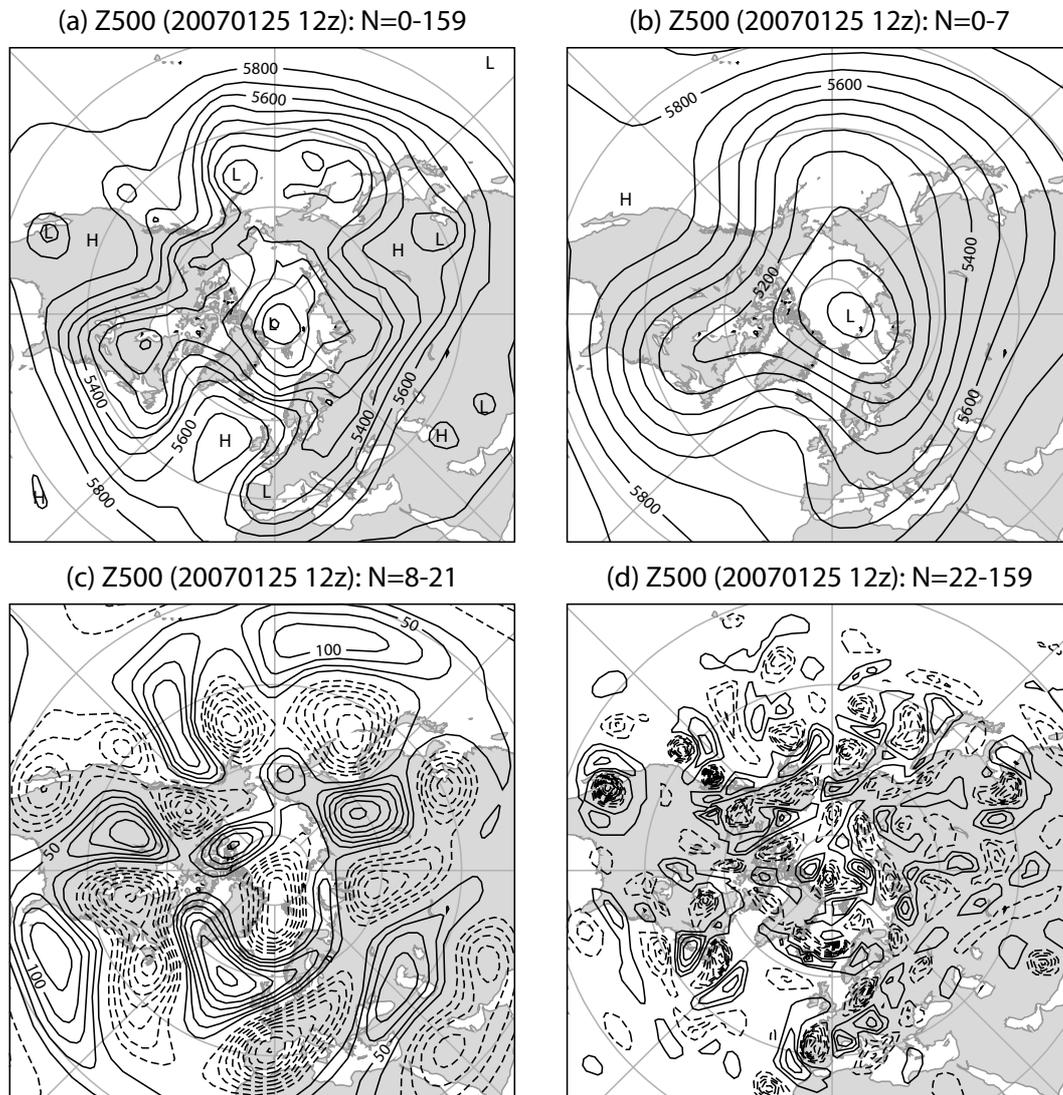
Figure 1: 500hPa geopotential height field (in m) at 12UTC on 25 January 2007 from the operational ECMWF analysis for various spectral triangular filters: (a) N = 0 − 159 (unfiltered, contour interval 100 m), (b) N = 0 − 7 (planetary waves, contour interval 50 m), (c) N = 8 − 21 (synoptic variability, contour interval 25 m) and (d) N ≥ 22 (sub-synoptic variability, contour interval 10 m).

associated Legendre function, which depends on latitude and total wavenumber, and an exponential function, which depends on zonal wavenumber and longitude. The actual filters were subjectively chosen such as to be most efficient in the band from $35°$–$65°$N.

In order to illustrate the characteristics of the spectral filters, the 500 hPa geopotential height field at 12 UTC on 15 January 2007 obtained from the operational ECMWF analysis has been spectrally decomposed into planetary, synoptic and sub-synoptic scales using the total wavenumber filter (Figure 1). The unfiltered field shows a pronounced blocking anti-cyclone over the central North Atlantic (Figure 1a); low-pressure systems can be found, south of the Aleutian Islands, over Newfoundland, and in the Nordic Seas; and a cut-off low is located just to the east of Baja California. The planetary-scale filter (Figure 1b) is efficiently capturing the large-scale structure of the atmospheric flow including the large-amplitude planetary-wave structure in the North Atlantic into which the blocking anti-cyclone is embedded. The synoptic-scale total wavenumber filter

