

A review of predictability and
ECMWF forecast performance, with
emphasis on Europe

T. Palmer, Č. Branković, R. Buizza,
P. Chessa, L. Ferranti,
B. Hoskins and A. Simmons

Research Department

December 2000

This paper has not been published and should be regarded as an Internal Report from ECMWF.
Permission to quote from it should be obtained from the ECMWF.



A review of predictability and ECMWF forecast performance, with emphasis on Europe

T. Palmer, C. Brankovic, R. Buizza, P. Chessa,
L. Ferranti, B. Hoskins, A. Simmons

1. Introduction

This review discusses issues of predictability and recent forecast performance over Europe, in a comparative sense. For example, are systematic and flow-dependent errors larger over Europe than elsewhere; has it been more difficult to improve forecast performance over Europe than elsewhere; in what sense is the general circulation different over Europe than, say, North America?

The topics discussed include: trends and variability in medium-range skill scores in section 2; statistics of error doubling times in section 3; features of the general circulation linked with high and low forecast skill scores (including persistence scores) in section 4; interannual variations in the dominant singular values of the 48-hour linearised propagator in section 5; systematic and flow-dependent error in section 6. Some other relevant topics are summarised in section 7; concluding remarks are made in section 8.

Some of the results are clear cut. For example, it is shown there is no objective evidence that it has been harder to improve forecast skill over Europe than over regions such as North America. On the other hand, the interpretation of other results are more open ended, raising questions such as: what is the dynamical significance of the flow-dependent error patterns revealed by singular value decomposition (SVD) techniques, and how do these patterns relate to the systematic error fields in the medium and extended range. Some nonlinear dynamical interpretation of these topics is offered in this paper, based on circulation regime analysis.

2. Trends and variability in medium-range predictive skill

The unusually poor quality of deterministic ECMWF forecasts for Europe in the summers of 1998 and 1999 led to speculation as to whether there were fundamental difficulties in forecasting for this region that might be identified either by comparison with the skill of forecasts for other regions or by evidence of a slower rate of improvement of forecasts for Europe than for elsewhere. To examine this, a study has been made of objective measures of the skill of day-five forecasts over the past decade. Comparisons have been made of the forecasts for various regions from several centres.

Operational analyses and forecasts of 500hPa height fields are received regularly from a number of weather services, and the forecasts are routinely verified in the same way as the operational ECMWF forecasts (*Niemenen*, 1983). The forecasts of a particular centre are normally verified against the analyses of that centre, and such verifications will be presented here. It can also be instructive to verify against the analyses of other centres or against observations. Some recent discussion of this has been given by *Simmons* (2000).

2.1 Annual-mean performance

Figs. 1 and 2 are based on archived monthly means of the root-mean-square errors of the day-five 500hPa height forecasts. Annual running means of these are plotted beginning in 1991 for the forecasts from ECMWF, the Met. Office, Deutscher Wetterdienst (DWD) and the National Centers for Environmental Prediction (NCEP), USA. Fig. 1 shows scores computed over Europe and North America, and over the extratropics of the Northern Hemisphere and the Southern hemisphere. Fig. 2 shows results for the North Atlantic, the North Pacific, East Asia and Australia/New Zealand. Figs. 3 and 4 show corresponding anomaly correlations.

These figures show that the ECMWF forecasts improved over all regions during the 1990s, as judged by both measures of accuracy. The same is largely true for the forecasts of the other centres. The ECMWF forecasts are generally the most accurate. ECMWF typically has less of a lead over the other centres for the northern hemispheric sector from North America eastward to Europe than for East Asia and the North Pacific. The skill of the ECMWF forecasts has improved sharply over the Southern Hemisphere during the past few years, over Australia/New Zealand in particular. Here though, the lead that developed over other centres has been reduced recently as a result of substantial improvements in the forecasts from the Met. Office and NCEP.

Superimposed on the trend towards higher skill is an interannual variability in forecast quality, particularly within the Northern Hemisphere. This variability is to a degree common to all centres. It is particularly large over the North Pacific and North America, where forecast accuracy was at a maximum in the annual mean centred on winter 1997/98, coincident with the latest and strong occurrence of El Niño. Studies by *Palmer* (1988) and *Simmons* (1988) have indicated that the Pacific/North American mode or teleconnection pattern is relatively stable when it is in the (positive) phase that predominates during El Niño, and that variations in medium- and extended-range forecast skill may be linked with this.

Root-mean-square errors vary substantially from region to region, being largest for the Southern Hemisphere as a whole, followed by Europe, and smallest for Australia/New Zealand. This, however, appears simply to reflect regional variations in the temporal and sub-regional spatial variance of the 500hPa height field. As shown by *Simmons* (1996) for example, the anomaly correlation coefficient computed over a sufficiently large number of cases is given to a good degree of approximation simply by one minus the mean-square error normalized by the variance of the forecasts or verifying analyses, for a forecast model which has low mean errors and a realistic representation of the level of variance. Anomaly correlations do not vary substantially from region to region in the mean, particularly for more recent ECMWF forecasts. This can be seen in Table 1, which shows the maximum and minimum annual-average root-mean-square errors and anomaly correlations recorded for ECMWF during the period illustrated in Figs. 1 to 4. Results are also shown for a persistence-normalized skill score discussed later. For each region the poorest results (highest error and lowest correlation) occur within the first year or so of the series, and the best results occur within the past two years or so. The range of minimum root-mean-square error varies from 36.1m to 70.3m, but the range of maximum anomaly correlation varies only from 77.0% to 80.8%. Also given in Table 1 are crude measures of relative improvement in these forecast scores. They are computed by dividing the difference between maximum and minimum values by the maximum improvement possible when scores were at their poorest. This gives an overestimate of improvement due to forecasting system changes alone, as it includes a contribution from interannual variability. The larger relative improvement of anomaly correlations compared with root-mean-

square errors is consistent with the approximate linear relationship between anomaly correlation and mean-square error.

| Region | Root-mean-square error | | | Anomaly correlation | | | Persistence-normalized skill | | |
|--------------|------------------------|-------------------|-------------------------|---------------------|-------------------|-------------------------------|------------------------------|-------------------|-------------------------------|
| | <i>Max</i> (m) | <i>Min</i> (m) | $\frac{Max - Min}{Max}$ | <i>Min</i> (%) | <i>Max</i> (%) | $\frac{Max - Min}{100 - Min}$ | <i>Min</i> (%) | <i>Max</i> (%) | $\frac{Max - Min}{100 - Min}$ |
| N.Hemisphere | 66.9 | 55.1 | 0.18 | 74.6 | 80.8 | 0.24 | 64.0 | 74.7 | 0.30 |
| Europe | 79.5 | 63.0 | 0.21 | 72.2 | 80.1 | 0.28 | 65.4 | 77.8 | 0.36 |
| N.America | 63.5 | 52.1 | 0.18 | 66.8 | 77.4 | 0.32 | 64.2 | 75.2 | 0.31 |
| E.Asia | 53.9 | 43.1 | 0.20 | 66.2 | 79.1 | 0.38 | 59.4 | 76.5 | 0.42 |
| N.Atlantic | 74.2 | 60.7 | 0.18 | 71.5 | 78.9 | 0.26 | 67.2 | 76.6 | 0.29 |
| N.Pacific | 69.4 | 57.1 | 0.18 | 71.2 | 80.5 | 0.32 | 61.8 | 76.0 | 0.37 |
| S.Hemisphere | 88.6 | 70.3 | 0.21 | 65.1 | 77.0 | 0.34 | 53.5 | 72.9 | 0.42 |
| Australia/NZ | 51.4 | 36.1 | 0.30 | 61.0 | 80.0 | 0.49 | 58.4 | 79.5 | 0.51 |

Table 1: Range and relative improvement of ECMWF annual running-mean day-5 500hPa height scores from 1991 to present

Overall, these results do not show Europe to be a region where, in the annual mean, forecast skill is unusually low or unusually difficult to improve. In fact, it can be seen in Table 1 that Europe has relatively high anomaly correlations and persistence-normalized skill scores, and the maxima do not represent isolated peaks in the time series, but rather occur during a recent period with sustained high annual-mean scores. The large root-mean-square errors of forecasts for Europe appear to be a consequence of the large variance of the 500hPa height field over this region, and they have been reduced proportionately by slightly more than have errors for other regions of the Northern Hemisphere since 1991.

2.2 Seasonal-mean performance

Although a problem in annual-mean performance does not show up for Europe, this could be because wintertime performance is sufficiently good to counterbalance relatively poor summertime performance. To examine this, Fig. 5 shows four-month running means of the anomaly correlations of day-five 500hPa forecasts for Europe and North America. Fig.6 presents the corresponding plots for the Northern Hemisphere and the Southern Hemisphere. These figures display a general seasonal variability in scores, with highest anomaly correlations in winter and lowest correlations in summer. The annual cycle over the Northern Hemisphere is more regular than over Europe or North America, and more regular than over the Southern Hemisphere. Overall, the upper panel of Fig. 5 does indicate a tendency for the ECMWF forecasts for Europe to have more of a lead over those of other centres in winter than summer, although this was not the case from 1995 to 1997. In addition, the performance of the ECMWF system over North America in late spring and summer has been relatively poor in recent years, especially the current one. This is the subject of continuing investigation.

The four-month mean score of the ECMWF forecasting system over Europe dropped to a sharp minimum in the summer of 1999, a net effect of spells of particularly poor forecasts within the May to August period. Over these four months the mean of the anomaly correlations of the ECMWF forecasts was well below the corresponding mean correlations of the Met. Office and NCEP forecasts, and on a par with that of the DWD forecasts, making this the worst four-month period on record overall for ECMWF relative to the other centres. ECMWF forecasts over North America were also poor in mid-1999. Over both regions, scores reached levels not seen since 1994, and this was the case also for ECMWF scores computed over the whole of the extratropics of the Northern Hemisphere. Southern Hemisphere scores were, however, at a high level.

Scores for Europe and North America were generally lower than normal in the winter of 1998/99. The summertime forecasts of ECMWF were also quite poor over Europe in 1998, despite scores for the Northern Hemisphere being higher than any earlier summer. Scores for the past winter have also been at a generally high level, the highest ever for Europe and the Northern Hemisphere in the case of ECMWF and the Met. Office.

The pronounced minimum in the anomaly correlations of the NCEP forecasts over the Northern Hemisphere seen in the summer of 1998 is known to be due to problems in the performance of a new version of the forecasting system introduced in June that year. The problems were remedied in the autumn. Also, the sharp drop in the Southern Hemisphere scores of recent DWD forecasts is known to be associated with problems in the quality of SATEM retrievals, which are not used by the other centres who instead assimilate radiances directly. The retrievals had to be removed temporarily from the DWD analysis system. Studies to date have not found a single such major source of error which can account for the poor performance of the ECMWF system in the summer of 1999, although several relevant sensitivities have been identified, as discussed in ECMWF/SAC/29(00)4c.

2.3 Means, medians and intraseasonal variability

An alternative representation of the seasonal performance of the ECMWF system over the past decade is given in Figs. 7 and 8. It is based on statistics of 120-day sequences of anomaly correlations of day-five 500hPa height forecasts. Results for Europe and the extratropics of the Northern Hemisphere are shown in Fig. 7. The red curves denote the medians of the running 120-day distributions of scores, and the blue curves the running (arithmetic) means of the scores. The latter simply give a finer-resolution picture than is provided by the four-month running averages of the monthly-mean scores that are presented in Figs. 5 and 6. The difference between the median and the mean curves is not large for the Northern Hemisphere, but is much more pronounced for Europe. The mean is distinctly lower than the median, due to the skewed distribution of correlations. The Fisher transform, $(0.5(\ln(1+a) - \ln(1-a)))$ where a is the anomaly correlation coefficient, is the standard way of transforming a set of correlations for statistical analysis¹. The green curves in Fig. 7 are obtained by calculating the means of the Fisher transforms of the daily correlations and transforming these means back to correlations. They are very close to the median curves, both for Europe and for the Northern Hemisphere. The medians (or means of the Fisher transforms) are generally slightly higher

1. The Fisher transform is not ideally suited for the correlations of predicted and analysed anomalies, as it gives large weight to correlations close to -100% as well as to correlations close to 100%. This is not important here, however, as only few forecasts give negative correlations at the five-day range.



for Europe than for the Northern Hemisphere, whereas the converse is the case for the means of the correlations, as noted earlier. Note however, that the difference between the median and the mean of the correlations does not vary much in time, so that both curves provide a similar picture of trends and interannual variability.

Fig. 7 also provides an indication of the variability in forecast skill within the 120-day periods. This is shown by shading the range between the 10th and 25th percentiles of the distribution and also the range between the 75th and 90th percentiles. As the ordinates are linear in the Fisher transform, the upper and lower shaded bands appear to be of similar width in a particular plot, although the width differs from one plot to the other. The much larger variability of forecasts over the smaller European domain is evident, as reflected in the larger differences between the medians and means of the correlations. Relatively poor performance in the summer of 1999 (and also in the summer of 1998 for Europe) comes more from a worsening of the scores of the poorer forecasts in the sample than from a worsening of the better forecasts in the sample. In the longer term, the overall improvement in Northern Hemisphere scores appears to be obtained across the range of forecast performance. In contrast, over Europe the poorer forecasts appear to have been improved the most, as seen especially in the jump in the lower band after 1994. This jump is most likely related to important model changes in April 1995 (Miller *et al.*, 1995).

Is the variability of forecast skill within a season larger over Europe than over other regions of the Northern Hemisphere? The answer is no, if skill is measured by the anomaly correlation of the day-five forecast. Fig. 8 shows plots as in Fig. 7 but for East Asia, the North Pacific, North America and the North Atlantic (and, for clarity, without the means of the Fisher transforms.) For each of these regions, the 10th and 25th percentiles of the distribution are typically lower than those for Europe, whereas less difference can be seen for the 75th and 90th percentiles. The difference between the medians and means for the various regions is quite similar, suggesting again that the operational practice of simply averaging correlations to produce monthly and longer-term means, though not statistically ideal, is sufficient for monitoring and comparative studies such as presented earlier.

2.4 Further remarks

It has been argued that Europe is not a particularly problematic region for forecast skill. This has been based on anomaly correlation being an appropriate measure for judging skill, since over a large enough sample of cases it is to good approximation simply related to the mean-square error normalised by the natural level of variance over the region in question. It must nevertheless be noted that the anomaly correlation does not give an invariably reliable measure of skill for an individual forecast over a particular region. A simple extreme example suffices to illustrate this. Suppose that the atmospheric situation evolves such that the flow becomes very close to the climatological mean over the region, and that the forecast predicts the observed (very weak) anomaly pattern but with the wrong sign. Then the root-mean-square error will be small and the forecast will correctly predict conditions close to the climatological mean, but the anomaly correlation will be the lowest possible, -100%. Conditions approaching this extreme may occur more often in regions with relatively low natural variability. Europe is a region of relatively high natural variability, and may thus be less prone than other regions for anomaly correlation to give unreliable indications of low forecast skill.

Another factor to be kept in mind is the possible effect of use of an unrepresentative or erroneous climatology on the calculation of anomaly correlations. The operational ECMWF verification has been running since

January 1980, soon after the start of operations, and has continued to use a climatology available at the time, produced by NCAR and based on the atlases of *Taljaard et al.* (1969) and *Crutcher and Meserve* (1970). This has the advantage of providing continuity and a reference state that is unbiased with respect to the different forecasting centres for which anomaly correlations have been compared. It is, however, shown in Annex that use of an inappropriate climatology can cause anomaly correlations to be overestimated. An absolute reduction in anomaly correlation of about 3% at the 60% level was illustrated by *Simmons* (1996) for northern hemispheric winter scores for 1995 when changing from the older NCAR climatology used operationally to the mean of the operational ECMWF analyses for the period from 1981 to 1995. The NCAR climatology is likely to be most accurate over land masses with good radiosonde coverage, so anomaly correlations computed from it may give an overestimation of correlations for the Atlantic, Pacific and Southern Oceans relative to those for regions such as Europe and North America.

In view of these uncertainties about anomaly correlation, an alternative skill score has been examined. It is defined by one minus the mean square error of the forecast normalized by the mean square error of persistence, multiplied by 100 to express it as a percentage. Overall, this score gives much the same picture as anomaly correlation, although it has a somewhat different interannual variability. Plots are not presented here, but minimum and maximum running annual-mean values and relative improvements are included in Table 1. For this persistence-normalized score, as well as for anomaly correlation, Europe is seen to be a region where the skill of day-five forecasts is above-average.

3. Error doubling times

One approach to understanding variations in forecasting-system performance is to estimate the variations in intrinsic predictability that arise from low frequency intraseasonal and interannual variations in the atmospheric general circulation. This can be investigated, for example, by constructing a simple index of the baroclinic instability of the flow in terms of mean latitudinal temperature gradient and static stability (the so-called "Eady Index"), or by constructing an index based on the amplification factors of the leading singular vectors used to construct the initial perturbations for ensemble forecasting. *Lorenz* (1982) estimated intrinsic forecast-error growth rates by fitting a simple error-growth model to the differences between successive forecasts verifying at the same time. *Lorenz*' study was based on forecasts from a 100-day period beginning 1 December 1980. His conclusions were re-examined by *Simmons et al.* (1995; see also *Simmons*, 1996) in the light of subsequent operational performance. Diagnosis of more recent forecasts gave estimates of intrinsic error doubling time of around 1.4 to 1.6 days, compared with the values of around 1.8 to 2 days found using the early operational forecasts. It was argued that this was likely to be due to the development of a more active (and more realistic) operational forecast model rather than to a shift in the predictability of the atmosphere itself. Intrinsic error growth rates in the extratropics were found to be somewhat slower in the Southern Hemisphere than in the Northern Hemisphere.