(a) Z500 (20070125 12z): M=0-159   (b) Z500 (20070125 12z): M=0-3

(c) Z500 (20070125 12z): M=4-14   (d) Z500 (20070125 12z): M=15-159
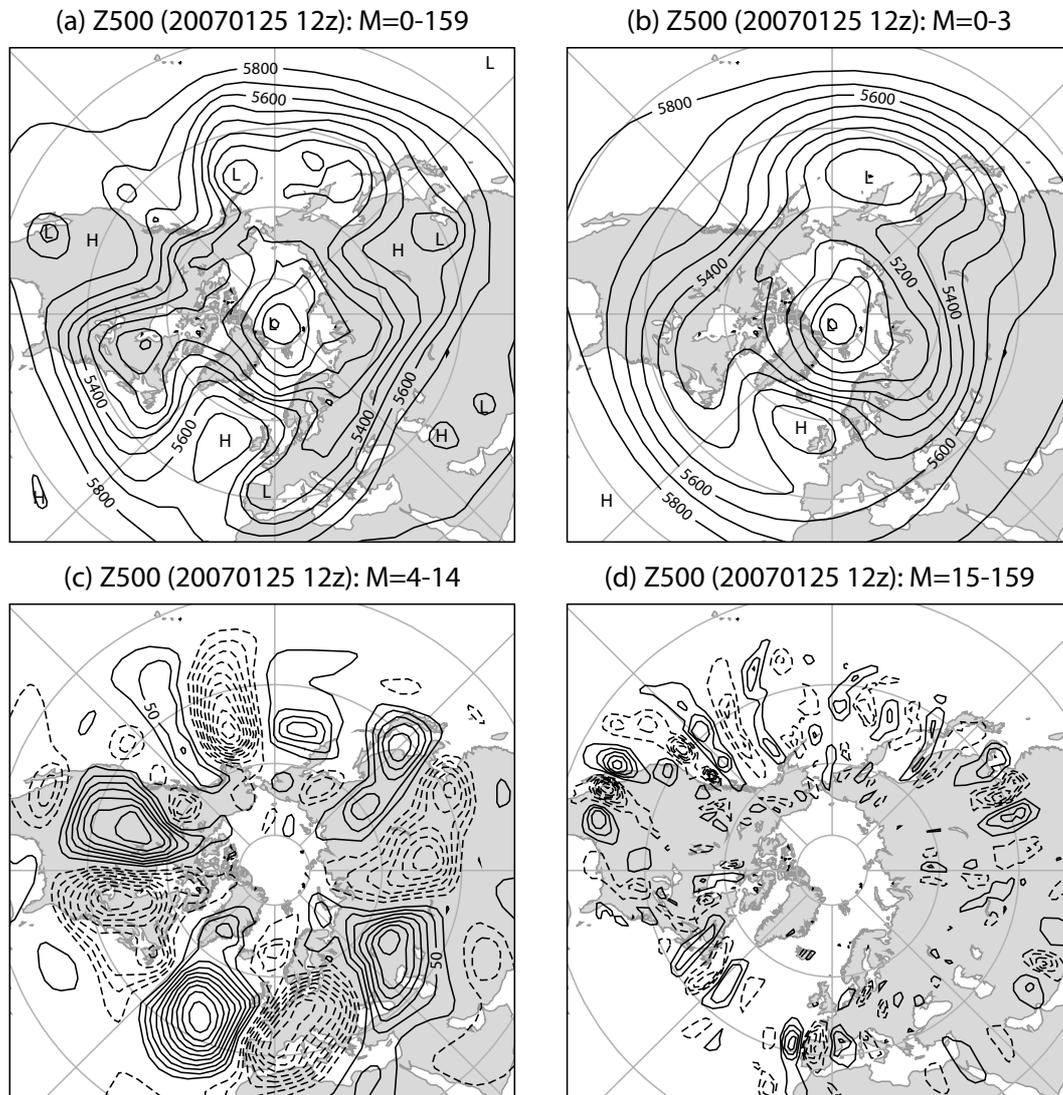
Figure 2: As in Fig. 1 but for various zonal spectra bands: (a) $M = 0 - 159$ (unfiltered, contour interval 100 m), (b) $M = 0 - 3$ (planetary waves, contour interval 50 m), (c) $M = 4 - 14$ (synoptic variability, contour interval 25 m) and (d) $M \geq 15$ (sub-synoptic variability, contour interval 10 m).

captures both, finer-scale features of the troughs and ridges such as for the North Atlantic blocking as well as synoptic features such as the the above-mentioned low-pressure systems (Figure 1c). The sub-synoptic total wavenumber filter, finally, captures very fine-scale features associated with synoptic-scale systems such as the cut-off low east of Baja California or the low-pressure system close to Newfoundland (Figure 1d).

The performance of the zonal wavenumber filter on 25 January 2007 is illustrated in Figure 2. As mentioned above the choice of the zonal wavenumber bands has been made having the mid-latitudes in mind. In fact, in the mid-latitudes the zonal wavenumber filter efficiently separates the different spatial scales. In other latitude bands the breakdown is less meaningful; over the Arctic, for example, meteorological phenomena of almost all scales appear in the M=0–3 band ('planetary scales').

3.2   Climatology

Some verification measures depend on a climatology. The anomaly correlation coefficient used extensively in the verification of deterministic forecasts considers anomalies with respect to a climatological mean. Many probabilistic scores evaluate the predicted probability for events such as an anomaly larger than one climatological standard deviation. In addition, probabilistic skill scores require as reference the score of the climatological distribution. Such skill scores are helpful to decide whether a forecast contains useful information beyond the climatological distribution. A good estimate of the climatological distribution is important in order to avoid "false skill" (Hamill and Juras, 2006). Here, we compute climatological statistics which depends on location and the day of the year. The climatology consists of daily fields of the mean, standard deviation of anomalies, and quantiles of anomalies. The climatology is based on ERA-40 analyses, which are expected to provide the most accurate available, consistent, long-term description of the atmosphere. In order to obtain good estimates globally, the climate is based on the 23 years 1979–2001 during which satellite data constrain the analysis well in the southern hemisphere (Uppala et al. 2005). For each day of the year, statistics are based on a 61-day window centred on the day of interest. The statistics are computed with weights which are maximum at the window centre and gradually decrease to zero at ± 30 days. Thus, $23 \times 61 = 1403$ dates contribute to the climate statistics of one day. Variable weights are superior to constant weights in terms of resolving the annual cycle and in filtering high-frequency sampling uncertainty. The climate of a particular wavenumber band is computed from the filtered analyses. For the climatological mean, one could alternatively compute the mean of the full fields and then apply the various filters. However, for the nonlinear statistics, such as standard deviation and quantiles, the computation of the statistic and the waveband filter are not commutative. Further details of the computation are given in the Appendix.