Table 2 presents error doubling times computed for the Northern Hemisphere for each season of the 1990s. Generally, doubling times are relatively short in autumn and summer, most variable in spring and longest in winter. This is broadly consistent with the seasonal variations seen in the scores based on monthly anomaly correlations presented in Figs. 3 and 4. Some of the interannual differences in doubling time can be related to known consequences of model changes. In particular, the shorter doubling times beginning autumn 1991 occurred following operational implementation of the Centre's T213L31 model, which was more active than

its T106L19 predecessor. The early T213L31 version in fact tended to overdevelop systems due to problems both with its new semi-Lagrangian advection scheme (revised in August 1992; *Ritchie et al.*, 1995) and with its parametrization of cloud/radiation interaction (revised in February 1993; *Morcrette*, 1993). This is consistent with the relatively low doubling times recorded for the first three seasons of 1992.

| | Winter Dec-Feb | Spring Mar-May | Summer Jun-Aug | Autumn Sep-Nov |
|------|-------------------|-------------------|-------------------|-------------------|
| 1990 | 1.76 | 1.57 | 1.73 | 1.51 |
| 1991 | 1.75 | 1.65 | 1.71 | 1.45 |
| 1992 | 1.52 | 1.49 | 1.43 | 1.45 |
| 1993 | 1.61 | 1.52 | 1.54 | 1.46 |
| 1994 | 1.57 | 1.50 | 1.54 | 1.42 |
| 1995 | 1.51 | 1.50 | 1.50 | 1.53 |
| 1996 | 1.64 | 1.59 | 1.54 | 1.46 |
| 1997 | 1.63 | 1.60 | 1.53 | 1.48 |
| 1998 | 1.60 | 1.48 | 1.51 | 1.46 |
| 1999 | 1.51 | 1.42 | 1.45 | 1.47 |

Table 2: Intrinsic error doubling times (days) for the Northern Hemisphere, following Lorenz(1982). Winter 1990 denotes the period from 1 December 1989 to 28 February 1990

Another noteworthy feature of Table 2 is that error doubling times for spring and summer are significantly lower in 1999 than in any other recent year, and the winter 1999 values are also relatively low. This cannot be linked in any obvious way to recent model changes, and provides some reassurance that the relatively poor ECMWF forecast performance in early and mid 1999 might have been due, in part at least, to an inherently less predictable synoptic regime. Some confirmation of this is provided by the day-5 Northern Hemisphere scores presented in the lower-left panels of Figs. 1 and 3, which show that the running annual-mean accuracy levels of both ECMWF and Met. Office forecasts were a minimum over the year to August 1999

4. Time-filtered variability, the skill of persistence forecasts, and flow patterns associated with high and low skill scores.

In the discussion in section 2 above, the skill of a persistence forecast was used to normalise the forecast model skill scores. Fig 9 shows the D+5 day rms of a persistence forecast for a number of regions. There is substantial multi-year variability. Can we expect periods where persistence error is large to correspond to periods of low predictability? How is variability in the skill of persistence linked with the transient statistics of the general circulation itself? To address this latter question, panels a)-c) of Fig 10 show diagnostics of transient variability from ERA-15, time filtered to show variability on timescales greater than 20 days (low pass), between 7 and 20 days (band pass), and less than 7 days (high pass). For all these timescales there is more variability over Europe than over North America, consistent with the larger rms forecast errors over Europe discussed in section 3.

The map of high pass variability shows the conventional representation of storm-track activity; such high-pass variability can therefore be thought of as representing synoptic scales. Consider the variability associated with slower-than-synoptic-scale variability. If such variability is dominated by low-pass activity, then one would expect the skill of a medium-range persistence forecast to be relatively skilful. Conversely, if slower-than-synoptic-scale variability is dominated by band-pass activity, then we might expect the skill of such a persistence forecast to be relatively poor. As such, the ratio of band- to low- pass activity should give a measure of the medium-range skill of a persistence forecast. It could be imagined that medium-range forecasting would be particularly difficult for years in which the ratio is large. Fig 10d shows a map of the ratio of band- to low- pass activity; compared with the variance maps in Fig 10a-c, this ratio is not have much spatial variability. In particular, whilst the variances themselves are larger over Europe than over North America, the ratio of variance has similar size over the two continents.

Focusing attention on Europe, Fig 11 shows a timeseries of this ratio of band-pass to low frequency variance averaged over Europe. As one might expect, this timeseries correlates reasonably well (on seasonal to interannual timescales) with the skill of a persistence forecast over Europe (Fig 9a); notice in particular the maxima in 1995 and 1999.

Whilst Figs 9 and 11 show interesting evidence of multi-seasonal behaviour, there is only a poor correlation between the rms skill of the operational forecast over Europe, and the ratio of band- to low- pass activity. For example, whilst the European rms error and anomaly correlation coefficient showed relatively poor skill scores in the winter 1998/99 period, there was no dip in forecast scores in 1995 when the ratio of band- to low- pass activity was large. Of course, one could deduce from this that the operational forecast system did exceptionally well during 1995 (i.e. overcoming the fact that, according to persistence, this was a difficult period to forecast). On the other hand, the skill of a persistence forecast alone need not be a strong indication of intrinsic predictability as the schematic in Fig 12 (based on specific pairs of trajectories on the *Lorenz*, 1963, attractor) shows.

This raises the question of how best to relate the component of interannual variability of skill fluctuations that is linked with variations in predictability, to the general circulation itself. What sort of flow patterns are associated with skilful European forecasts (cf. *Palmer and Tibaldi*, 1986)? As mentioned in section 2, there is a tendency for medium range skill scores over North America to be correlated with the phase of the Pacific/North American teleconnection pattern. Similar results hold for the North Atlantic Oscillation; Fig 13 (from *Ferranti et al.*, 2000) shows the dominant singular vector pairs¹ derived from a singular value decomposition (SVD) between the verifying analysis and the magnitude of D+5 forecast error. The pattern associated with the magnitude of forecast error (right singular vector) shows that Europe is not a dynamically-homogeneous region. Overall, rms errors are enhanced during periods where this NAO-like pattern is in its negative phase (the phase where the Eady index of baroclinic instability is relatively small). However, there is a region covering most of the UK where rms errors decrease during periods of negative NAO. One can surmise that the general decrease in rms error over the UK associated with negative NAO-like flow is associated with a

1. In this paper, the term "singular vector" is used to describe the relevant elements of a singular value decomposition of a general matrix, and not just to the linearised dynamical propagator. Whilst the latter application is discussed in section 5, singular vectors derived from an SVD of statistical/empirical covariance matrices are discussed in this section and in section 6

decrease in storm-track activity. However, the general increase in rms error over northern and southern Europe (and most of the East Atlantic) could be associated with a stronger barotropic instability as discussed in *Palmer (1988)* and *Simmons (1988)* for the negative PNA pattern. This is discussed further below.

5. Interannual variations in singular values of the 48-hour linearised dynamical propagator

One potentially useful tool for studying interannual variations in the predictability of the atmosphere is the ensemble prediction system (EPS). In practice, however, the inclusion of evolved singular vectors and stochastic physics (both of which affect ensemble dispersion) over the last couple of years has made the EPS itself too non-stationary to be useful for this type of study. On the other hand the computation of the dynamical singular vectors themselves has not changed in the last three years, and we plot in Figs 15-16 some statistics of the dominant singular values from DJF 96/97 to DJF 98/99 and from JJA 97 to JJA 99.

First six equally sized regions are defined in the northern hemisphere (see Fig 14). A method is used to assign an operational EPS northern hemisphere singular vector to one of these regions. Specifically, each day, 25 singular vectors are computed and used as initial conditions in the EPS. For a given singular vector, the finite-time total energy $TE(x)$ is computed at each gridpoint x . The maximum value TE_{max} of TE is identified. A set Σ of gridpoints x is defined such that $TE(x) \geq 0.5TE_{max}$. The singular vector is assigned to the region in Fig 14 which has the largest intersection with Σ .

Fig 15 shows the seasonal average of the maximum singular value for each region (and for the rms average over all regions). In winter, the largest singular values occur for the Pacific and Atlantic regions consistent with the position of the two climatological jet maxima just upstream of these regions (cf. *Buizza and Palmer, 1985*). The seasonal averages of the maximum singular value over Europe and North America are very similar, with Europe having somewhat larger values in winter.

There is clear evidence of variations in the seasonal average of the maximum singular values (see Fig 16). For example, the dominant winter singular values over the Pacific have been monotonically rising from 96/97 to 98/99. However, this does not match the interannual variation in DJF D+5 skill over North America (see Fig 1), though it does match the variation in the skill of a DJF D+5 persistence forecast over North America (see Fig 9). Overall these statistics of 48-hour singular values are not well correlated with medium-range skill score variations. From a dynamical point of view this is not too surprising. Singular vector growth tends to be associated with regions of strong baroclinic instability (dominant singular vector locations are highly correlated with regions of strong "Eady index"; *Buizza and Palmer, 1985*). However, medium-range error appears to be associated as much with some form of combined baroclinic/barotropic growth associated with large-scale weather regime patterns. For example, (as shown in *Palmer, 1988*; *Simmons 1988*) D+7 forecast error over the Pacific/ North American region is larger during periods of negative PNA index (ridge over the north Pacific), despite the fact that the Eady index over the Pacific is anomalously small during such periods of negative PNA index. (These results have been confirmed using SVD analysis on forecasts data from the 1990s; *L. Ferranti*, personal communication.) It was found (in the two papers cited above) that differences in barotropic growth could explain the relative instability of the negative PNA pattern.

In a further paper, *Molteni et al. (1990)* described the manifestation of this relative instability in terms of nonlinear circulation regime dynamics (see also section 6 below). Specifically, it was found that in the PNA

region, there was a single well-defined regime with positive PNA index. On the other hand, there were multiple less-well-defined regimes for negative PNA index.

In contrast with the medium-range results discussed above, short-range D+2 forecast error over the Pacific region is larger during periods of positive PNA index; Palmer and Tibaldi, 1986. It can therefore be expected that there may be some relationship between dynamical singular values and error doubling times (as shown in Table 2). This appears to be the case (though this is hardly significant in view of the limited number of available years). For example, in the years in which a comparison can be made, the smallest doubling times occurred in summer 1999, and winter 1998/99; these were also seasons with the largest regional-mean singular value.

A clear and consistent disagreement between the singular values and the doubling times concerns the effect of the seasonal cycle. The summer singular values are uniformly smaller than the winter singular values, consistent with the reduction in the Eady index in summer. On the other hand the doubling times are clearly smaller in summer. A possible explanation for this discrepancy is the lack of moist diabatic processes in the operational singular vector computation. It is known that such processes increase singular values. Such diabatic processes will be included in the operational singular vector computations when tropical singular vectors are added in 2001.

One may ask whether the baroclinic singular vector instabilities used in the EPS are the most appropriate ones to trigger the more barotropic instability apparently associated with variations in large-scale flow patterns, and hence medium-range predictability. By the time the singular vectors have evolved at D+2, they have acquired significant barotropic structure (*Buizza and Palmer, 1995*), and as such have the capability to trigger the barotropic regime instabilities. However, one could ask whether singular vector perturbations should be designed with respect to some "regime-trigger" optimality. For example, sensitivity perturbations with respect to regime patterns (cf. *Oortwijn and Barkmeijer, 1995; Palmer and Corti, 1997*) could be envisaged as EPS perturbations. One problem with this approach is the need to optimise with respect to a forecast range (e.g. D+5) where linear assumptions are wholly inappropriate.

The performance of the EPS, e.g. in producing enhanced spread during periods of negative PNA and NAO index, can be tested by performing an SVD analysis between the D+5 verification and ensemble spread; one would expect the same dependence as with an SVD analysis between D+5 verification and magnitude of error (see Fig 13). However, as mentioned at the beginning of this section, the non-stationarity of the EPS makes such an analysis problematic. Running a fixed EPS system on selected ERA-40 initial conditions might be worthwhile in this respect.

6. Systematic and flow-dependent error in the medium and extended-range

Medium-range systematic error has fallen significantly over the last 20 years. Fig 17 shows the D+10 500hPa height systematic error for DJF over the last 20 years, based on 5 year averages. From 1980-84 the mid-latitude systematic error was associated with excessive zonal flow across the Pacific, and across the east Atlantic into Europe. From 1985-89 this westerly bias continued over the Pacific, but weakened over the Euro-Atlantic sector, where the pattern of systematic error resembled somewhat the NAO. Since 1995, the magnitude of systematic error has been relatively weak, though a pattern correlation with its counterpart ten

years earlier, is evident. Why has the pattern of systematic error been so robust, even though the amplitude has reduced significantly?

Although now small in the medium range, systematic error is not negligible in the extended range. Fig 18 shows the seasonal-timescale systematic error (at T63 resolution) from four model versions that were operational during the period 1995-2000 (i.e. associated with panel d) of Fig 17). (The first panel in Fig 18 is for the version of the model used in both ERA-15 and PROVOST projects.) It can be seen that the qualitative character of these seasonal-timescale systematic errors has survived many model cycles and resembles the pattern of medium-range systematic error from models operational in the 1980s. Why should this be?

The fact that a particular model cycle has large systematic error on the seasonal timescale, but relatively insignificant systematic error in the medium range, might reflect the fact that the medium-range diagnostics were obtained from a model with higher resolution than the (T63) resolution used to integrate on the seasonal timescale. However, Fig 19 shows that for cycle 18r6 (used to make the ECMWF AMIP2 integrations), the seasonal-timescale systematic error is not fundamentally dependent on resolution. Why is this systematic error pattern so robust?

Comparison between the medium range and seasonal timescale systematic error suggests that in recent years the “onset” of systematic error has been delayed in the forecast range. Whilst once it started *ab initio*, now it appears beyond the medium range. What sort of mechanisms might be responsible for such a delayed onset?

In recent years, the focus in diagnosis of medium-range error has moved from systematic to flow dependent. Fig 20 shows the dominant singular vector pair from an SVD between 500hPa height verifying analyses, and D+7 forecast error, based on an analysis by *Ferranti et al.* (2000). The forecasts used in this analysis are based on the period winter 1992/93 to 1998/99. The pattern associated with the leading singular vector¹ from the verification is very similar (over the Euro-Atlantic sector) to the pattern of systematic error associated with the early versions of the model operational during the 1980s. Although this pattern does not correlate perfectly with the classical North Atlantic Oscillation pattern, we follow the terminology of *Ferranti et al.* (2000) and refer to it as an NAO pattern. By comparing the left and right singular vectors, it can be seen that at D+7, the forecast error is anti-correlated with the analysis anomaly; in other words the model tends to underpredict the verifying analysis anomaly.

Fig 21 shows the probability density function (pdf) of the analysis and the D+7 forecast projected onto the left singular vector shown in Fig 20 - the systematic error projected onto this pattern is also included in this pdf. It can be seen that there is an underestimation of the observed pdf particularly for negative values of the NAO-like pattern. This underestimation is one explanation for the anticorrelation between the left and right singular vectors.

However, there is a second explanation possible for this anticorrelation. Suppose the pdf of the analyses and the model agreed perfectly, but that there was no predictability in fluctuations in the singular vector. Then for any anomalous value of the verification, the most likely value of the forecast would be at the mode of the pdf.

1. As discussed in the previous footnote, the singular vectors arising from this type of statistical SVD analysis are unrelated to the dynamical singular vectors discussed in section 5.

This is just as plausible an explanation of the anticorrelation of the left and right singular vectors. But if the explanation lies in limited predictability, why should the SVD analysis pick out a pattern that was spatially correlated with the model systematic error?

In fact, these two explanations may not be completely independent of one another if fluctuations from positive to negative values of the left singular vector are associated with some sort of regime instability. Indeed regime instability could also help explain why systematic and flow-dependent errors have some correlation with the dominant modes of variability of the atmosphere. Below some speculative remarks on a possible such explanation.

Fig 22 is a schematic (mechanical) analogue of behaviour that is observed in the *Lorenz* (1963) model which has manifest regime behaviour (cf. *Molteni et al.* 1993; *Palmer* 1999). The stationary apparatus is bolted to a board which has weak frictional contact with a table. The regimes are represented by cups, the system state vector by the ball. After more or less coming to rest in one of the cups, the ball is dropped back into the funnel. The climatic pdf of the system arises from multiple drops; if the ball lies in a cup most of the time, then this pdf will be bimodal. The ridge onto which the ball is dropped represents a regime instability which limits the predictability of regime transitions. This limited predictability means that a flow dependent SVD analysis on two such systems (system A being "truth", system B being the "forecast model") would produce a singular vector in the direction between the regimes (the dominant EOF of the system), and with the forecast error singular vector anticorrelated with the analysis singular vector. The reason for this is simple: for an ensemble of cases where the system A ball is dropped and goes to the left hand cup system, there will be a fraction of cases where the system B ball goes to the right hand cup.

A fan is now introduced to represent a systematic error in system B's formulation. For strong enough fan speed (cf. 1980-85 ECMWF model formulation) the board slides along the floor, and system B drifts uniformly relative to system A *ab initio*. For weak forcing, (cf. 1995-1999 ECMWF model formulation) the forcing only has an impact when the ball is falling onto the ridge. In particular, whilst the ball is in the cup (i.e. during a period of regime residence) the system is insensitive to the erroneous forcing. Hence the influence of the erroneous forcing will not be strongly manifest in forecast experiments on timescales comparable with the regime residence time (cf. the so-called delay in onset of systematic error).

The climatic manifestation in system B's response to the weak forcing is to increase the pdf associated with one regime, and to decrease the pdf associated with the other regime. The systematic error of system B will therefore point in the direction of the system's dominant EOF, and is therefore correlated with the flow-dependent singular vector. This result is not sensitive to the direction of erroneous forcing (possibly explaining why different atmospheric forcings e.g. associated with tropical convection or orography, can give rise to similar responses). Which regime's pdf increases depends on the structure of the apparatus near the regime instability. The sensitivity of the systematic error to the erroneous forcing depends on the projection of the forcing onto the regime instability.

There is substantial evidence that the northern winter circulation has non-gaussian regime structure (following work by Ghil and coworkers; see also *Molteni et al.*, 1990). Fig 23 and 24 show regimes from a 45 year NCEP dataset based on a cluster analysis in a phase-space spanned by the leading two EOFs of the northern hemisphere (*Corti et al.*, 1999). Over the whole dataset, cluster A is the most populated cluster; indeed by using only recent data, cluster A is overwhelmingly the most populated cluster (consistent with independent

estimates of recent climate change; *Corti et al.*, 1999). Cluster D, on the other hand, is very much a secondary regime.

There is an approximate spatial correspondence between the systematic and flow-dependent error patterns and the cluster A and D anomalies, lending support to the notion that both the systematic and dominant flow-dependent error could be interpreted in terms of a model bias towards the most densely populated atmospheric regime, and away from the least densely populated regime. On the other hand, the cluster analysis in *Corti et al.* (1999) was performed not only over the whole hemispheric domain, but also in an extremely truncated subspace of phase space, and may therefore not be adequate for a more quantitative assessment of European systematic and flow dependent error from a regime perspective. A Euro-Atlantic regime analysis has been performed by *Chessa* (2000), motivated by earlier work of *Vautard* (1990). Six principal regimes were identified, shown in Fig 25. The population frequencies calculated from verifying analyses and the D+10 EPS control forecast for each of the six clusters taken over the 3 winters DJF 1996/97 to 1998/99 are given in Table3.

| | 1 | 2 | 3 | 4 | 5 | 6 |
|-----|------|------|------|------|------|------|
| Ana | 0.22 | 0.09 | 0.14 | 0.19 | 0.17 | 0.20 |
| EPS | 0.26 | 0.07 | 0.12 | 0.20 | 0.12 | 0.23 |

Table 3: Population frequency of the six clusters shown in Fig 24 over the winters 1996/7 to 1998/99 from the analysis (top row) and the D+10 EPS control forecast (bottom row)

It can be seen that clusters 1 and 6 (both of which are associated with enhanced zonal flow) are overpopulated at D+10, whilst cluster 5 (with an enhanced ridge over Europe) in particular is underpopulated.

We have so far discussed model error in terms of some erroneous constant forcing vector. However, it can be noted that a model could also show a bias in regime frequencies, overestimating dominant regimes and underestimating secondary regimes, if the model underestimates transient variability (see Fig 25; from *Molteni and Tibaldi*, 1990). Is there any evidence for such an underestimation in transient activity in the ECMWF model? Firstly, all operational versions of the ECMWF model simulate the intraseasonal Madden-Julian Oscillation (MJO) in the tropics very poorly, and this error is not improved with increasing resolution (*L. Ferranti*, personal communication). It is known that failure to simulate the MJO affects the ability of the model to simulate and predict large-scale regime transitions in the extratropics, particularly on timescales beyond 10 days (*Ferranti et al*, 1990). Although the amplitude of the MJO is largest in the warm-pool area of the tropical west Pacific, it is known that atmospheric perturbations in this region can influence European climate (*Palmer and Mansfield*, 1984; *Ferranti et al*, 1994)

Secondly, it is known that the ECMWF model fails to simulate the known $k^{-5/3}$ spectrum of variability (*Nastrom and Gage*, 1985) on scales up to several hundred kilometres; rather, the model has a k^{-3} (or steeper) spectrum down to the truncation limit (*M. Hortal*, personal communication). *Lilly* (1983, 1998) has speculated that quasi-two dimensional kinetic energy associated with mesoscale convective systems (MCSs) could propagate to larger scales, giving rise to such a shallow spectrum. It is likely that there may be other sources of small scale forcing which could generate an inverse cascade, orography being an obvious example.

7. Other topics

A number of possibly relevant topics are mentioned briefly below:

7.1 Data sensitivity, role of radiosondes

A number of 10-day forecasts have been made showing the impact of denying observing systems in the ECMWF data assimilation (*Bouttier and Kelly, 2000*). On the basis of these experiments, it was found that Europe is more sensitive than the northern hemisphere as a whole, to a loss of radiosonde and aircraft data. The sensitivity to satellite data is about the same. As such, Europe depends more on radiosonde data than on satellite data, and therefore benefits less from progress in the use of satellite data than other areas such as USA. As Europe relies a lot on radiosonde data, we may have been hurt more by the demise of the radiosonde network in recent years. Most notable is the decrease in coverage over Siberia. This is likely to hurt directly the medium-range forecast performance at 5-7 day range or later.

7.2 Storm-track diagnosis

Differences in the general circulation statistics between Europe and North America may be highlighted by new ways of representing storm-track activity. A new diagnostic, developed at Reading University, is based on the idea of meridional parcel displacement computed by dividing enstrophy by mean potential vorticity gradient. Meridional storm-track parcel displacements were found to be substantially larger over Europe than over North America.

7.3 The role of Greenland

Discussions of the poor scores in the summer 99, raised the question of whether Greenland plays a unique role in the general circulation. Numerical experiments have been performed at Reading University on the role of Greenland. It was found that the response of the general circulation to Greenland was quite different to the "canonical" orographic response (e.g. associated with the Rockies). In particular, removing Greenland produced an anomalous downstream low in time-mean surface pressure (located to the north of Scotland). The northward deflection of storm tracks to the lee of Greenland appears to be an important component of its effect on the mean circulation over Europe. Possibly Greenland has a significant influence on the frequency of occurrence of blocking-like flow regimes over Europe.

7.4 Targeted observations

A study has been made of the European Christmas storm period, and whether additional targeted observations could have made over the North Atlantic region that might have significantly improved predictive skill. This work is continuing, and will be reported elsewhere.

7.5 Medium-range evolution of SST

During Spring and Summer it is known that on a time scale of five days there may be considerable changes in SST in the North Atlantic area, which may have impact on the forecast. Since in our forecasting system SST is kept fixed, it was thought worthwhile to do experiments where SST is updated once a day using analysed sea surface temperature fields. In total 74 forecasts with CY21R4 were performed from the months of May,



August, September, October, November and December of 1999. Although according to the significance tests there was a significant improvement in the scores of 1000 mb geopotential over the European Area in the late medium range, the improvement was nevertheless very small (*Janssen, 2000*).

7.6 Surface pressure bias

Meteorological analysts have frequently reported systematic problems with surface pressure observations over the North American continent and from ships. There may be many reasons for the occurrence of systematic differences in observed and first-guess surface pressure, notable problems with model orography, local problems at stations, different observational practices, different ship height when loaded or not, etc. Rather than trying to solve these problems by doing a detailed analysis, it was thought worthwhile to follow the relatively straightforward approach of applying a surface pressure bias correction to all land stations, ships and drifting buoys. The surface pressure bias was obtained from two months of obs minus fg differences. Surface pressure observations were corrected using these biases and analyses and forecasts were produced. In total there are only 38 cases, 16 cases from a period in July and August 1999 and 22 cases from December 1999 and January 2000. Preliminary results on scores over the European area for 1000 mb geopotential height are positive. Although for a small area such as Europe there are not enough cases to be fully confident of the statistical significance, there is also an improvement over the Northern and Southern Hemisphere. More definite results will be reported by *Dragosavac, Janssen and Persson* (in progress).

7.7 Transplant analysis experiments

A number of cases of poor ECMWF forecasts have been rerun using the ECMWF model from UKMO analyses. In general results point to the importance of the initial conditions in determining forecast performance (consistent with earlier results). However, preliminary results from a recent poor period from ECMWF (July 2000) has suggested model sensitivity; this work is continuing.

8. Conclusions

A number of topics relating to the performance of the ECMWF operational forecasts over Europe have been discussed in this paper.

In section 2, it was shown that whilst rms forecast errors are generally larger over Europe than North America, anomaly correlation coefficients or persistence normalised rms measures of skill do not show significant differences between the two regions. Moreover, there is no objective evidence that recent developments in the ECMWF forecast system have had less impact over Europe than over other comparable parts of the northern hemisphere such as North America.

Seasonal-mean error doubling times, computed using the *Lorenz (1982)* parametrisation were discussed in section 3. Despite the fact that the summer flow is less baroclinically unstable, error doubling times are generally smaller in summer than in winter. There is significant interannual variability in values; for example, doubling times for winter 1998/99 and for spring and summer 1999 (a period of poor forecast performance), are relatively low.

In section 4, an analysis of 500hPa transient variability of the general circulation was made, motivated by interannual variations in the skill of a persistence forecast. On all time-ranges studied, the transient variability was larger over Europe than over North America, consistent with larger rms forecast errors over Europe. A measure of transient variability was defined (the ratio of band- to low-pass variance), whose interannual variability correlated well with interannual variability in the skill of a persistence forecast. On the other hand, there was no strong correlation between interannual variations in this measure of transient variability, and medium-range forecast skill scores.

To explore links between skill scores and circulation patterns, results were shown from an SVD analysis between the general circulation and the magnitude of forecast error. The forecast error pattern correlated most strongly with a pattern that projected onto the North Atlantic Oscillation (NAO). The forecast error pattern itself was not uniform over the European sector. For a negative NAO pattern, forecast error was relatively large over northern and southern Europe, and relatively small over a band including the UK. From a dynamical point of view, this indicates that Europe is not a particularly homogeneous region from which to study variability in area-mean skill scores.

Statistics of the dominant dynamical singular vectors, as computed in the EPS, were analysed in section 5. Singular values tend to be largest in regions of strong baroclinic instability, typically over the North Pacific and North Atlantic sectors. With the current dry formulation, singular values are smaller in summer than in winter. This is inconsistent with the fact that error doubling times are smaller in summer, and suggests that EPS singular vector computations, particularly in summer, should be based on moist dynamics, with a contribution to the initial norm arising from uncertainties in the analysis of moisture. Interannual variations in dominant singular values are partially correlated with error doubling times, but not with medium-range forecast skill.

Systematic and flow-dependent errors were discussed in section 6. Although systematic errors in the medium-range are now small, they are still significant on the seasonal timescale. Patterns of systematic error from recent seasonal model integrations resemble medium-range systematic error from model versions of the 1980s. This suggests that the general pattern of systematic error is robust, though its onset now occurs later in the forecast range. Current medium-range flow-dependent errors resemble the systematic model errors found in seasonal timescale integrations.

It has been suggested that a fruitful way to understand dynamically the interrelationship between flow-dependent and systematic error may be through an analysis of hemispheric and regional circulation regimes. A crucial aspect in this analysis appears to be the identification and quantification of the (finite-time) instabilities associated with regime transitions. Such instabilities are not simply related to the more baroclinic instabilities picked up by the EPS singular vector analysis. Preliminary work to try to analyse instabilities associated with regime instability has been undertaken by *Oortwijn and Barkmeijer (1995)* based on a Euro-Atlantic blocking pattern, and *Corti and Palmer (1997)* based on transitions related to the NAO and PNA patterns. In both these studies the sensitivity (i.e. gradient) of prescribed patterns with respect to perturbations 5 days earlier was found, using a 3-level quasi-geostrophic model (and its adjoint). Composites of low and high sensitivity transitions were made, and EOFs of the flow which distinguished between cases of low and high sensitivity were found. For example, the sensitivity in transitions in the NAO pattern were found to be related to the phase of the upstream PNA pattern - unpredictable transitions being associated with a negative PNA index.



There are a number of ways in which this work could be extended. Firstly, it would be useful to compute a pdf such as that shown in Fig 23 (and corresponding regimes as shown in Fig 24), based on integrations of the ECMWF model. Secondly, the use of the 6 Euro-Atlantic regimes found by Chessa (see Fig 25) would provide a more regional basis to the analysis of regime sensitivity. Finally, code has recently been developed (J. Barkmeijer, personal communication) to calculate so-called forcing singular vectors, i.e. quasi-constant forcing perturbations which have optimal impact on energy variance at some prescribed forecast time. Initial testing of this code is taking place in a 3-level quasi-geostrophic model (T. Iversen, personal communication), but code is also being developed for the IFS.

From a regime perspective, the character of the systematic error over the Euro-Atlantic region is still characterised by excessive population of the most densely populated zonal regimes. This is hardly significant on the 10-day timescale, but is more serious on the seasonal timescale (being comparable with the magnitude of interannual variability), and is likely to be problematic when 30-day ensemble forecasting is recommenced. An overpopulation of the dominant regimes is consistent with a model with weak transient variability, and it has been noted that one component of the general circulation which the ECMWF fails to simulate is the $k^{-5/3}$ spectrum observed on scales of hundreds and kilometres and less. If this spectrum is associated with an inverse cascade of mesoscale energy, as has been suggested, then some sort of stochastic parametrisation of such mesoscale variability may be needed (Palmer, 2000). To date, there is no evidence that the current stochastic physics in the EPS has a significant impact on systematic error, though modifications to this scheme based on explicit potential vorticity forcing (based on a project in collaboration with the Meteorological Office) may lead to different results.

A number of further topics relevant to the issues discussed in this paper were summarised in section 7.

In general, studies of the predictability of Euro/Atlantic flow can be put on a firmer footing with the ERA-40 dataset. Such a dataset will allow a more rigorous definition of (3-dimensional) regime structure, and forecasts (including EPS forecasts) run from these analyses will allow variations in predictability to be studied without the problem of a changing forecast system.

References

- Bouttier, F. and G. Kelly, 2000: Observing system experiments in the ECMWF 4D-Var system. *Proceedings of the second CGC/WMO workshop on the impact of observing systems on NWP*, Toulouse, France, 6-8 March 2000.
- Brankovic, C. and T.N.Palmer, 2000: Estimates of seasonal predictability and predictive skill from ECMWF PROVOST ensemble integrations. *Quart.J.Roy.Meteor.Soc.*, **126**, 2035-2067
- Brankovic, C. and D. Gregory, 2000: Impact of horizontal resolution on seasonal integrations. *Clim. Dynamics*. Submitted.
- Buizza, R. and T.N.Palmer, 1995: The singular vector structure of the atmospheric global circulation. *J. Atmos.Sci.*, **52**, 1434-56.
- Chessa, P., 2000: Classification and verification of the ECMWF EPS perturbed forecasts using pre-defined weather regimes. *ECMWF Operations Department Internal Memorandum (30/7/2000)*.

- Corti, S., F. Molteni and T.N.Palmer, 1999: Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature*, **398**, 799-802
- Corti, S. and T.N.Palmer, 1997: Sensitivity analysis of atmospheric low-frequency variability. *Quart.J.Roy.Meteor.Soc.*, **123**, 2425-2447.
- Crutcher, H.L., and J.M. Meserve, 1970: Selected level heights, temperatures and dew points for the Northern Hemisphere. *NAVAIR Atlas 50-IC-52*, Chief Naval Operations, Washington DC, 132pp.
- Ferranti, L., T.N.Palmer, F.Molteni and E.Klinker, 1990: Tropical-extratropical interaction associated with the 30-60 day oscillation, and its impact on medium and extended-range predictability. *J. Atmos.Sci.*, **47**, 2177-2199.
- Ferranti, L., F. Molteni and T.N.Palmer, 1994: Impact of localised tropical and extratropical SST anomalies in ensembles of seasonal GCM integrations. *Quart. J.R.Meteor. Soc.*, **120**, 1613-1645.
- Ferranti, L., E.Klinker, A.Hollingsworth and B.J.Hoskins, 2000: Diagnosis of systematic errors dependent on flow pattern. To be submitted.
- Janssen, P.A.E.M., 2000: Impact of SST changes on the atmospheric forecast. *ECMWF Research Department Internal Memo R60.8/PJ/52*.
- Lilly, D.K., 1983: Stratified turbulence and the mesoscale variability of the atmosphere. *J.Atmos.Sci.*, **40**, 749-761.
- Lilly, D.K., G.Bassett, K.Doegemeier and P. Bartello, 1998: Stratified turbulence in the atmospheric mesoscales. *Theor. Comput. Fluid Dynamics*, **11**, 139-153.
- Lorenz, E.N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505-513.
- Miller, M., M. Hortal and C. Jakob, 1995: A major operational forecast model change. *ECMWF Newsletter*, **83**, 2-8.
- Molteni, F., Tibaldi, S. and T.N.Palmer, 1990: Regimes in the wintertime circulation over northern extratropics. I: Observational evidence. *Q.J.R. Meteorol.Soc.*, **116**, 31-67.
- Molteni, F. and S. Tibaldi, 1990: Regimes in the wintertime circulation over northern extratropics. II: Consequences for dynamical predictability. *Quart. J. R. Meteorol. Soc.*, **116**, 1263-1288.
- Molteni, F. L.Ferranti and T.N.Palmer, 1993: A dynamical interpretation of the global response to equatorial Pacific SST anomalies. *J. Clim.*, **6**, 777-795.
- Morcrette, J.-J., 1993: Revision of the clear-sky and cloud radiative properties in the ECMWF model. *ECMWF Newsletter*, **61**, 3-14.

Nastrom, G.D. and K.S. Gage, 1985: A climatology of atmospheric wavenumber spectra observed by commercial aircraft. *J. Atmos.Sci.*, **54**, 435.

Nieminen, R., 1983: Operational verification of ECMWF forecast fields and results for 1980-1981. *ECMWF Tech. Rep.*, **36**, 48pp.

Oortwijn, J. and J. Barkmeijer, 1995: Perturbations that optimally trigger weather regimes. *J. Atmos.Sci.*, **52**, 3932-3944.

Palmer, T.N., 1988: Medium and extended range predictability and stability of the Pacific North American mode. *Q. J. R. Meteorol. Soc.*, **114**, 691-713.

Palmer, T.N., 1999: A nonlinear dynamical perspective on climate prediction. *J. Clim.*, **12**, 575-591.

Palmer, T.N., 2000: A nonlinear dynamical perspective on model error: a proposal for nonlocal stochastic-dynamic parametrisation in weather and climate prediction models. *Q. J. R. Meteor.Soc.*, submitted.