The climatology of the standard deviation of differently filtered Z500 fields based on ERA-40 is shown in Figure 3. In agreement with previous studies on low-frequency intraseasonal variability (or so-called 'weather regimes') (e.g., Blackmon et al. 1977, Gulev et al. 2002), enhanced planetary wave activity can be found at the exit regions of the North Atlantic and North Pacific storm tracks. Moreover, the high-latitude maxima coincide with the northern centres of action of two of the most prominent teleconnection patterns of the Northern Hemisphere: the North Atlantic Oscillation and the Pacific North America pattern (e.g, Wallace and Gutzler 1981). The synoptic wavenumber filters (Figure 3b,d) are very efficient in capturing the Z500 variability associated with the passage of synoptic systems within the two main storm tracks of the Northern Hemisphere (Blackmon et al. 1977, Gulev et al. 2002). Finally, the distribution of sub-synoptic Z500 variability is similar to its 'synoptic counterpart', except that the degree of 'localization' in the main storm tracks is less pronounced (see also Ayrault et al. 1995, Gulev et al. 2002).
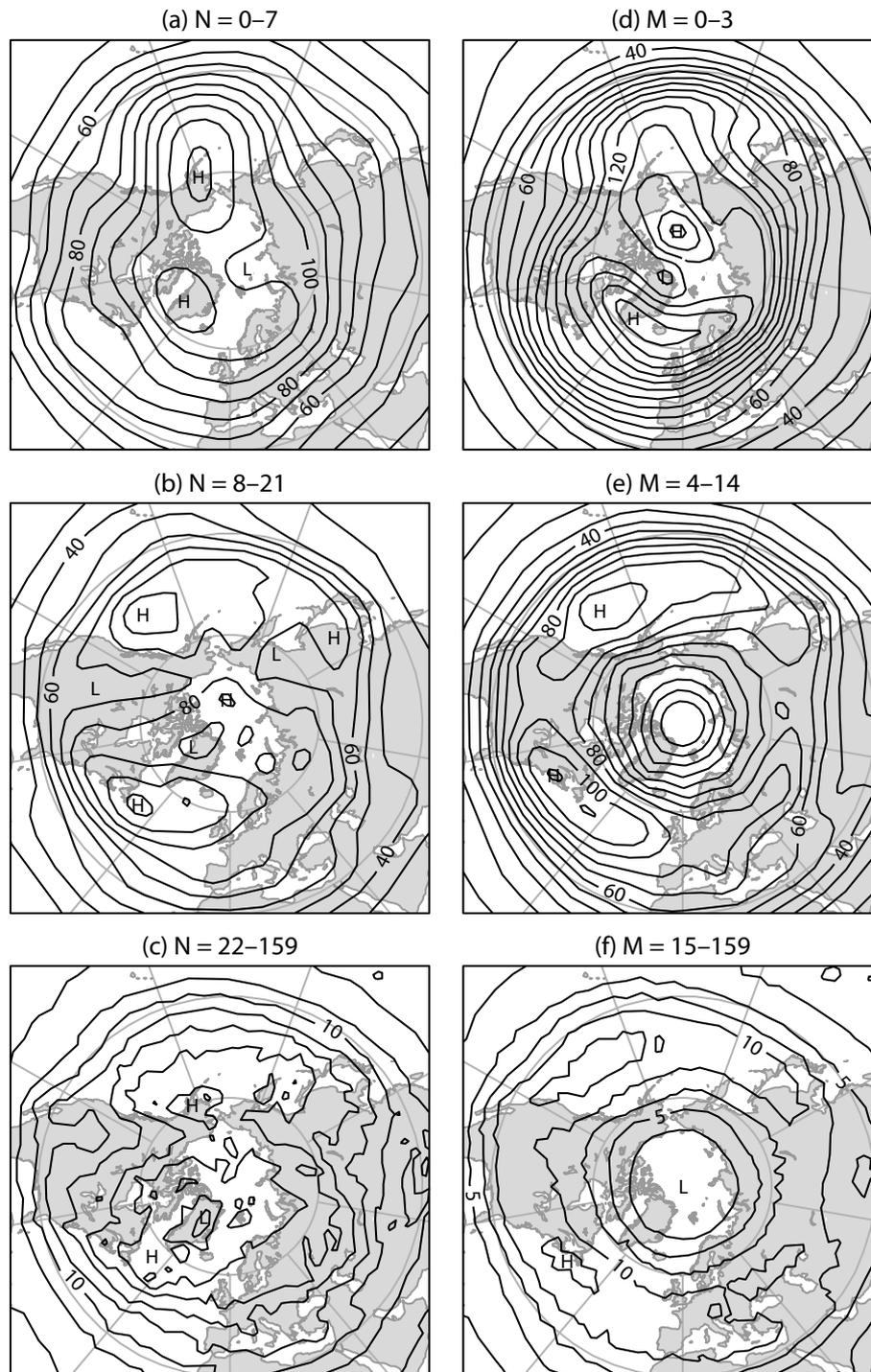
Figure 3: Climatology of the standard deviation of differently filtered 500hPa geopotential height fields (in m) based on ERA-40 data for winters (December–March) of the period 1979–2001: (a) triangular N=0–7 (planetary waves, (b) triangular N=8–21 (synoptic variability), (c) triangular N ≥22 (sub-synoptic variability), (d) zonal M=0–3 (planetary waves), (e) zonal M=4–14 (synoptic variability) and (f) zonal M ≥15 (sub-synoptic variability). The contour interval is 10 m for planetary waves (a,d), 5 m for synoptic-scale waves (b,e) and 2.5 m for sub-synoptic-scale waves. See text for details on how the climatology has been computed.

3.3 Verification

The ensemble forecasts are evaluated using operational ECMWF analyses as surrogate for the true state of the atmosphere. In order to verify a particular spectral band, forecasts and verifying analyses are first filtered with the respective waveband filter. All verification statistics are computed on a regular 2.5 $^\circ$ × 2.5 $^\circ$ grid with fields triangularly truncated at wavenumber 63 prior to the spectral to gridpoint transformation. Preliminary tests with triangular truncation at wavenumber 159 and 1.5 $^\circ$ × 1.5 $^\circ$ grid indicate that the probabilistic scores for upper air fields like geopotential and temperature are very similar for the two resolutions (not shown). When the verification statistics are averaged over a region, weights proportional to the cosine of the latitude are used in order to approximate integration on the sphere.

Ensemble forecasts are evaluated by the level of agreement between the ensemble standard deviation, a.k.a. spread, and the Root Mean Square error of the ensemble mean. The two quantities should be equal for a perfectly reliable ensemble prediction system and a sufficiently large sample size (Leutbecher and Palmer 2007). Furthermore, we consider the Brier Score and the Ranked Probability Score (Candille and Talagrand 2005). The Ranked Probability Score is defined here for 10 climatologically equally likely categories. It is equal to the sum of the Brier Scores of exceeding the 1st, 2nd, . . . , 9th climate decile.

4 Results

4.1 Winter 2006/07

4.1.1 Spread-skill relationship

As mentioned in the previous section a reliable ensemble prediction requires that on average the root mean square error (RMSE, hereafter) of the ensemble mean (i.e., the deterministic forecast error) matches the spread of the ensemble about the ensemble mean (i.e., the forecast uncertainty). Previous work has revealed that for Northern Hemisphere Z500 the ECMWF EPS is overdispersive (i.e., the spread exceeds the RMSE) in the short-range; in the far medium-range beyond D+5[2], the EPS appears underdispersive (Buizza et al. 2005). The above results are based on *unfiltered* Z500 fields including all atmospheric scales captured by the model. Now the question arises how the different spectral ranges contribute to the mismatch between spread and skill in the short-range and far medium-range.

The RMSE of the Z500 ensemble mean at D+2 is shown in Figure 4a–c for three different spectral ranges (using the total wavenumber filter). Also shown is the corresponding ensemble spread about the ensemble mean (Figure 4d–f). Evidently, Z500 forecast error in the short-range is most pronounced on synoptic scales, particularly in the North Atlantic and North Pacific storm track regions. The D+2 ensemble spread for the different spectral bands suggest that most of the overdispersiveness of the ECMWF EPS reported by Buizza et al. (2005) is due to excessively large amounts of spread on synoptic scales in the North Atlantic and North Pacific storm track regions (Figure 4b,e). On planetary scales, if anything, the ECMWF EPS seems underdispersive (Figure 4a,d). Too little spread on planetary scales in the Northern Hemisphere extratopics suggested by Figure 4, however, might simply be an artifact coming from the tropics, where the EPS is known to be underdispersive, when the total wavenumber filter is used for low wavenumbers (see above). On sub-synoptic scales the spread of the ECMWF EPS seems to be well calibrated at D+2.