Palmer, T.N. and D.A.Mansfield, 1984: Response of two atmospheric general circulation models to sea-surface temperature anomalies in the tropical East and West Pacific. *Nature*, **310**, 483-485.

Palmer, T.N. and S. Tibaldi, 1986: Forecast skill and predictability. *ECMWF Technical Memorandum No. 127*

Ritchie, H., C. Temperton, A. Simmons, M. Hortal, T. Davies, D. Dent, and M. Hamrud, 1995: Implementation of the semi-Lagrangian method in a high resolution version of the ECMWF forecast model. *Mon. Wea. Rev.*, **123**, 489-514.

Simmons, A.J., 1988: Barotropic instability, and anomalies of the extratropical northern hemisphere winter circulation. *Pontificiae Academiae Scientiarum Scripta Varia*, **69**, 437-468.

Simmons, A.J., R. Mureau and T. Petroliaqis, 1995: Error growth and predictability estimates for the ECMWF forecasting system. *Quart. J. Roy. Meteor. Soc.*, **121**, 1739-1771.

Simmons, A.J., 1996: The skill of 500hPa height forecasts. *Proceedings of 1995 ECMWF Seminar on Predictability, Vol. 1*, 19-68.

Simmons, A.J., 2000: Objective verification of deterministic forecasts. *Proceedings of 1999 ECMWF Seminar on Diagnosis of Models and Data Assimilation Systems*.

Taljaard, J.J., H. Van Loon, H.L. Crutcher and R.L. Jenne, 1969: Climate of the Upper Air: Southern Hemisphere. Vol. 1, Temperatures, Dew Points and Heights at Selected Pressure Levels. *NAVAIR Atlas 50-IC-55*, Chief Naval Operations, Washington DC, 135pp.

Vautard, R., 1990: Multiple weather regimes over the North Atlantic. Analysis of precursors and successors. *Mon. Wea. Rev.*, **138**, 2056-2081.

Annex 1: Effect of erroneous climatology on anomaly correlation

Let c denote a true climatology, giving anomaly correlation coefficient AC and c' denote an erroneous or unrepresentative climatology, giving anomaly correlation coefficient AC' . Let f denote a forecast value and a the corresponding value of the verifying analysis, with an overbar denoting an average over area and over a sample of forecasts sufficiently large that the averages of the analyses and forecasts (assumed essentially unbiased) are negligibly different from the true climatology.

The difference in the climatologies is assumed to be uncorrelated with the analysed and forecast anomalies computed with respect to the true climatology. Then

$$\overline{(f - c')(a - c')} \approx \overline{(f - c)(a - c)} + \overline{(c - c')^2}$$

$$\overline{(f - c')^2} \approx \overline{(f - c)^2} + \overline{(c - c')^2}$$

and

$$\overline{(a - c')^2} \approx \overline{(a - c)^2} + \overline{(c - c')^2}$$

If mean errors are unimportant the anomaly correlation with respect to the erroneous climatology is given by

$$AC' \approx \frac{\overline{(f - c')(a - c')}}{\sqrt{\overline{(f - c')^2}} \sqrt{\overline{(a - c')^2}}} \approx AC + (1 - AC) \frac{\overline{(c - c')^2}}{A^2} \quad (1)$$

where A^2 is the variance of the sample, either $\overline{(a - c)^2}$ or $\overline{(f - c)^2}$, as it is assumed that the forecast model gives an accurate simulation of variance as well as the climatological mean. It is also assumed in [\(1\)](#) that the mean-square error of the erroneous climatology is small compared with the variance of the sample: $\overline{(c - c')^2} \ll A^2$.

Equation (A.1) shows that the anomaly correlation with respect to the erroneous climatology, AC' , is larger than the true anomaly correlation, AC . Using the 1981-1995 mean of the operational ECMWF analyses for c and the older NCAR climatology used in operational verification for c' it is found for the extratropics of the Northern Hemisphere that

$$\frac{AC' - AC}{1 - AC} \approx 0.08$$

This corresponds to an absolute difference of 3% in correlations at about the 60% level and 5% at about the 40% level, as illustrated by *Simmons* (1996).

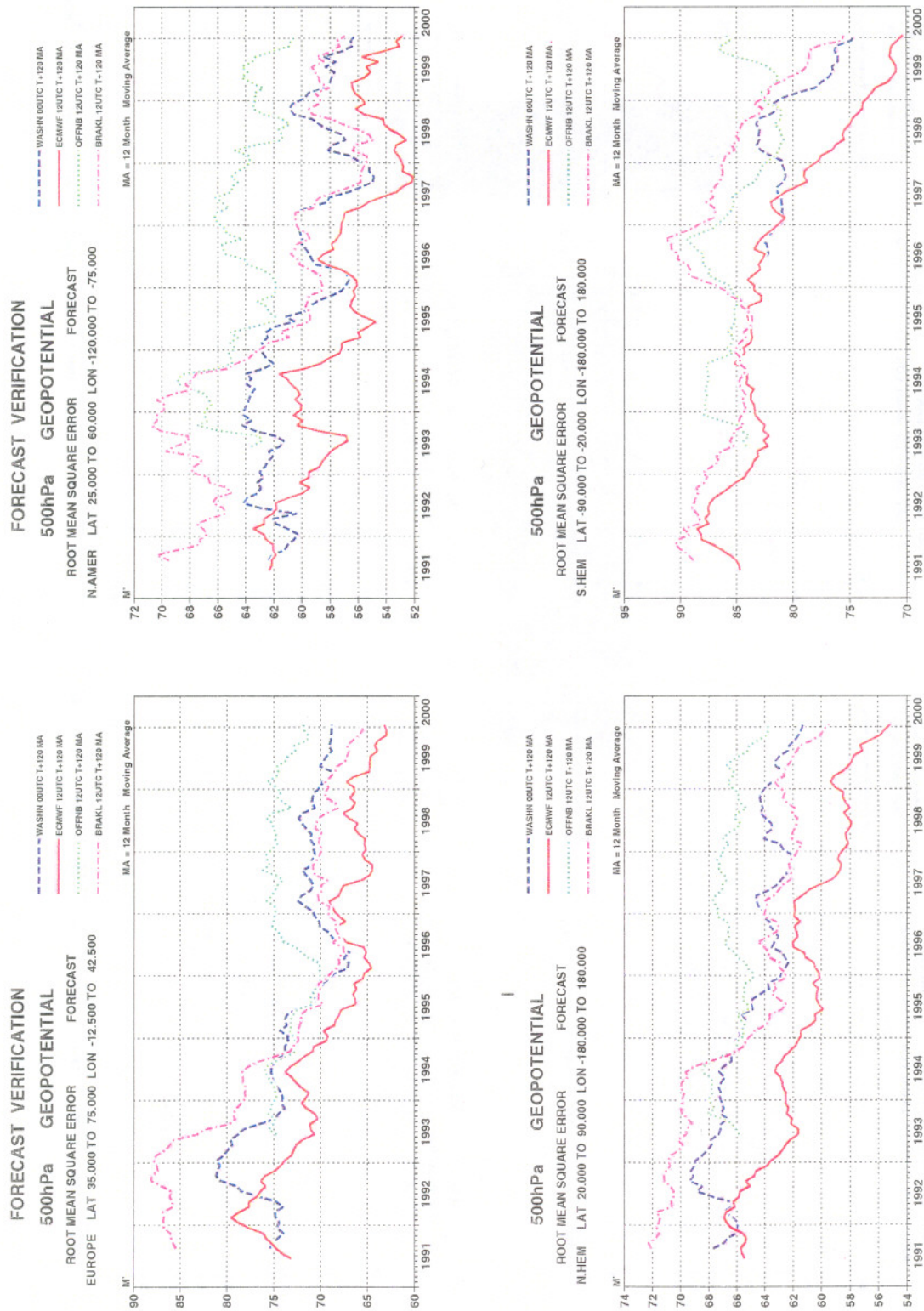


Figure 1: Time series of root-mean-square errors of day-5 operational 500hPa height forecasts from ECMWF (red solid), NCEP (blue dashed), DWD (green dotted) and the Met. Office (magenta chained), for Europe (upper left), North America (upper right), the Northern Hemisphere (lower left) and the Southern Hemisphere (lower right). Monthly-mean scores available since 1991 are smoothed by applying 12-month running means.

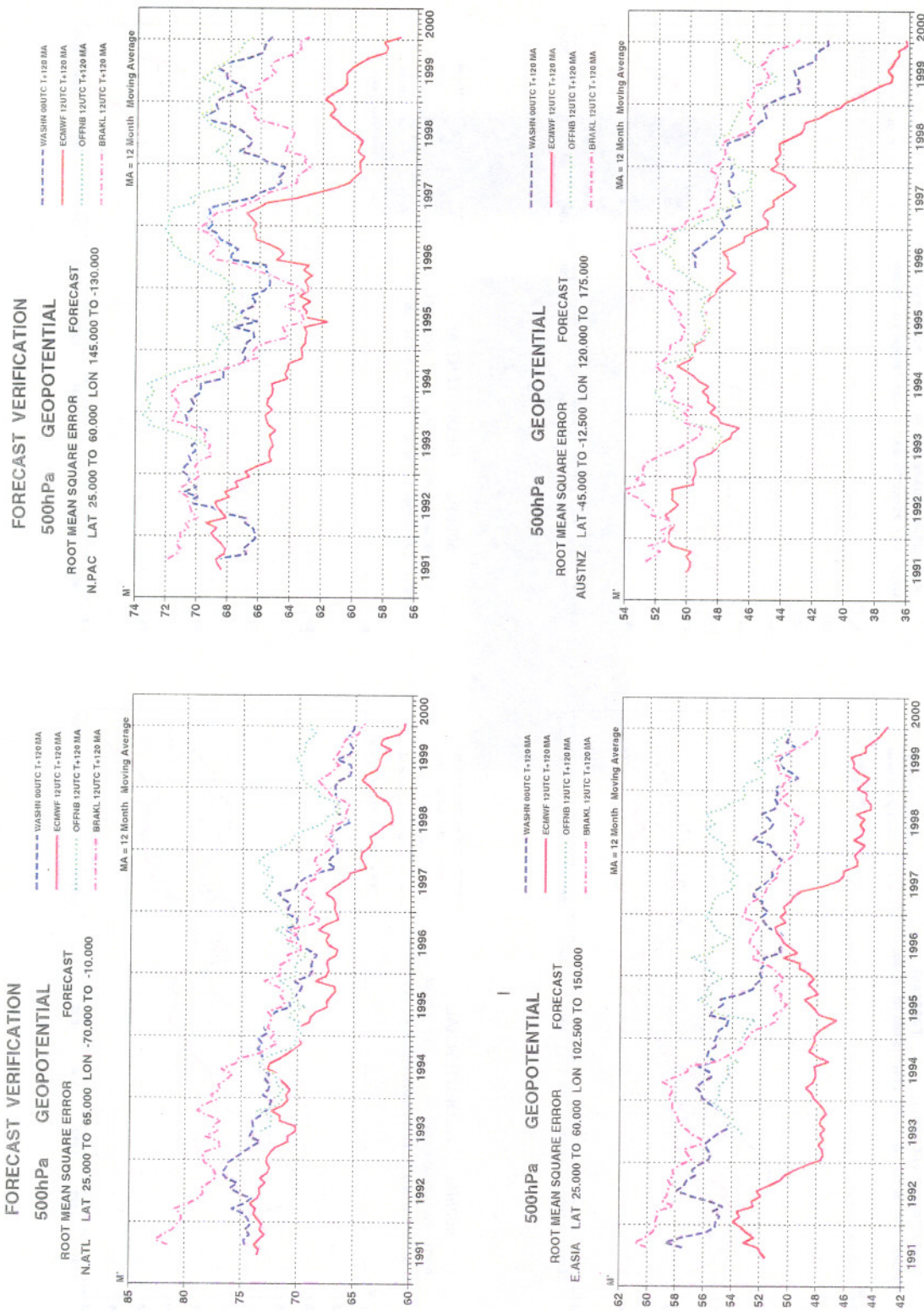


Figure 2: Time series of root-mean-square errors of day-5 operational 500hPa height forecasts from ECMWF (red solid), NCEP (blue dashed), DWD (green dotted) and the Met. Office (magenta chained), for the North Atlantic (upper left), North Pacific (upper right), East Asia (lower left) and Australia/New Zealand (lower right). Monthly-mean scores available since 1991 are smoothed by applying 12-month running means.

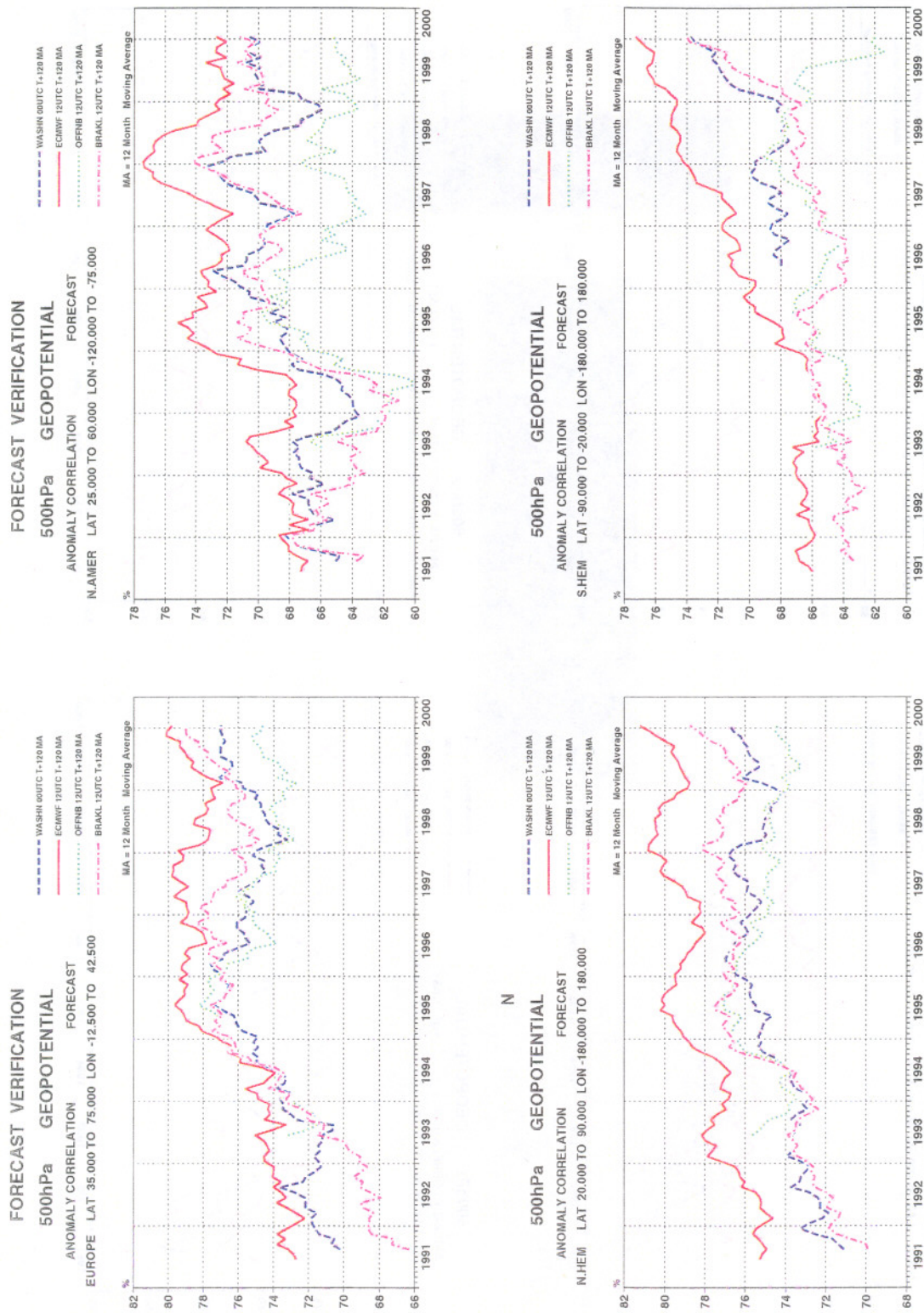


Figure 3: Time series of anomaly correlations of day-5 operational 500hPa height forecasts from ECMWF (red solid), NCEP (blue dashed), DWD (green dotted) and the Met. Office (magenta chained), for Europe (upper left), North America (upper right), the Northern Hemisphere (lower left) and the Southern Hemisphere (lower right). Monthly-mean scores available since 1991 are smoothed by applying 12-month running means.