---

[2]As is common practice in the NWP community, D+n denotes a n-day forecast.
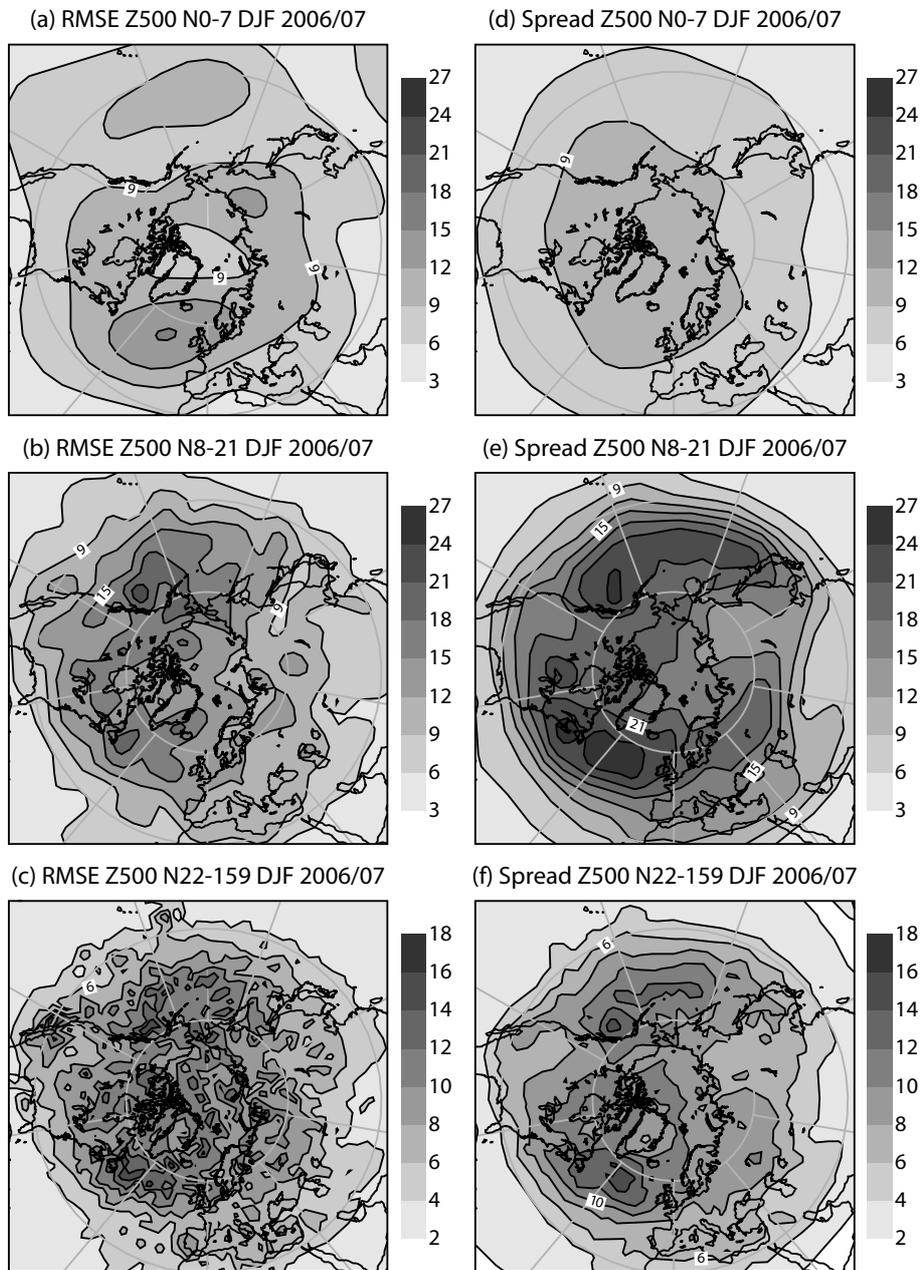
Figure 4: Root mean square error (m) of the ensemble mean at D+2 for different spatial scales: (a) N=0–7 (planetary-scale), (b) N=8–21 (synoptic-scale) and (c) N ≥22 (sub-synoptic scale). Also shown is the ensemble spread about the ensemble mean (in m) at D+2 for (d) N=0–7, (e) N=8–21 and (f) N ≥22.

Although the ECMWF EPS is overdispersive on synoptic scales in the short-range (Figure 4), by D+5 the ensemble spread matches the deterministic forecast error very well, both in terms of magnitude and geographical distribution (Figure 5).
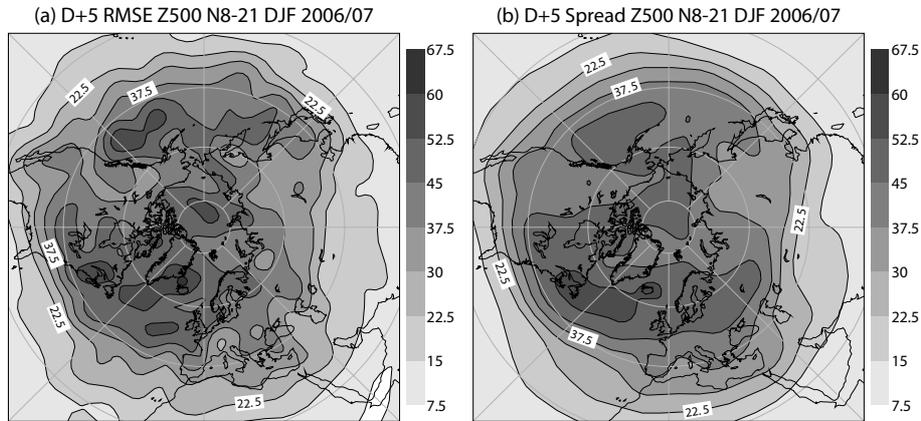
Figure 5: As in Fig. 4c,d but for D+5 forecasts. Notice the different contour interval.

The 'spread-skill relationship' of the ECMWF EPS for the winter of 2006–07 for the three different spectral ranges and all forecast steps from D+1 to D+15 is shown in Figure 6 for the Northern Hemisphere mid-latitudes. For unfiltered Z500 fields (upper panels) the following behaviour of the EPS is confirmed: the EPS is overdispersive early on in the forecast; by D+5 the spread matches the skill; and beyond D+5 the EPS becomes underdispersive. Figure 6 reveals that the overdispersivness in the short-range up to D+4 is primarily due to an overdispersive ensemble on synoptic scales; during early parts of the forecast the agreement between Z500 spread and skill is much better on planetary and sub-synoptic scales. Beyond D+5, both synoptic and planetary scales contribute to the overconfidence of the ensemble. Too little spread beyond D+5 in the mid-latudes for Z500 might partly be of tropical origin (Leutbecher et al. 2007).

Also shown in Figure 6 are 'spread-skill curves' for the unfiltered and spectrally filtered meridional wind component at the 850 hPa level (V850 hereafter) at the mid-latitudes of the Northern Hemisphere. In terms of spread-skill relationship the conclusions are very similar to those for Z500 with the exception that for V850 spread and skill match very well beyond D+8 or so. Another interesting feature revealed by Figure 6 is that the three different spectral band contribute differently to the total error/spread for V850 compared to Z500.

Finally, it is noted that for any band of latitudes $\phi_1 \le \phi \le \phi_2$, the three zonal wavenumber bands provide an exact decomposition of the RMS error and ensemble standard deviation of the unfiltered fields, i.e.

$$s_{\text{unf.}} = \left[s_{\text{plan.}}^2 + s_{\text{syn.}}^2 + s_{\text{sub-syn.}}^2 \right]^{1/2}. \tag{1}$$

Here, s denotes either the RMS error of the ensemble mean or the ensemble standard deviation and the indices unf, plan, syn, sub-syn refer to the unfiltered fields and planetary, synoptic and subsynoptic wavebands, respectively. This decomposition can be derived from the orthogonality relationship of the spherical harmonics when integrating along a circle of latitude. For the total wavenumber bands the corresponding decomposition is only exact if the entire sphere is considered.