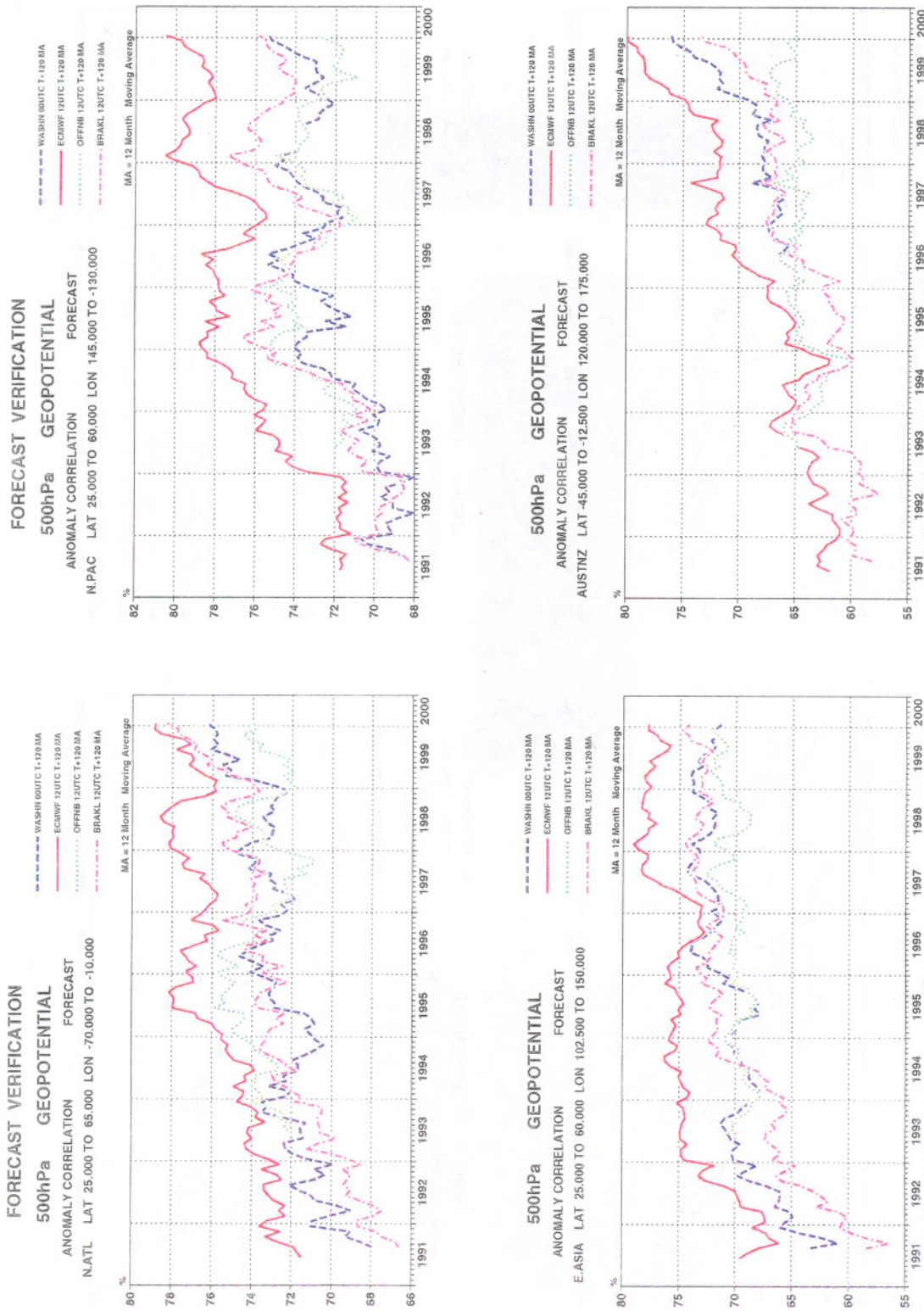


Figure 4: Time series of anomaly correlations of day-5 operational 500hPa height forecasts from ECMWF (red solid), NCEP (blue dashed), DWD (green dotted) and the Met. Office (magenta chained), for the North Atlantic (upper left), North Pacific (upper right), East Asia (lower left) and Australia/New Zealand (lower right). Monthly-mean scores available since 1991 are smoothed by applying 12-month running means.

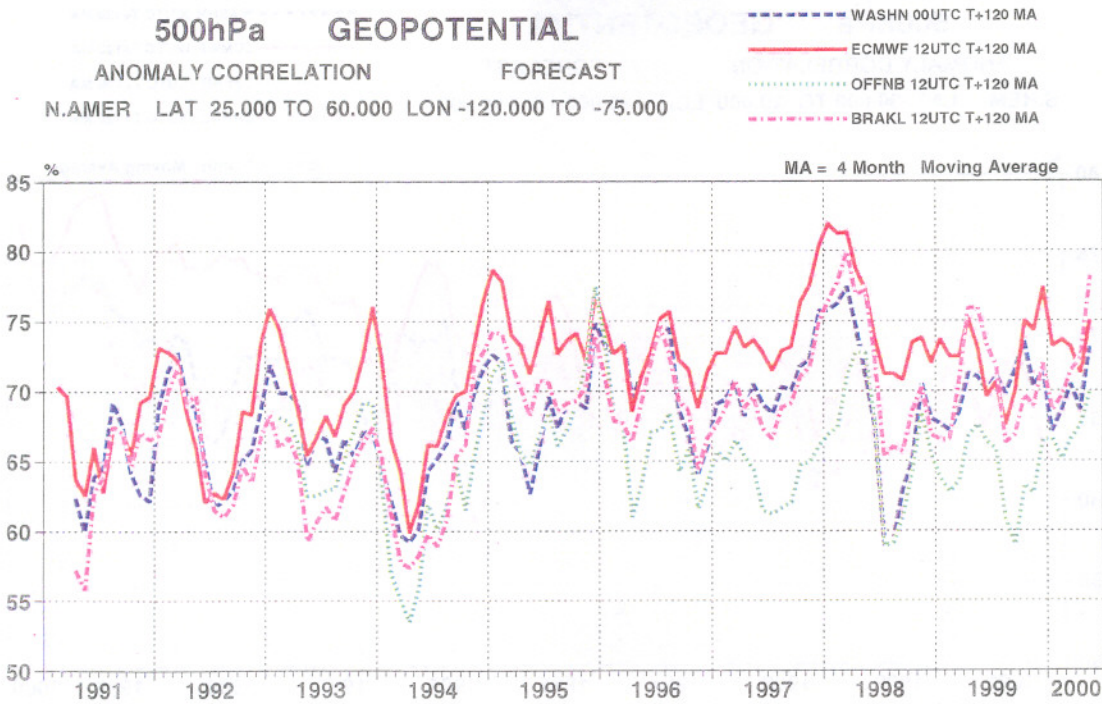
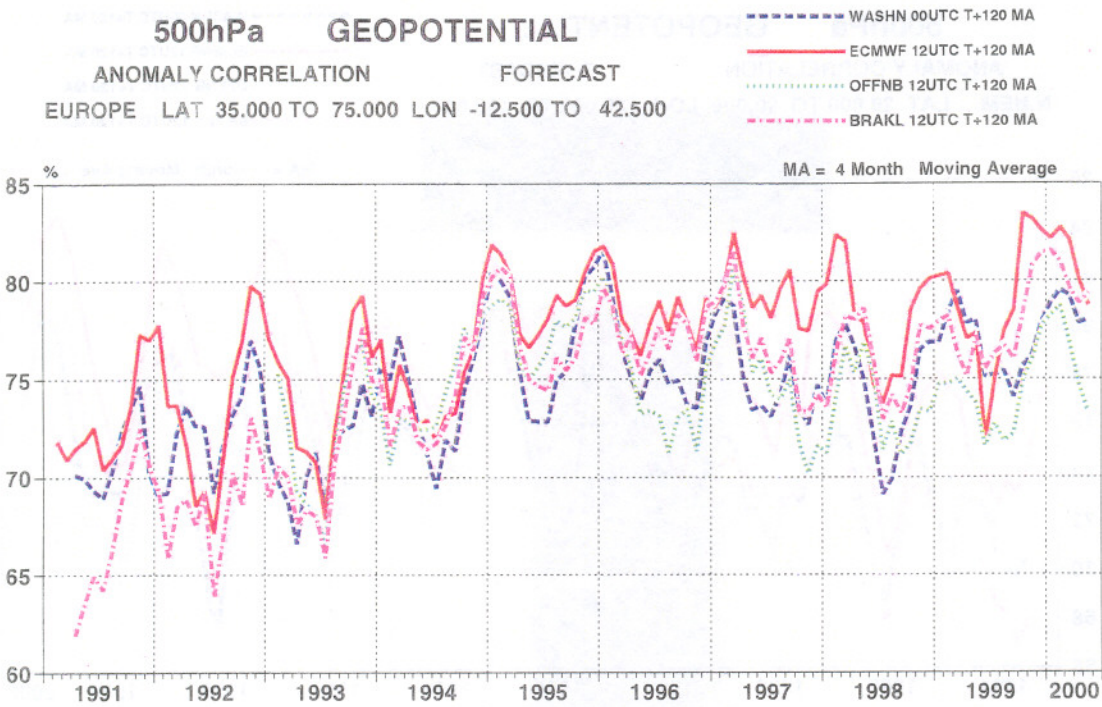


Figure 5: Time series of anomaly correlations of day-5 operational 500hPa height forecasts from ECMWF (red solid), NCEP (blue dashed), DWD (green dotted) and the Met. Office (magenta chained), for Europe (upper) and North America (lower). Monthly-mean scores available since 1991 are smoothed by applying 4-month running means.

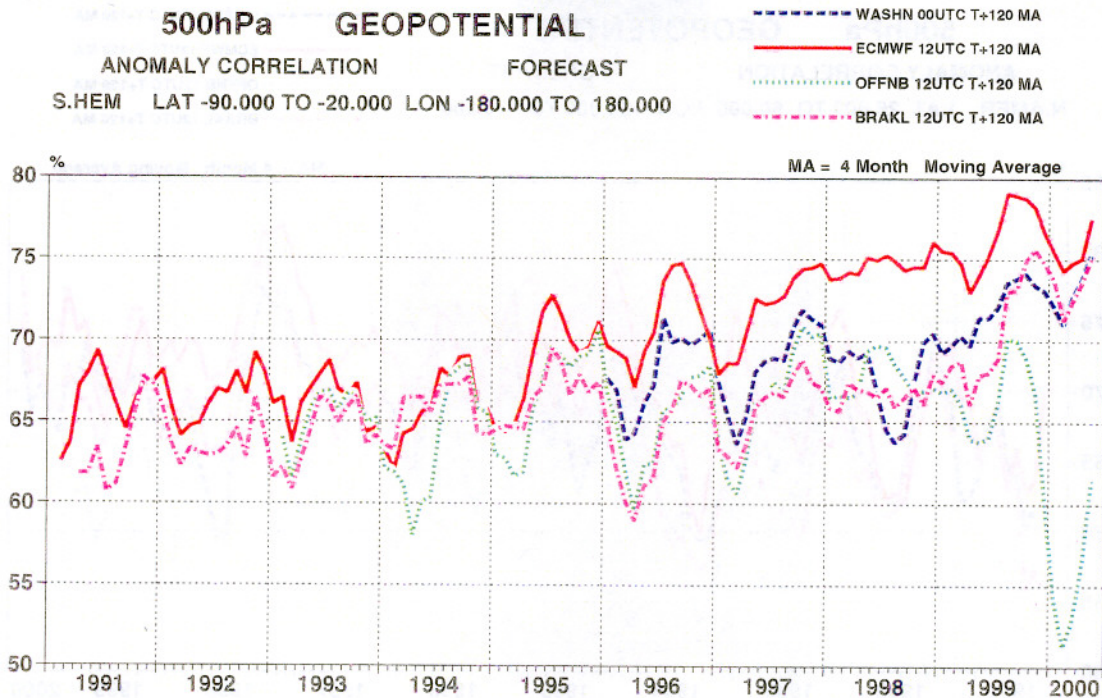
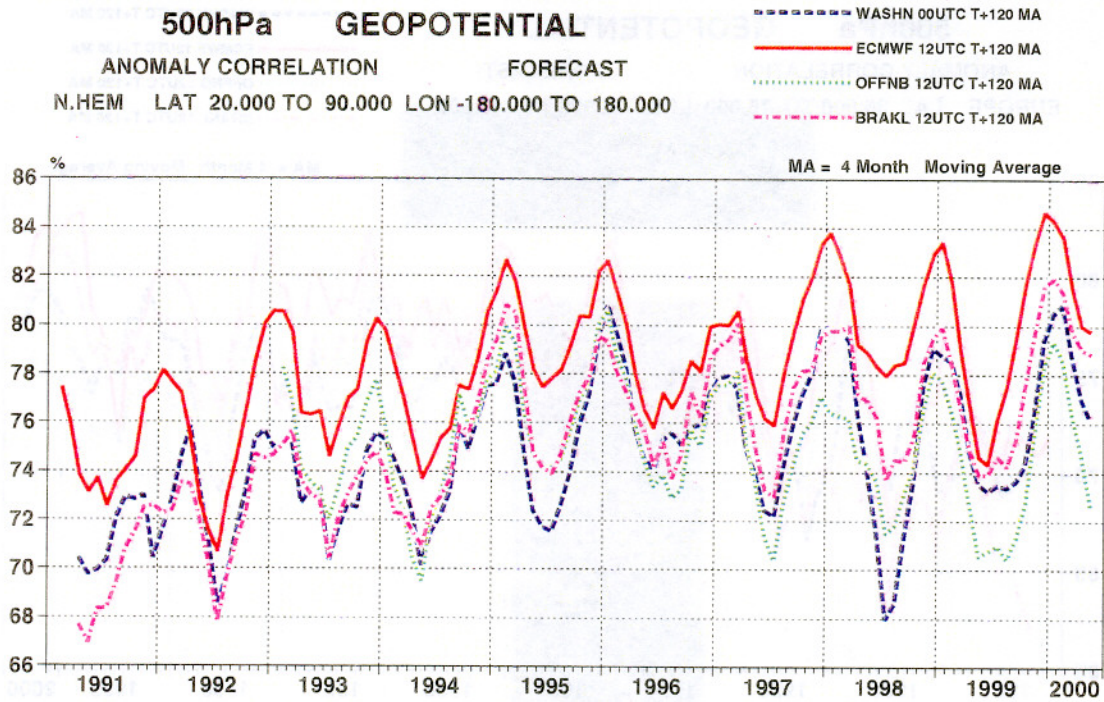


Figure 6: Time series of anomaly correlations of day-5 operational 500hPa height forecasts from ECMWF (red solid), NCEP (blue dashed), DWD (green dotted) and the Met. Office (magenta chained), for the Northern Hemisphere (upper) and the Southern Hemisphere (lower). Monthly-mean scores available since 1991 are smoothed by applying 4-month running means.

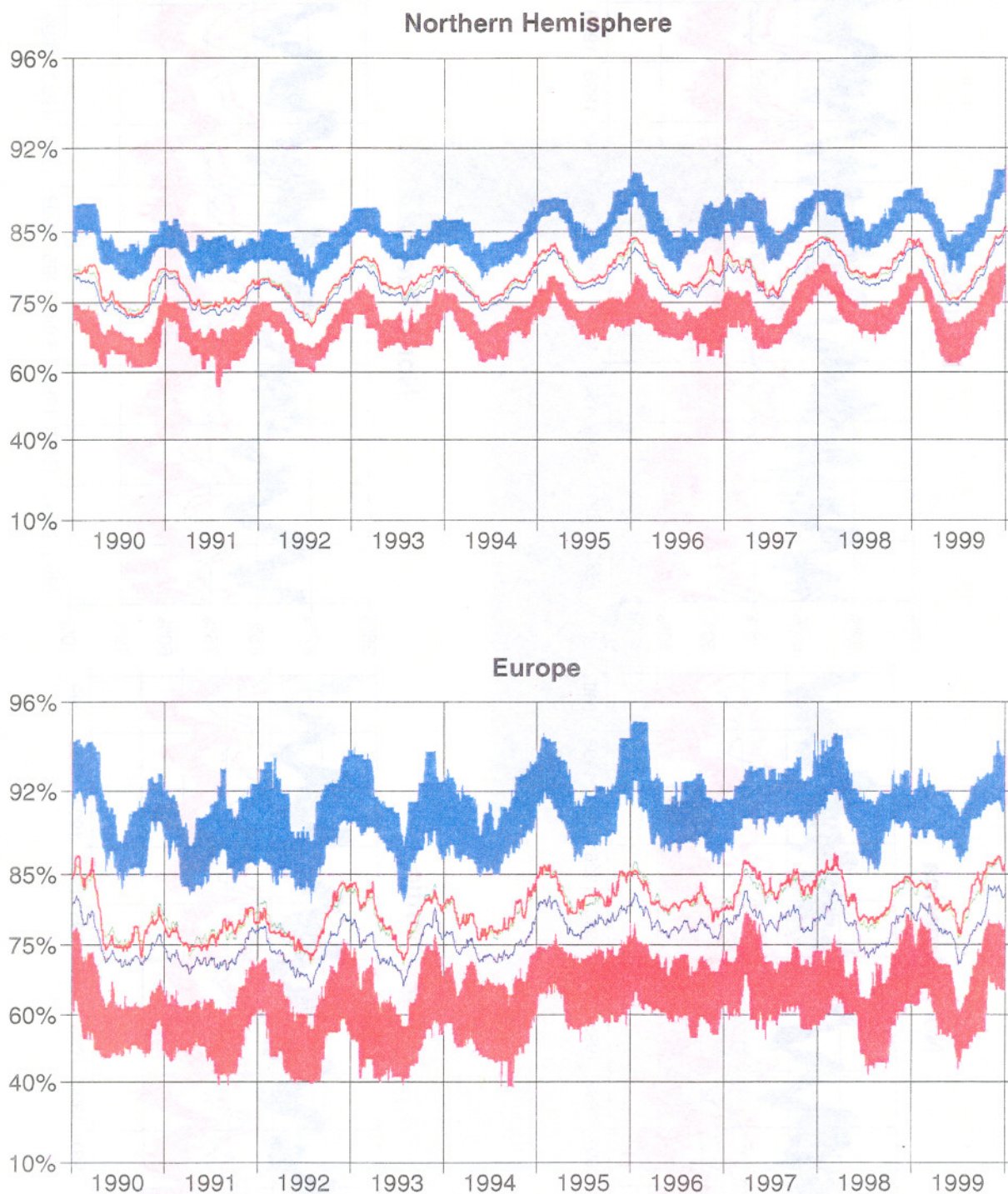


Figure 7: Time series of anomaly correlations of day-5 operational ECMWF 500hPa height forecasts from 1990 to 1999 for the Northern Hemisphere (upper) and Europe (lower). The red line shows the median of running 120-day sequences of daily correlations and the blue line the corresponding 120-day running (arithmetic) means of the correlations. The pink shaded band indicates the range of correlations which lie between the 10th and 25th percentiles of the 120-day distributions. The light blue band shows the range between the 75th and 90th percentiles. The ordinates are linear in the Fisher transform of the anomaly correlation. The green line (partly hidden by the red line) is derived from means of the Fisher transforms of the correlations.

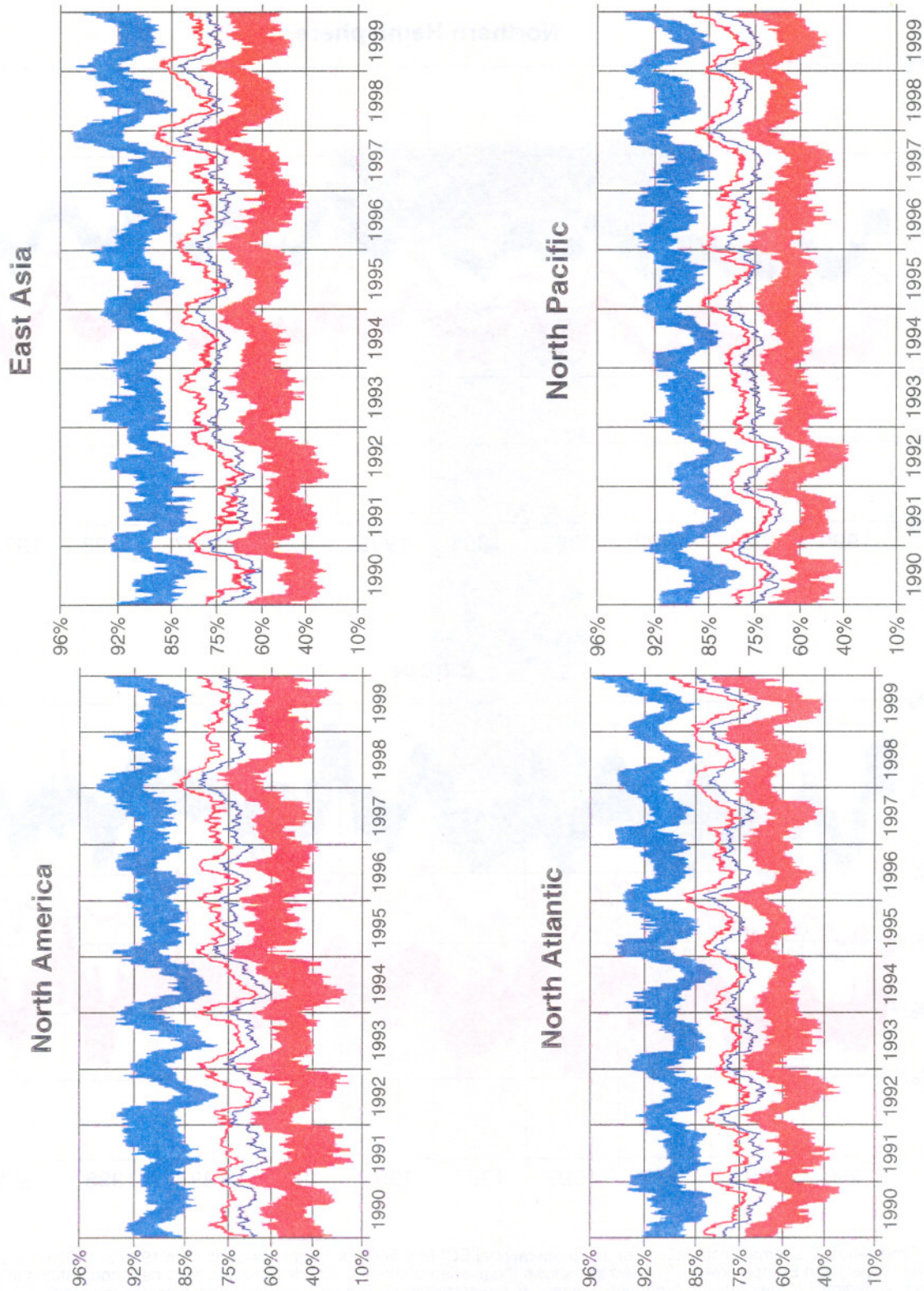


Figure 8: Time series of anomaly correlations of day-5 operational ECMWF 500hPa height forecasts from 1990 to 1999 for North America (upper left), East Asia (upper left), the North Atlantic (lower left) and the North Pacific (lower right). The red line shows the median of running 120-day sequences of daily correlations and the blue line the corresponding 120-day running (arithmetic) means of the correlations. The pink shaded band indicates the range of correlations which lie between the 10th and 25th percentiles of the 120-day distributions. The light blue band shows the range between the 75th and 90th percentiles. The ordinates are linear in the Fisher transform of the anomaly correlation.

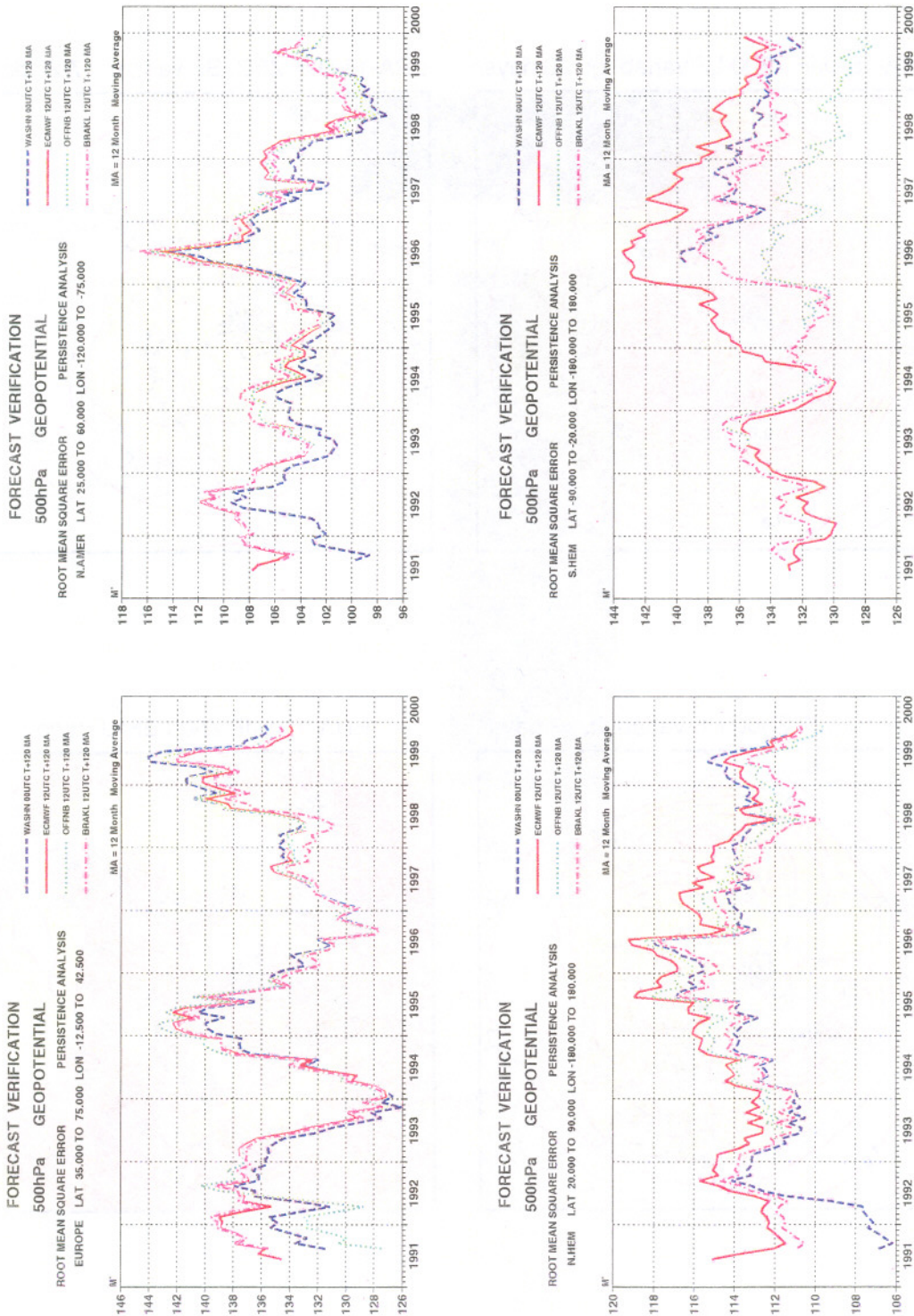
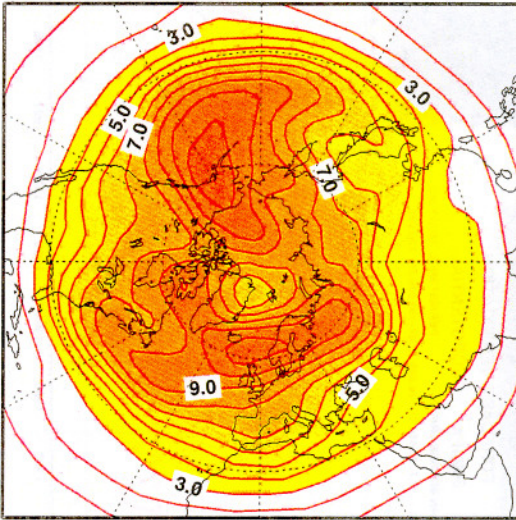
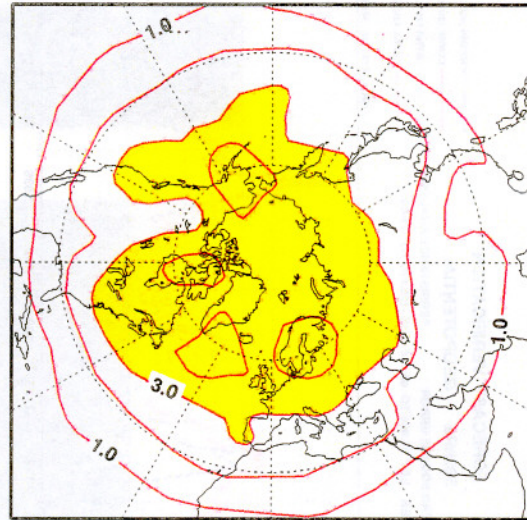


Figure 9: RMS error associated with D+5 persistence forecasts for Europe, North America, Northern Hemisphere and Southern Hemisphere, based on 500hPa height from different operational analyses.

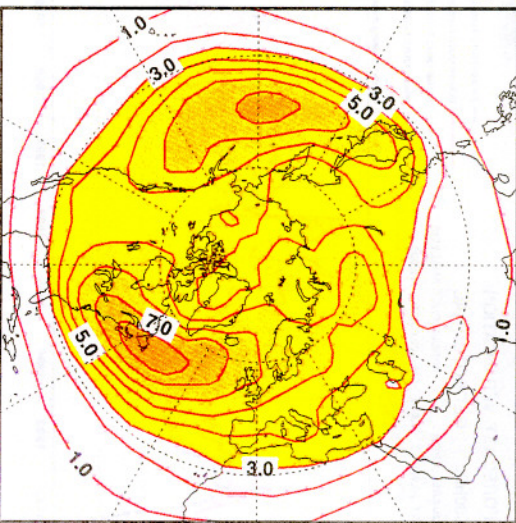
ERA-15 DJF Z500 LF variab. gt. 20 days



ERA-15 DJF Z500 BF variab. 7-20 days



ERA-15 DJF Z500 HF variab. lt. 7 days



ERA-15 DJF Z500 BF/LF ratio

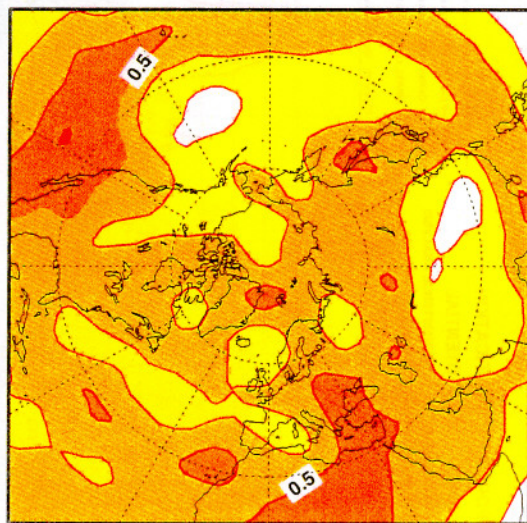


Figure 10: Diagnosis of DJF transient variability of 500hPa height from ERA-15. Top left: low-frequency variability (timescales greater than 20 days). Top right: band-pass variability (between 7 and 20 days). Bottom left: high-pass variability (less than 7 days). Bottom right: ratio of band-pass/low-pass variability.

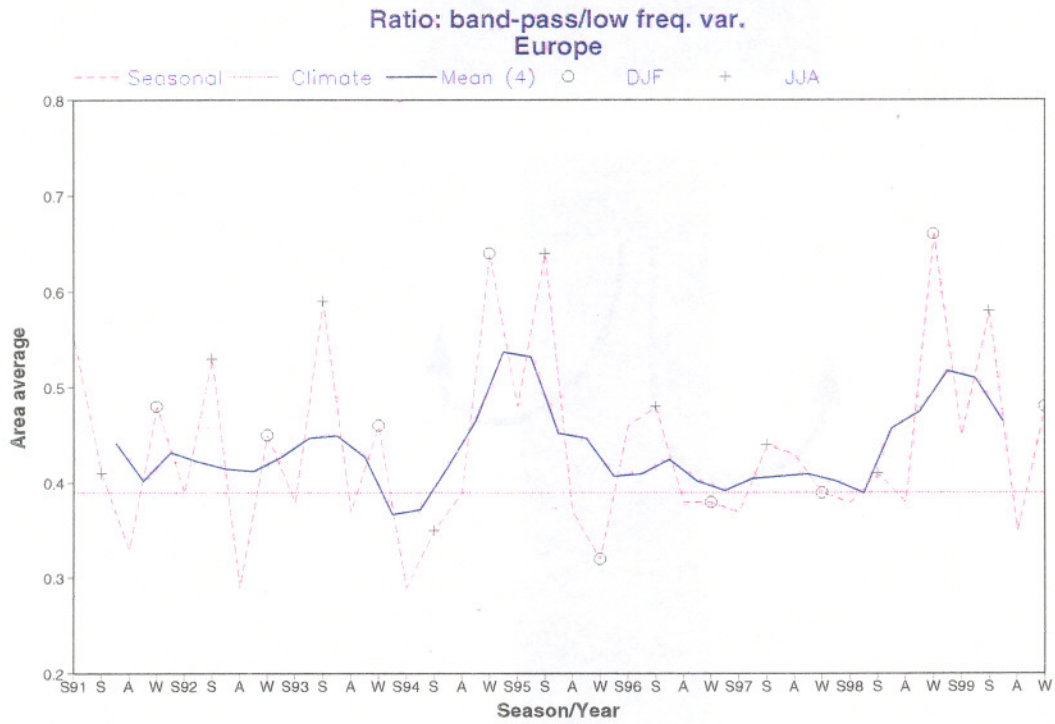


Figure 11: Timeseries of ratio of band pass to low pass transient variability based on ECMWF analyses. Summer and winter values shown in red line; four point running mean shown as blue line.

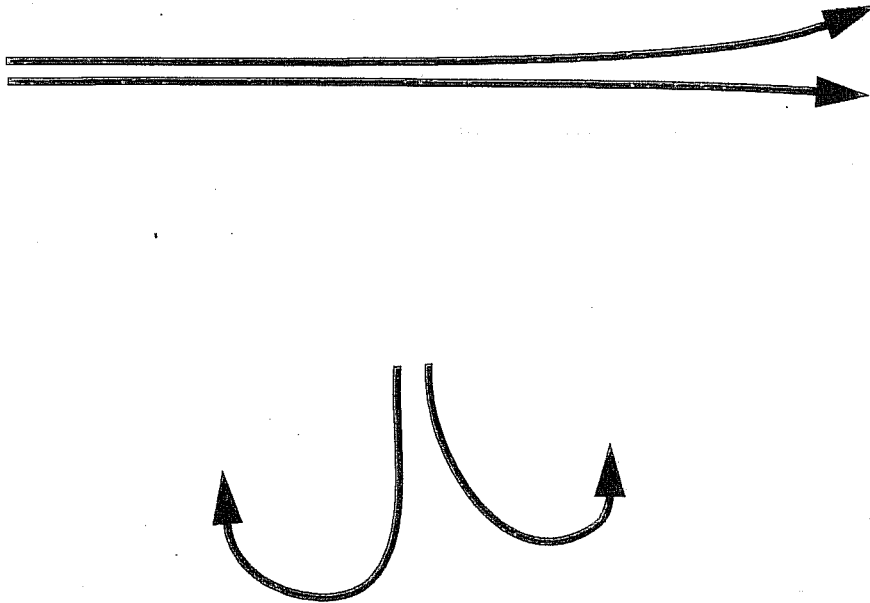


Figure 12: Schematic (based on actual phase-space trajectories from the Lorenz, 1963, attractor) showing that predictability need not be well estimated by the skill of persistence. Two pairs of trajectories are shown. For each pair, let one trajectory be "truth", the other be the "forecast". For the upper pair, the forecast error is relatively small, though the error of a persistence forecast is relatively large.