4.1.2 Probabilistic forecast skill

Figure 7a,b shows the Ranked Probability Skill Score (RPSS) of Z500 and V850 filtered with the zonal wavenumber filters and unfiltered for the Northern Hemisphere mid-latitudes. For both fields, the probabilistic skill increases significantly with increasing spatial scale. The skill of the Z500 and V850 for a given wavenumber band is quite similar in each of the three wavenumber bands. For Z500 the limit of no skill is reached at
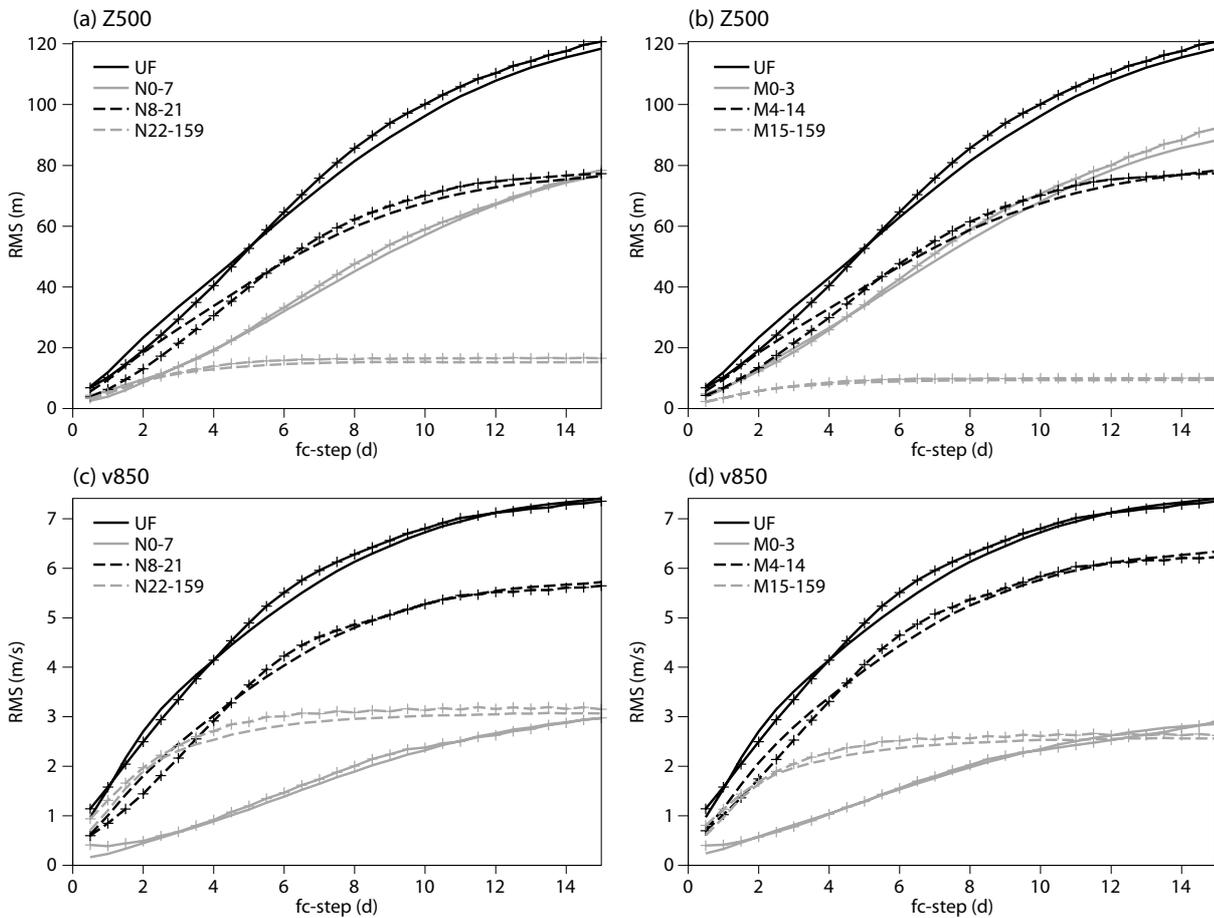
Figure 6: Ensemble standard deviation (no symbols) and ensemble mean RMS error (symbols) for the Northern Hemi-sphere mid-latitudes (35 °N–65 °N) and DJF06/07. Geopotential at 500 hPa (a,b) and meridional wind component at 850 hPa (c,d); decomposition by total wavenumber (a,c) and by zonal wavnumber (b,d).

about D+8 and D+13 for the sub-synoptic and synoptic bands, respectively. The skill in the planetary scales extends beyond the forecast range of 15 days.

The skill of the unfiltered fields lies between the skill of the synoptic waveband and that of the planetary waveband for Z500. This is consistent with the decomposition of the ensemble spread shown in Figure 6b where the spread is dominated by about equal contributions from the synoptic waveband and the planetary waveband. For the meridional wind component, the skill of the unfiltered fields is similar to that of the synoptic waveband. This is consistent with the fact that the ensemble spread for V850 is dominated by the contribution from the synoptic waveband (Figure 6d).

The Ranked Probability Score is a mean over the Brier scores for several events. Therefore, it is of interest to examine how the Brier scores depend on the anomaly threshold defining the event. Figure 7c,d shows the Brier Skill Score (BSS) for large positive Z500 anomalies ($> +1.5$ climatological standard deviations) and large negative Z500 anomalies (< -1.5 climatological standard deviations). For the three wavebands (sub-synoptic, synoptic and planetary), the BSS for large positive anomalies and small negative anomalies are rather similar. However, the BSS of the full fields for large positive anomalies is significantly larger than the BSS for large negative anomalies. The root cause of this asymmetry is the differences in the distribution of horizontal scales between cyclones and anticyclones with the latter being more large-scale. As the larger scales are more persistent and hence more predictable it follows that the skill for large positive events has to be larger than the
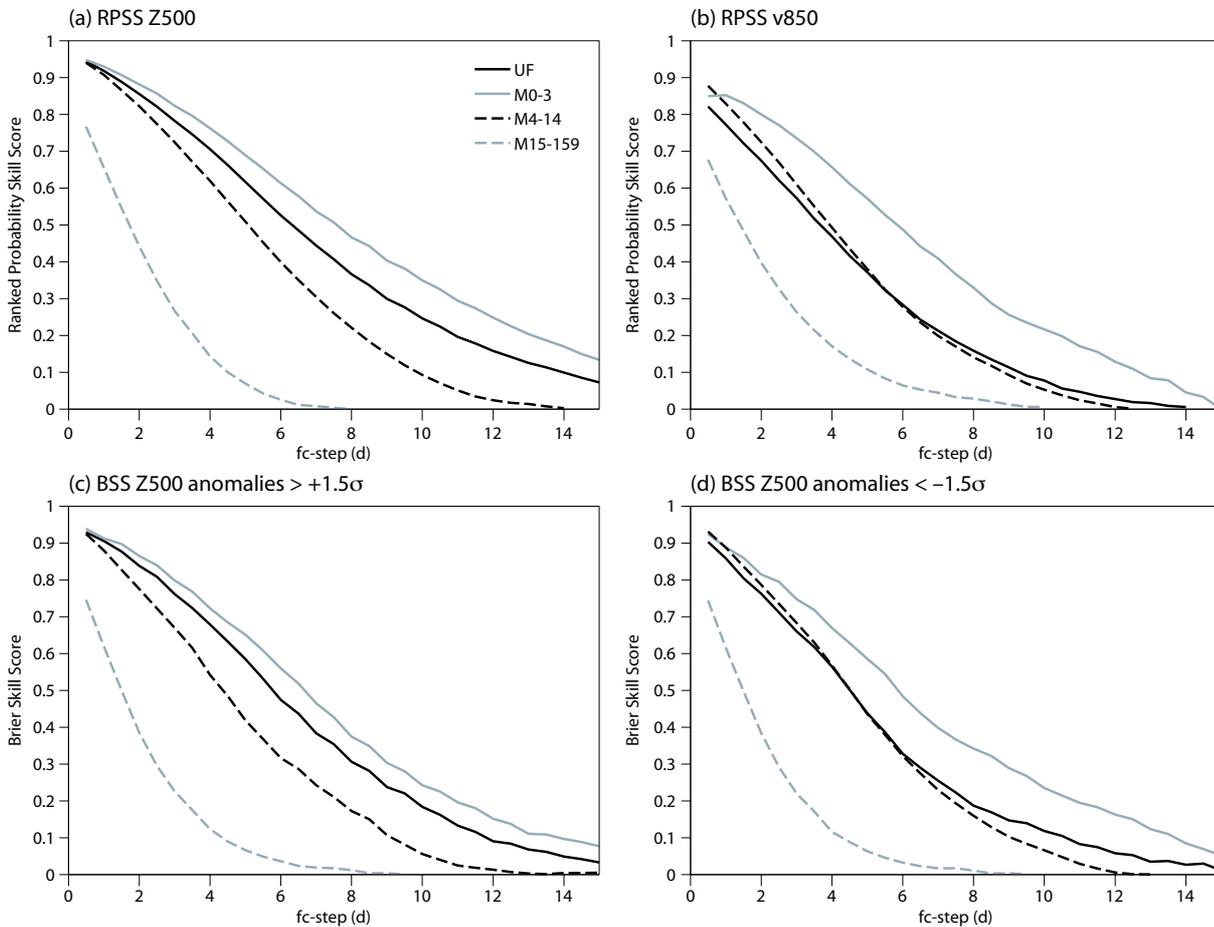
Figure 7: Probabilistic scores for the Northern Hemisphere mid-latitudes (35 °N–65 °N) and DJF06/07. RPSS of geopotential at 500 hPa (a), RPSS of meridional wind component at 850 hPa (b), Brier Skill Score for anomalies of 500 hPa geopotential above +1.5 climatological standard deviations (c), and Brier Skill Score for anomalies of 500 hPa geopotential below −1.5 climatological standard deviations (d).

skill for large negative events.

4.2 Evolution of skill: 1994 – 2007

Now, we investigate how the probabilistic skill of the EPS in the different wavebands evolved from 1994 until 2007 for Z500. As a reference, the probabilistic scores have also been computed for the unperturbed control forecast of the EPS and the higher-resolution deterministic forecast. Here, we focus on the RPSS and summarise results in terms of the lead time when the RPSS reaches 0.3. For the unfiltered Z500 fields, the lead time has increased by about 2 days over the period 1994–2007 (Figure 8a).

Consistent with the steady increase in skill for the unfiltered fields, there has been an increase in skill in all three wavebands from planetary to sub-synoptic scales (Figure 8b–d). This increase in skill is shadowed by a steady though smaller increase in skill of the deterministic forecasts. There appears to be a significant amount of interannual variability in the planetary and synoptic wavebands but not in the sub-synoptic waveband.
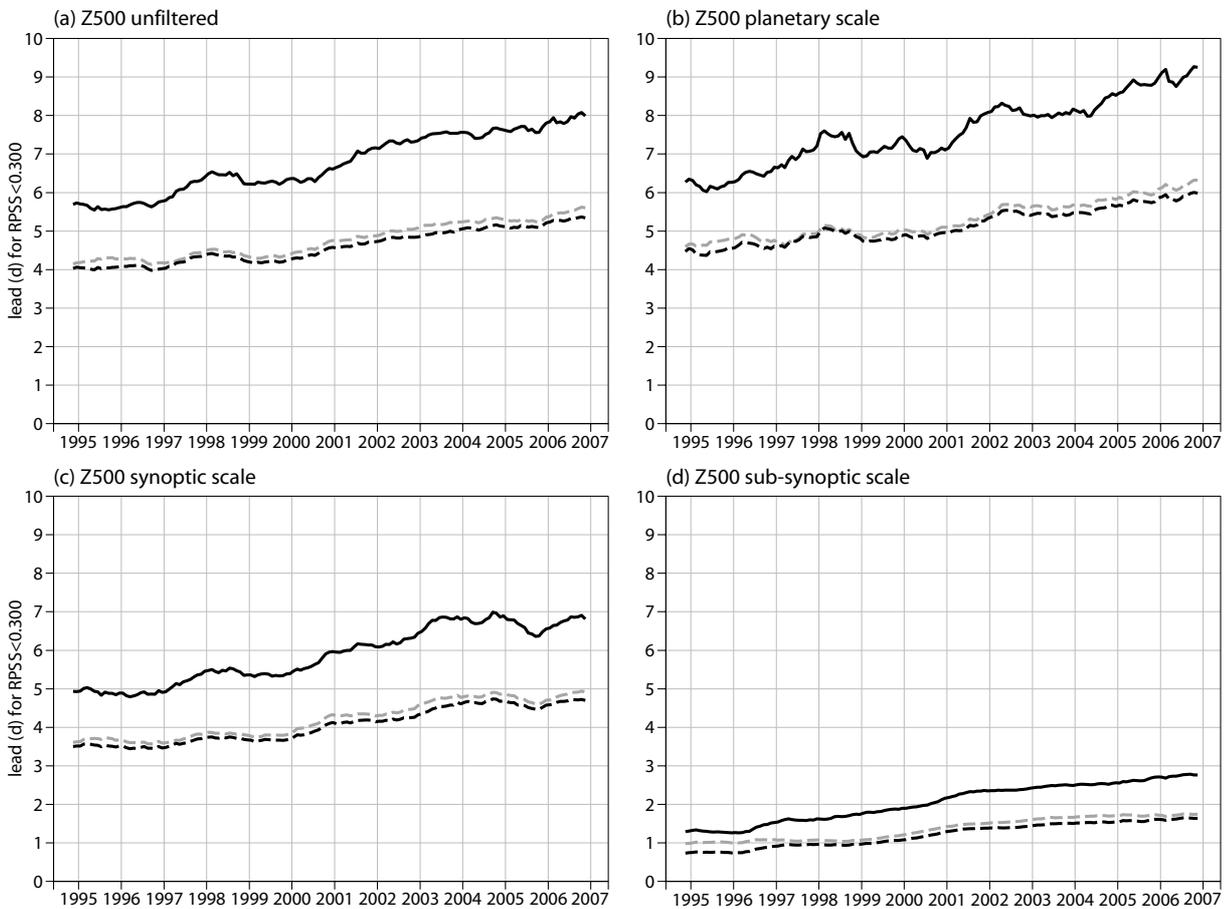
Figure 8: Lead time at which the Ranked Probability Skill Score of 500 hPa geopotential reaches 0.3 in the Northern Hemisphere mid-latitudes (35 °N–65 °N). Solid-black: Ensemble, dashed-black: control forecast, dashed-grey: deterministic high-resolution forecast. (a) unfiltered fields, (b) zonal wavenumber 0–3, (c) zonal wavenumber 4–14, (d) zonal wavenumber ≥ 15. Monthly scores have been smoothed with a 12-month moving average prior to computing the lead time.