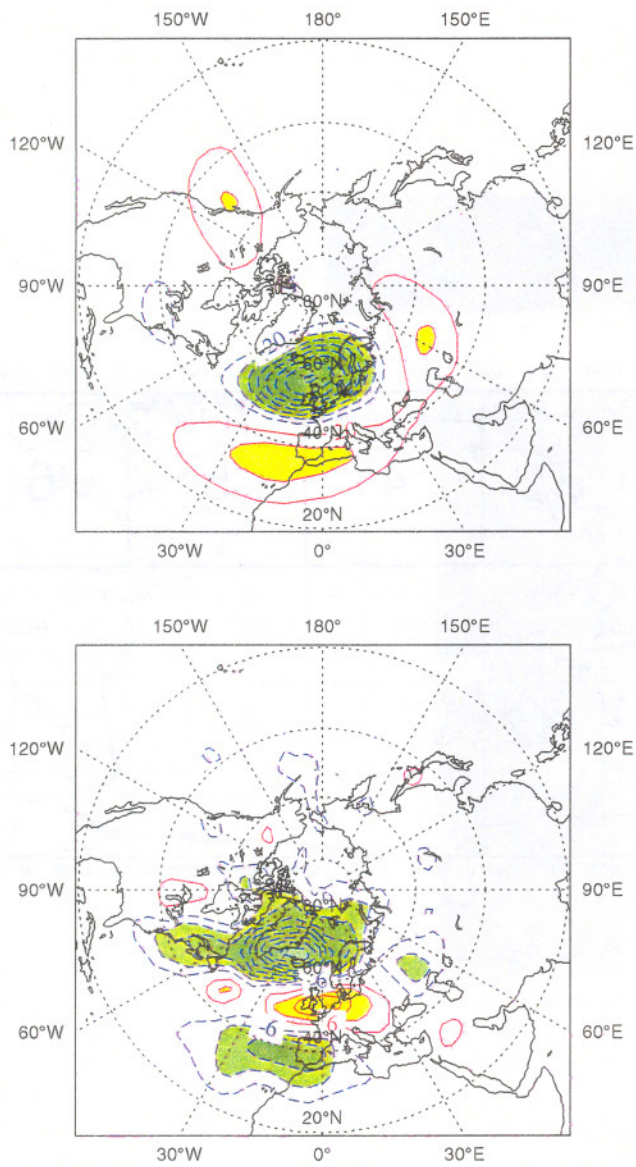


Figure 13: Leading pairs of singular vectors associated with a regional SVD applied to 500hPa height anomalies of the verifying analysis and the magnitude of 500hPa height forecast error at D+7 (contour 10m). From Ferranti et al, 2000.

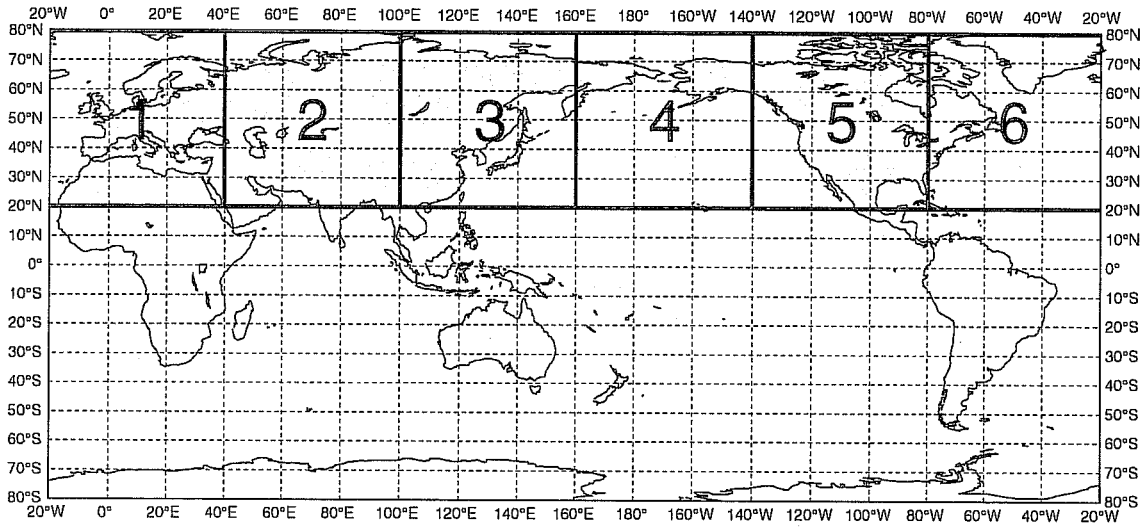


Figure 14: Position of the 6 regions used to categorise the operational dynamical singular vectors. (1=Europe, 2=Central Asia, 3=East Asia, 4=Pacific, 5=North America, 6=Atlantic).

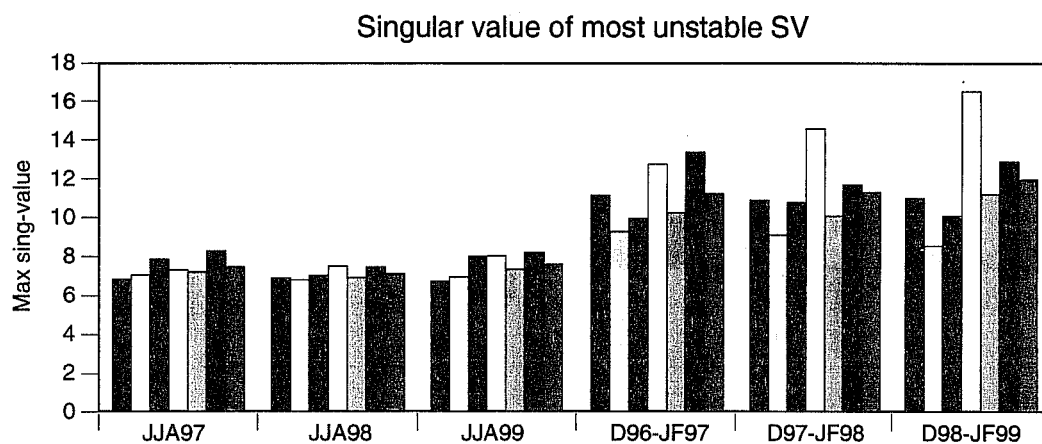


Figure 15: Seasonal-mean average of the dominant singular value for the regions shown in Fig 14. The bars from left to right are for the regions: Europe, Central Asia, East Asia, Pacific (white bar), North America, Atlantic, average over the 6 regions.

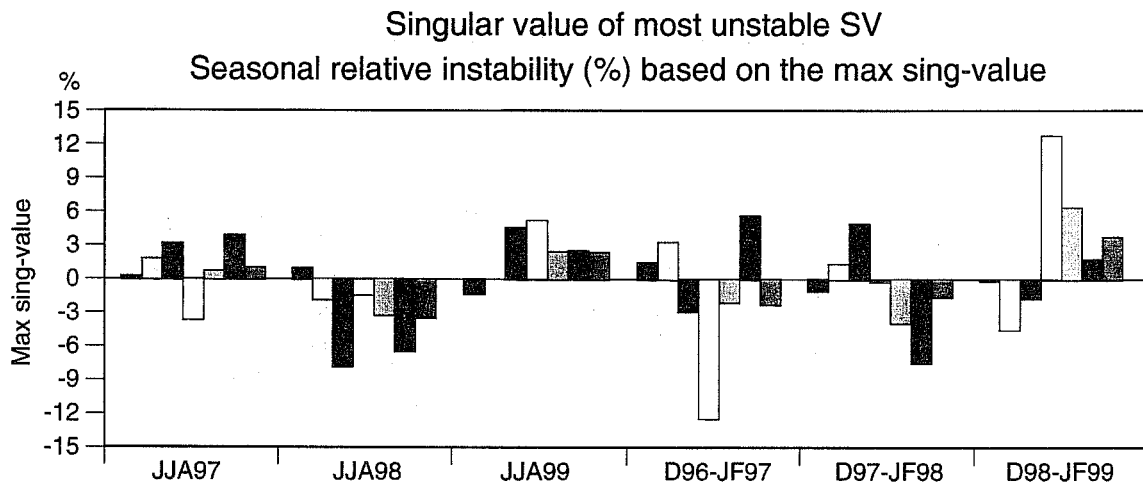
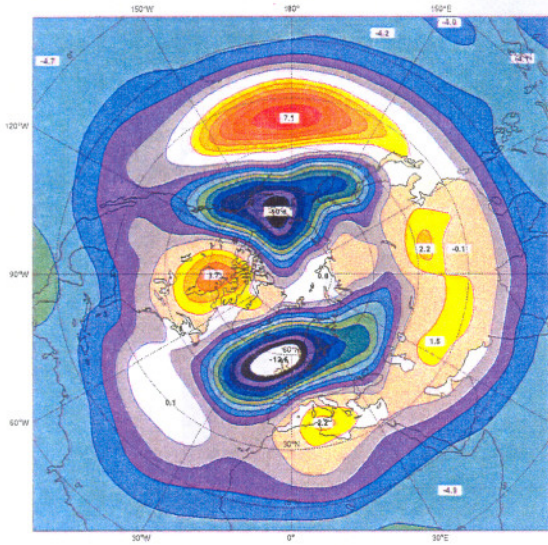
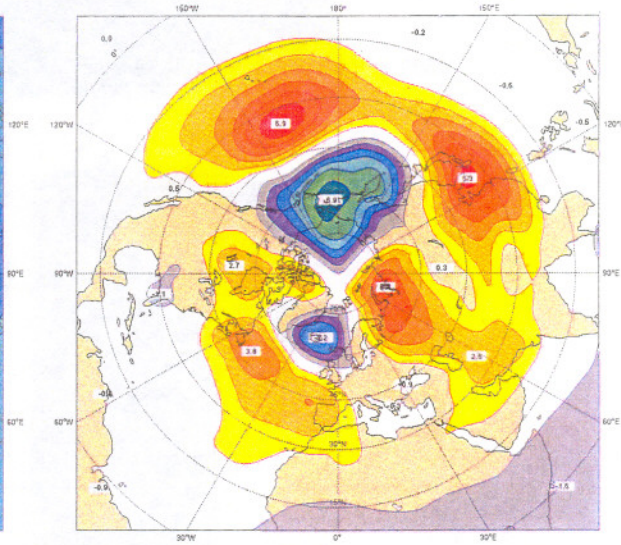


Figure 16: As figure 15 but expressed as a percentage anomaly. The bars from left to right are for the regions: Europe, Central Asia, East Asia, Pacific (white bar), North America, Atlantic, average over the 6 regions.

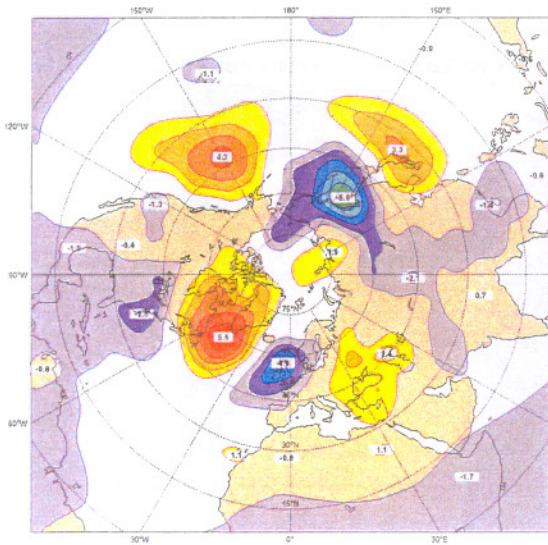
Z500 error D10 DJF 1980-1984; contours every: 1 (dam)



Z500 error D10 DJF 1985-1989; contours every: 1 (dam)



Z500 error D10 DJF 1990-1994; contours every: 1 (dam)



Z500 error D10 DJF 1995-1999; contours every: 1 (dam)

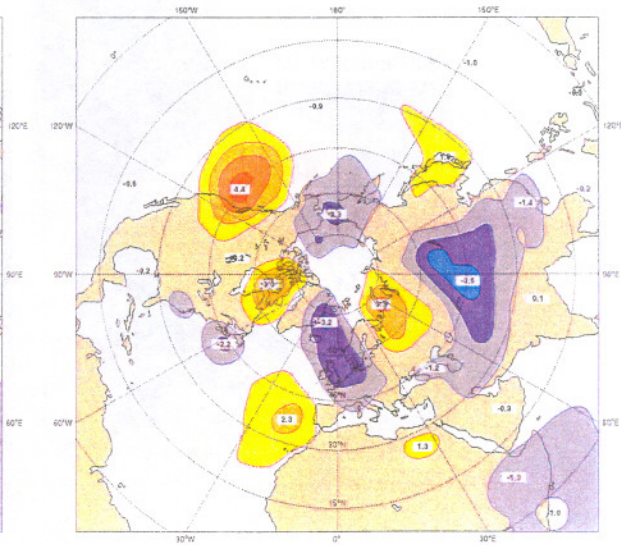


Figure 17: D+10 DJF 500hPa height systematic error (1 dam) for the ECMWF operational model. Top left:1980-84. Top right: 1985-1989. Bottom left: 1990-1994. Bottom right: 1995-1999.

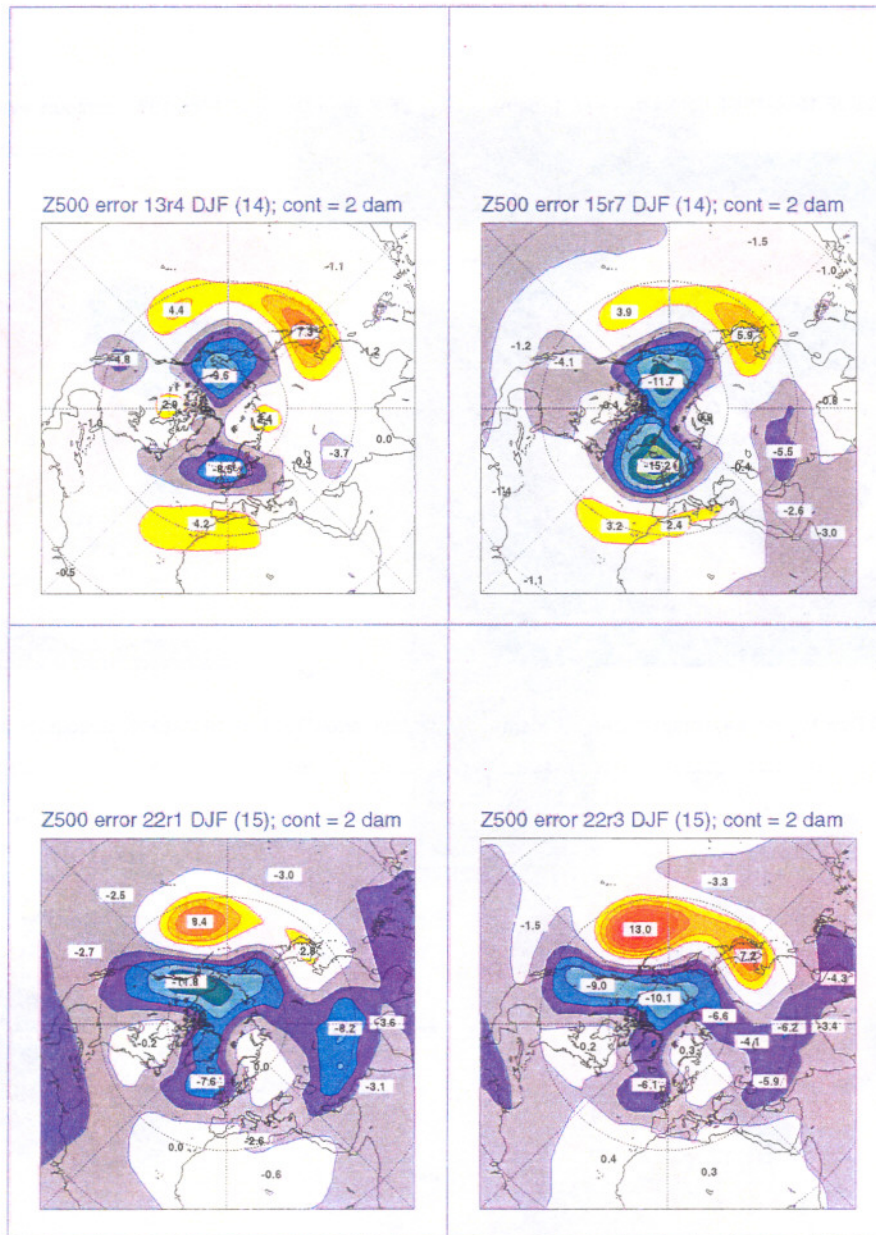


Figure 18: Days 1-90 mean DJF systematic error in 500hPa height (2 dam) based on 1 experiment per year over the ERA-15 period. Integrations made with specified observed sea surface temperature.