5   Discussion

A scale-dependent verification of the ECMWF is presented for three different spectral bands: planetary, synoptic and sub-synoptic scales. An assessment of the relationship between spread and skill shows that the 'overdispersiveness' of the ECMWF EPS in the short-range (up to about D+4) is primarily due to excessive amouts of spread on synoptic scales. This implies that the uncertainty predicted by the EPS in the short-range is too large for synoptic-scale features such as extratropical cyclones. Given the importance of reliable warnings for the occurrence of severe extratropical cyclones, particularly in the range up to D+4 where most of the predictability of such extreme events is concentrated (e.g., Jung et al. 2004, Jung et al. 2005), too much spread on synoptic scales is a significant short-coming. Excessive amounts of spread on synoptic scales might largely be explained by the way the initial perturbations are constructed. In particular since the singular vector technique used in the ECMWF EPS leads to perturbations in the short-range that peak on synoptic scales (see Fig. 12 in Buizza and Palmer 1995) a reduction of the amplitude of the singular vector initial perturbations appears to be desirable. On the other hand, without compensating for reduced amplitude by model effects, a reduced amplitude of the initial perturbations will lead to an even more underdispersive ensemble in the late medium-range.
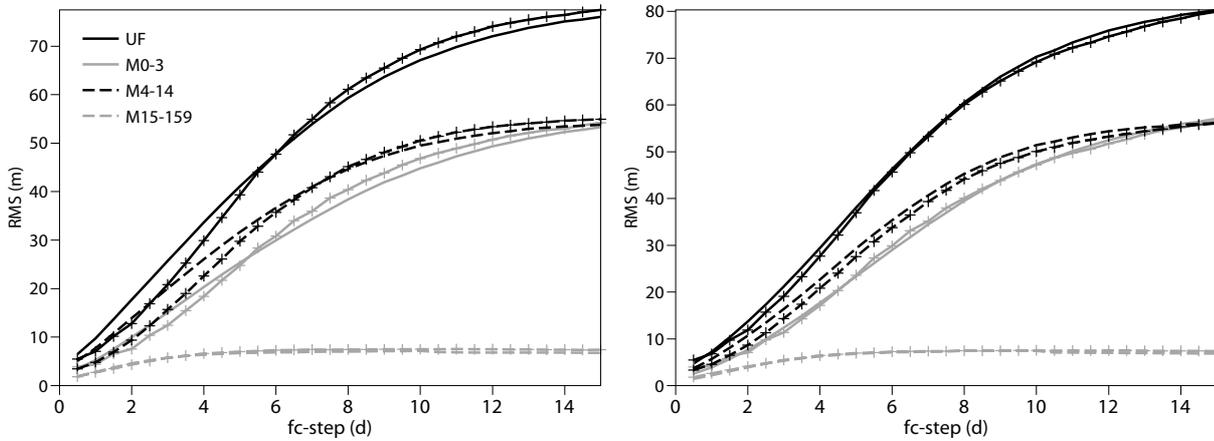
Figure 9: Same as Figure 6b, but for 69 cases during the period June–September 2007: (a) model cycle 32R2 and (b) model cycle 32R3.

The size of the initial perturbations in the ECMWF EPS was reduced by 30% on 9 November 2007. This was possible due to use of a more active model cycle leading to larger perturbation growth [3] (Bechtold et al. 2007). The beneficial impact of these changes for the spread-skill relationship, particularly on synoptic scales, is highlighted by Figure 9, which is based on 69 cases during the period June–September 2007. Evidently, the overdispersiveness of the ECMWF EPS, particularly on synoptic scales, has been substantially reduced with the introduction of the new model cycle. Moreover, the reduction of the initial perturbation amplitude did not lead to too little spread in the medium-range.

Appendix: Computation of the climatology

The climatological statistics is computed from analyses for $N_Y = 23$ years. For day ν in month μ, the climatology is computed from data $N_{1/2}$ days before the day to $N_{1/2}$ days after. The choice of $N_{1/2}$ has to seek a compromise between well resolving the annual cycle and a sufficient sample size. We have choosen $N_{1/2} = 30$. Due to the variable weights this fairly large window still resolves the annual cycle well. Climatologies were also computed with $N_{1/2} = 10$ and 5. They still show some noise on time scales smaller than one month but do not seem to contain additional information about the annual cycle.

The dates are arranged in a periodic manner in order to obtain a continous annual cycle without a jump at New Year. It is convenient to express the dates in terms of Julian day number. Let $J_{\text{YMD}}(j_Y, j_M, j_D)$ denote the Julian day number for day j_D in month j_M and year j_Y. The climate is computed from data between 1 Jan 1979 and 31 Dec 2001. In order to formulate the periodicity, we define $J_0 = J_{\text{YMD}}(1979, 1, 1) = 2443875$ and

---

[3]Changes included a revision to the convection scheme and a reduction of vertical diffusion

$\Delta J = J_{\mathrm{YMD}}(2002,1,1) - J_0 = 8401$. Any date J is mapped to date

$$J' = J_0 + [(J - J_0) \mod \Delta J] \tag{2}$$

which always falls between 1 Jan 1979 and 31 Dec 2001.

The climatology for day ν and month μ is computed from dates

$$D_{jk}(\mu, \nu) = (J_{\mathrm{YMD}}(1978 + k, \mu, \nu) + j)' \tag{3}$$

where index $j = -N_{1/2}, -N_{1/2} + 1, \ldots, N_{1/2}$ specifies the distance in days from the centre of the data window and index $k = 1, 2, \ldots, N_Y$ specifies the year. Thus, the total sample size is $N_Y(2N_{1/2} + 1) = 1403$. The statistics are computed with a weighted average designed to damp high-frequency sampling uncertainties. The weights depend on the distance j from the centre of the window

$$w_j = \frac{3(N_{1/2} + 1)}{N_Y(2N_{1/2} + 1)(2N_{1/2} + 3)} \left[1 - \left(\frac{j}{N_{1/2} + 1} \right)^2 \right] \tag{4}$$

Let x_{jk} denote data at date D_{jk}. Then, the mean and variance are computed as

$$\mathrm{mean} = \sum_{k=1}^{N_Y} \sum_{j=-N_{1/2}}^{+N_{1/2}} w_j x_{jk} \tag{5}$$

$$\mathrm{var} = \sum_{k=1}^{N_Y} \sum_{j=-N_{1/2}}^{+N_{1/2}} w_j \left(x'_{jk} \right)^2, \tag{6}$$

where $x'_{jk} = x_{jk} - \mathrm{mean}_j$ denotes the anomaly about the climatological mean.

The probability distribution of anomalies about the mean is constructed with the aid of the family of distributions

$$p_\varepsilon(x) = \sum_{k=1}^{N_Y} \sum_{j=-N_{1/2}}^{+N_{1/2}} w_j K_\varepsilon \left(x - x'_{jk} \right), \tag{7}$$

where ε is considered to be a small positive number and

$$K_\varepsilon(x) = \begin{cases} 0 & \text{if } |x| > \varepsilon, \\ (2\varepsilon)^{-1} & \text{if } |x| \le \varepsilon. \end{cases} \tag{8}$$

The CDF is defined at the data values $\widetilde{x} \in \{x'_{jk} \mid j = -N_{1/2}, \ldots, N_{1/2}, k = 1, \ldots, N_Y\}$ by integrating p_ε and then taking the limit $\varepsilon \to 0$

$$\mathrm{CDF}(\widetilde{x}) = \lim_{\varepsilon \to 0} \int_{-\infty}^{\widetilde{x}} p_\varepsilon(x) \, dx. \tag{9}$$

If the weights w_j were all equal, the CDF would assume the values given by Hazen's plotting positions (cf. Wilks 2006). For intermediate data points x', the CDF is defined by linearly interpolating the probability between the closest data points enclosing x'. Therefore, quantile values are obtained through linear interpolation of the CDF at the data points. Note, that the definition of the CDF (9) is consistent with the definition of the variance of anomalies in the sense that $\lim_{\varepsilon \to 0} p_\varepsilon$ will have the variance given by Eq. (6).

References

Ayrault F, Lalaurette F, Joly A, Loo C. 1995. North Atlantic ultra high frequency variability. *Tellus* **47A:** 671–696.

Bechtold P, Koehler M, Jung T, Leutbecher M, Rodwell M, Vitart F. 2007. Cycle 32r3: A step forward in predicting atmospheric variability. ECMWF Newsletter **114:**. (pages CHECK ???)

Blackmon ML. 1976. A climatological spectral study of the 500 hPa geopotential height field of the Northern Hemisphere. *J. Atmos. Sci.* **33:** 1607–1623.

Boer GJ. 1984. A spectral analysis of predictability and error in an operational forecast system. *Mon. Wea. Rev.* **112:** 1183–1197.

Boer GJ. 2003. Predictability as a function of scale. *Atmos.-Ocean* **41:** 203–215.

Casati B, Wilson LJ. 2007. A new spatial-scale decomposition of the Brier score: Application to the verification of lightning probability forecasts. *Mon. Wea. Rev.* **135:** 3052–3069.

Blackmon ML, Wallace JM, Lau N-C, Mullen SL. 1976. An observational study of the northern hemisphere wintertime circulation. *J. Atmos. Sci.* **34:** 1040–1053.

Buizza R. 2006. The ECMWF Ensemble Prediction System. In T. N. Palmer and R. Hagedorn (Eds.), *Predictability of Weather and Climate.*pp. 459–488. Cambridge University Press.

Buizza R, Miller M, Palmer TN. 1999. Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.* **125:** 2887–2908.

Buizza R, Bidlot J-R, Wedi N, Fuentes M, Hamrud M, Holt G, Vitart F. 2007. The new ECMWF VAREPS (Variable Resolution Ensemble Prediction System). *Quart. J. Roy. Meteor. Soc.* **133:** 681–695.

Buizza R, Houtekamer PL, Pellerin G, Toth Z, Zhu Y, Wei M. 2005. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems *Mon. Wea. Rev.* **133:** 1076–1097.

Buizza R, Miller M, Palmer TN. 1999. Stochastic representation of model uncertaininies in the ECWMF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.* **125:** 2887–2908.

Buizza R, Palmer TN. 1995. The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.* **52:** 1434–1456.

Candille G, Talagrand O. 2005. Evaluation of probabilistic prediction systems for a scalar variable. *Quart. J. Roy. Meteor. Soc.* **131:** 2131–2150.

Charney JG. 1948. On the scale of atmospheric motions. *Geofys. Publ.* **17:** 1–17.

Epstein ES. 1969. A scoring system for probability forecasts of ranked categories. *J. Appl. Meteor.* **8:** 985–987.

Gulev SK, Jung T, Ruprecht E. 2002. Climatology of interannual variability in the intensity of synoptic-scale processes in the North Atlantic from the NCEP-NCAR reanalysis data. *J. Climate* **15:** 809–828.

Jung T, Klinker E, Uppala S. 2004. Reanalysis and reforecast of three major storms of the twentieth century using the ECMWF forecasting system. Part I: Analyses and deterministic forecasts. *Meteor. Appl.* **11:** 343–361.

Jung T, Klinker E, Uppala S. 2005. Reanalysis and reforecast of three major storms of the twentieth century using the ECMWF forecasting system. Part II: Ensemble forecasts. *Meteor. Appl.* **12:** 111–122.

Jung T, Leutbecher M. 2007. Performance of the ECMWF forecasting system in the Arctic during winter. *Quart. J. Roy. Meteor. Soc.* **133:** 1327–1340.

Leutbecher M, Palmer TN. 2007. Ensemble forecasting. *J. Comp. Phys* **???:**, ???–??? (in press). DOI: 10.1016/j.jcp.2007.02.014. Also available as ECMWF Tech. Memo. 514.

Leutbecher M, Buizza R, Isaksen L. 2007. Ensemble forecasting and flow-dependent estimates of initial uncertainty. Proceedings of Workshop on *Flow-dependent aspects of data assimilation*, ECMWF June 2007, 185–201.

Molteni F, Buizza R, Palmer TN, Petroliagis T. 1996. The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.* **122:** 73–119.

Mureau R. 1990. The decrease of the systematic error and the increased predictability of the long waves in the ECMWF model. ECMWF Technical Memorandum 167. ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom.

Puri K, Barkmeijer J, Palmer TN. 2001. Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. *Quart. J. Roy. Meteor. Soc.* **127:** 709–731.

Simmons AJ. 2006. Observations, assimilation and the improvement of global weather prediction—some results from operational forecasting and ERA-40. In T. N. Palmer and R. Hagedorn (Eds.), *Predictability of Weather and Climate*.pp. 459–488. Cambridge University Press.

Simmons AJ, Hollingsworth A. 2002. Some aspects of the improvement of skill of numerical weather prediction. *Quart. J. Roy. Meteor. Soc.* **128:** 647–677.

Uppala S, Kallberg PW, Simmons AJ, Andrae U, Da Costa Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Holm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf J-F, Morcrette J-J, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, and Woollen J. 2005. The ERA-40 re-analysis *Quart. J. Roy. Meteor. Soc.* **131:** 2761–2779.

von Storch H, Zwiers FW. 1999. Statistical Analysis in Climate Research. Cambridge University Press, Cambrige, 484pp.

Wallace JM, Gutzler DS. 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. wea. Rev.* **109:** 784–812.

Wilks DS. 2006. Statistical Methods in the Atmospheric Sciences. Academic Press. 627 pp.