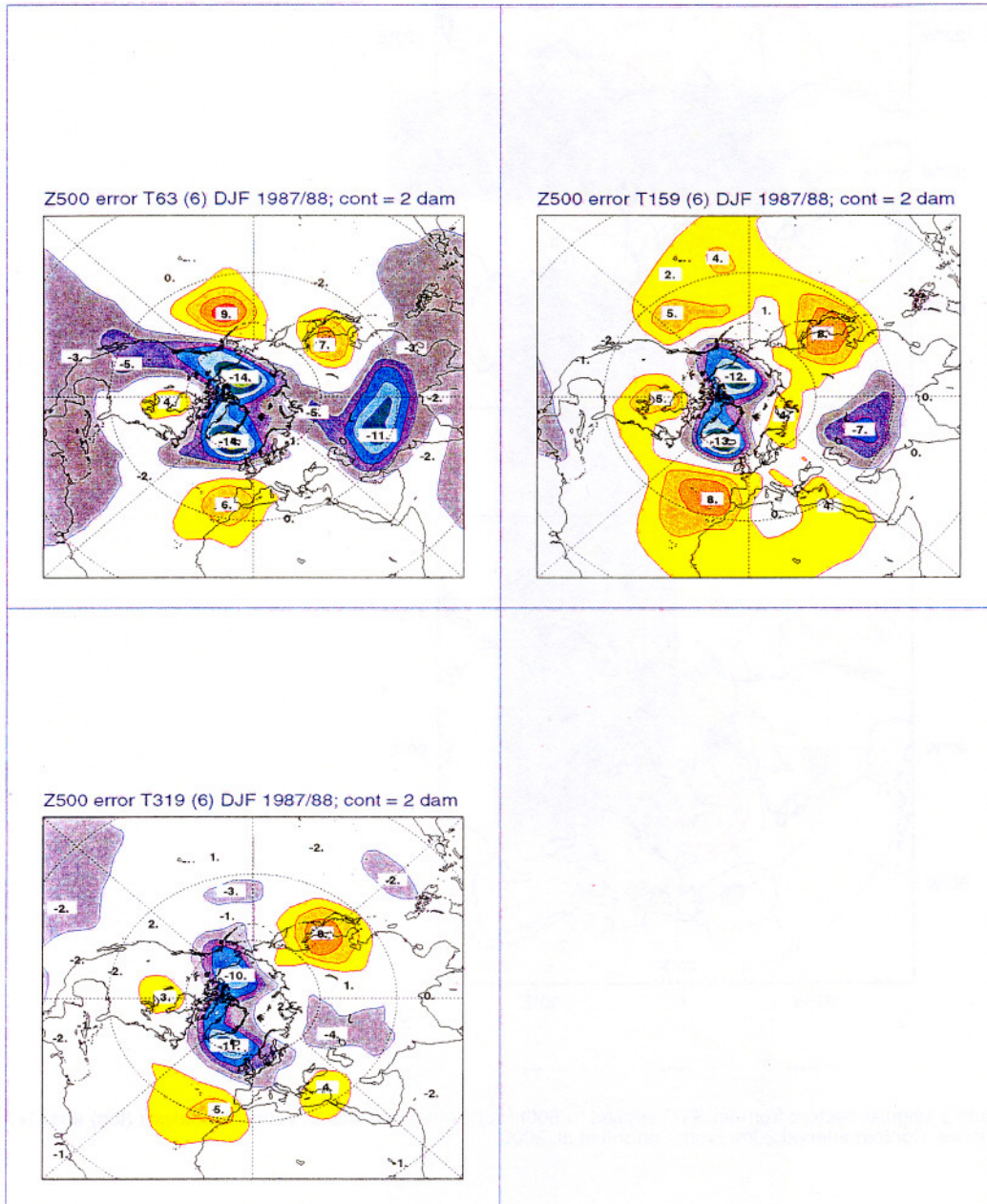


Figure 19: Impact of horizontal resolution on the seasonal timescale systematic error of cycle 18r6 (AMIP11). Top left: T63. Top right, TL159. Bottom, TL319. As Figure 18 but for an ensemble of 6 integrations made with SST for 1987/88.

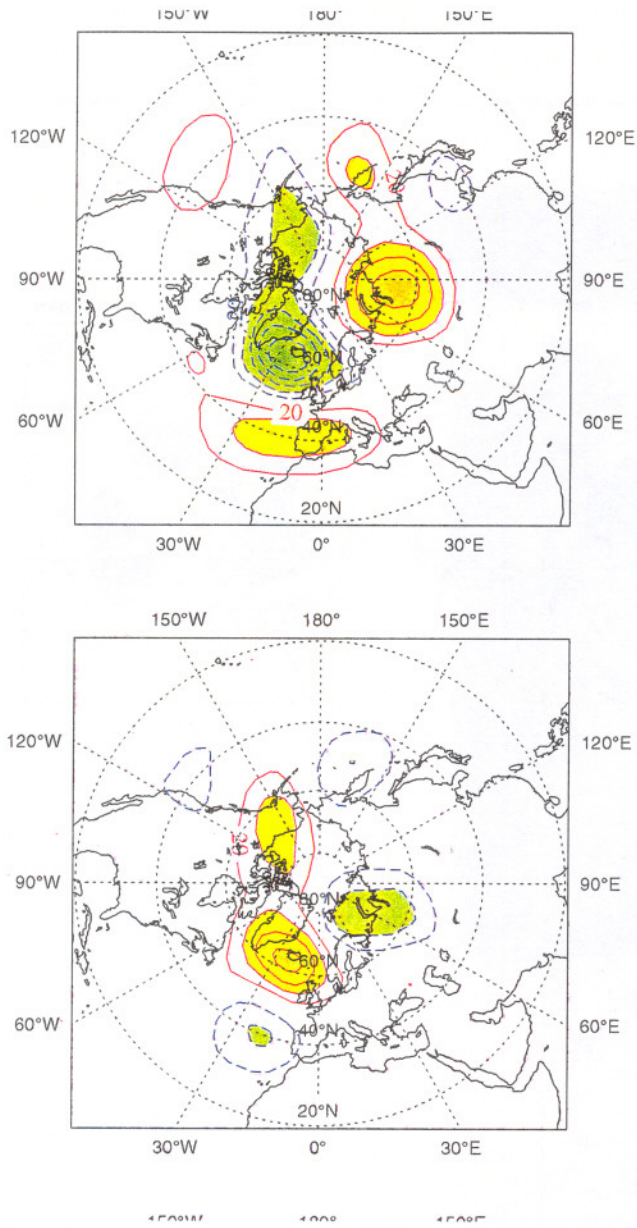


Figure 20: Leading singular vectors from an SVD applied to 500hPa height anomalies of verifying analysis (top) and D+7 forecast error (bottom) anomalies. Contour interval 20m. From Ferranti et al, 2000.

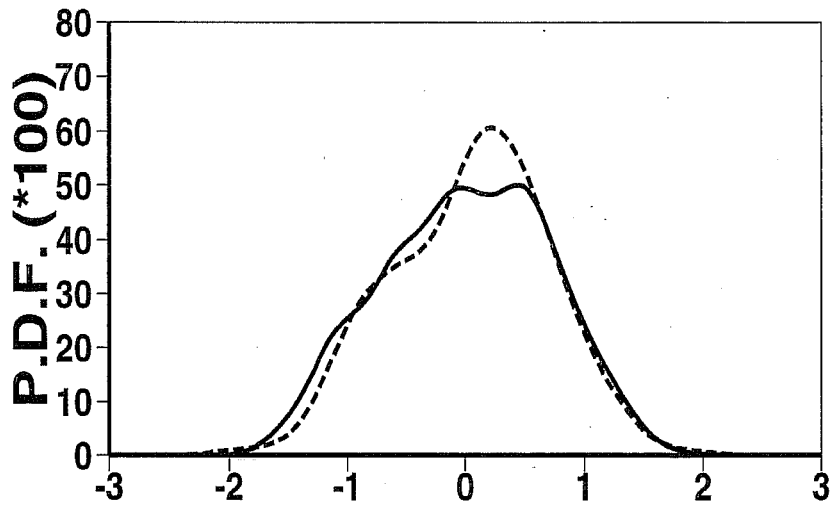


Figure 21: Probability density function of analysis (solid line) and D+7 forecast (dashed line) of daily projections onto the singular vector shown in Fig 20.

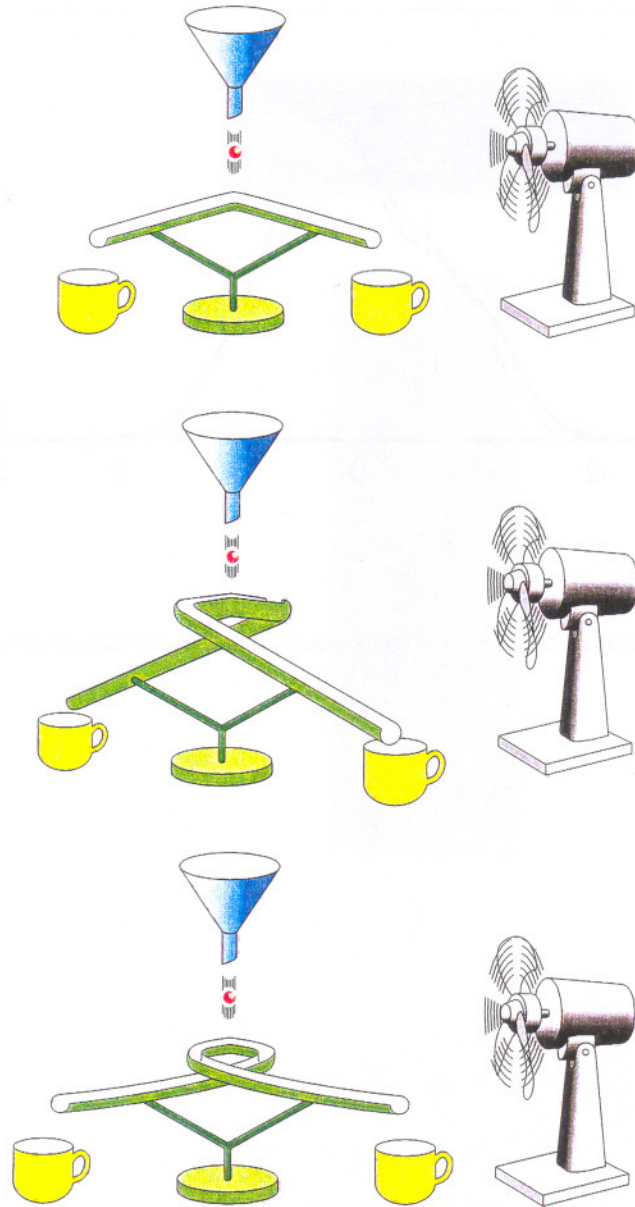


Figure 22: Schematic illustration of the link between regime predictability, instability and the response to a systematic forcing (representing a possible systematic misrepresentation of the equations of motion) in a nonlinear dynamical system (motivated by the Lorenz, 1963, system). The response of the system to the applied forcing is principally in the direction of the dominant EOF of the system, a result which is not sensitive to the direction of forcing (top). The sensitivity of the system to the applied forcing is dependent of the structure of the system in some neighbourhood of the regime instability (middle). It is possible (though under rather contrived conditions) for the response of the system to be opposite to the direction of applied forcing (bottom). From Palmer, 1999.

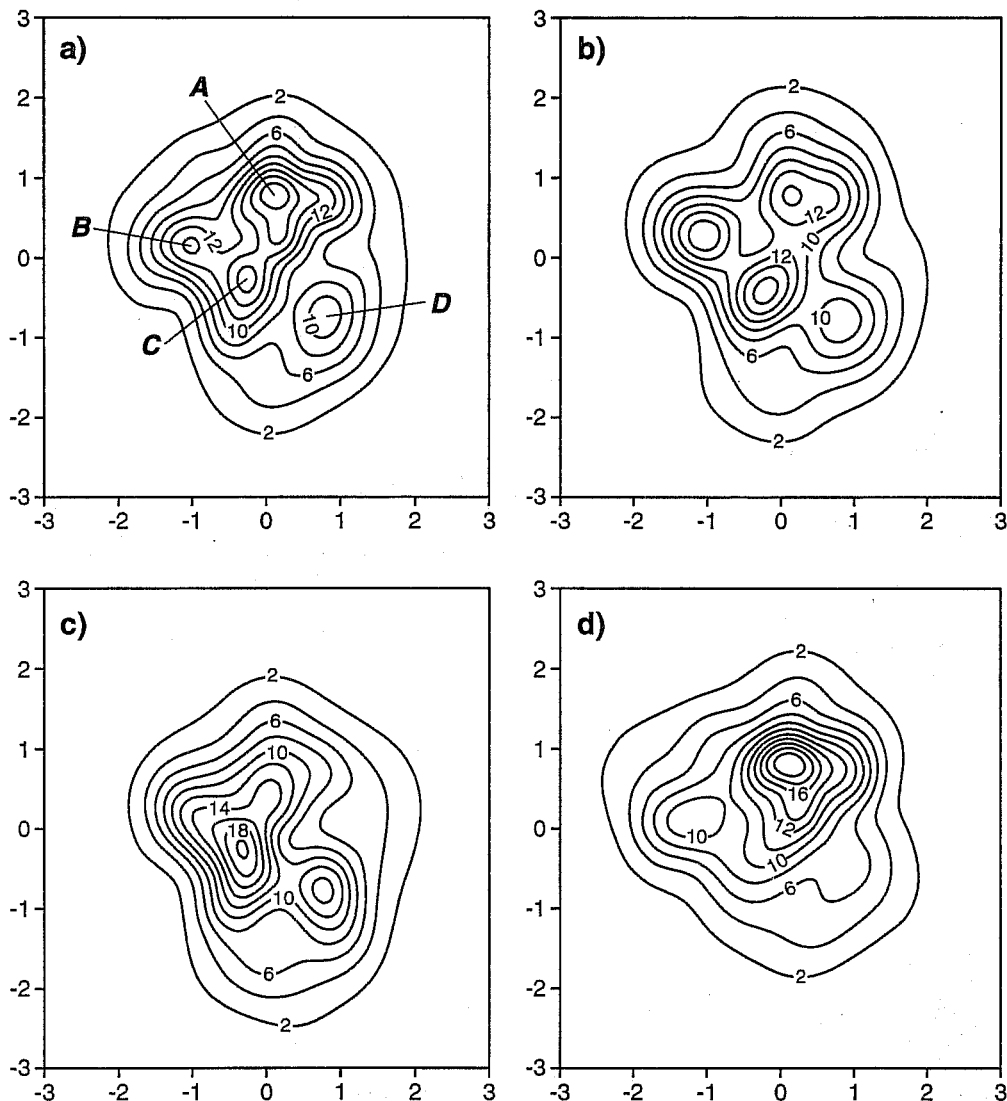


Figure 23: Atmospheric state vector pdf based on NCEP monthly-mean 500hPa height in the space spanned by the two dominant atmospheric EOFs computed from detrended monthly mean timeseries. a) using all data from the 1949-94 period. b) Using data from the 1949-94 period from which all months in which El Niño/La Niña events occurred have been removed. c) Using data from the 1949-71 sub-period. d) Using data from the 1971-94 sub period. The quantities plotted on the horizontal and vertical axes are projection coefficients of height onto the first and second EOFs respectively. The same detrended EOFs are used in all four panels. From Corti et al (1999).

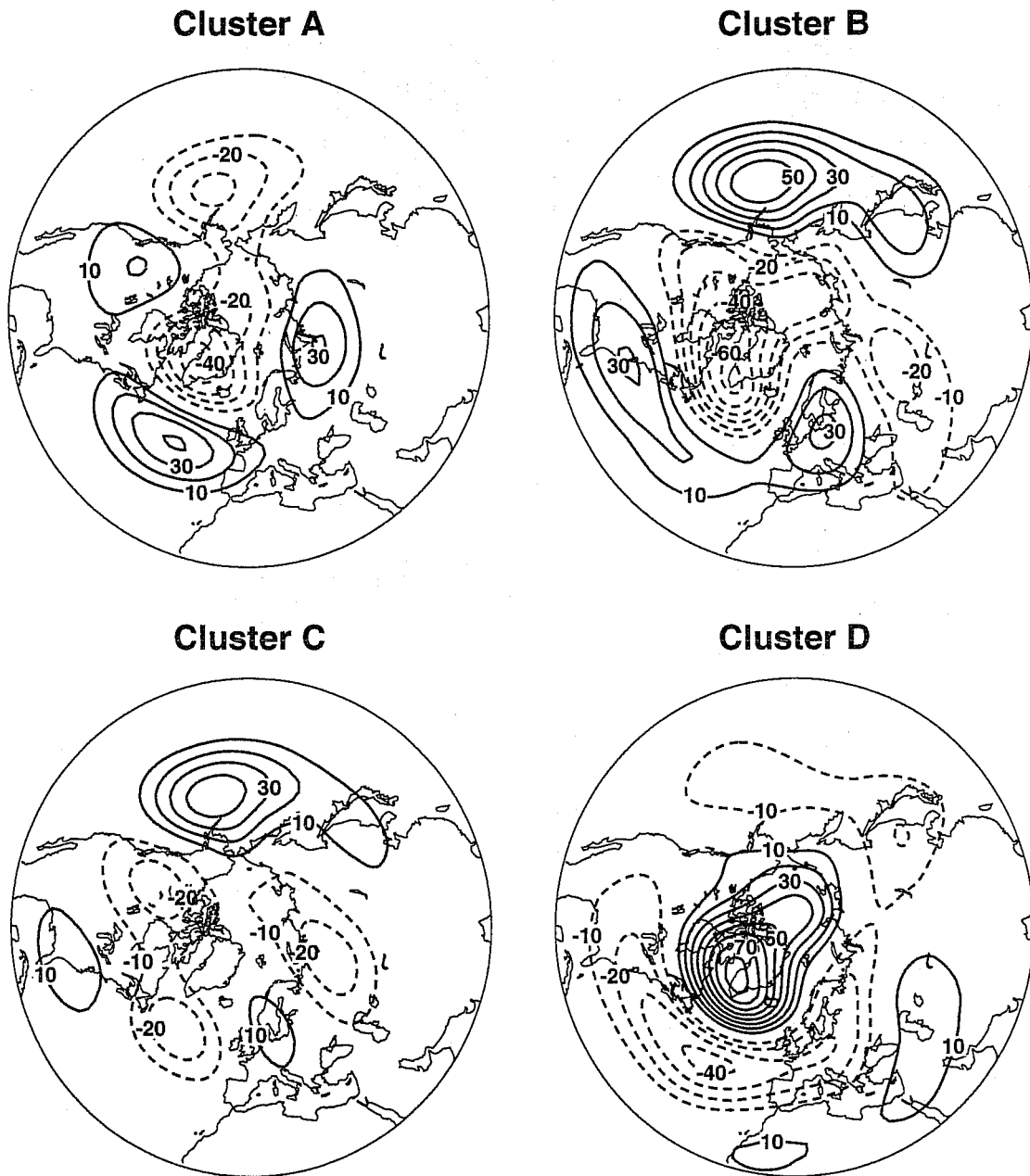


Figure 24: Geographical patterns (500hPa height anomaly) of the four pdf maxima in Figure 23a. From Corti et al (1999).

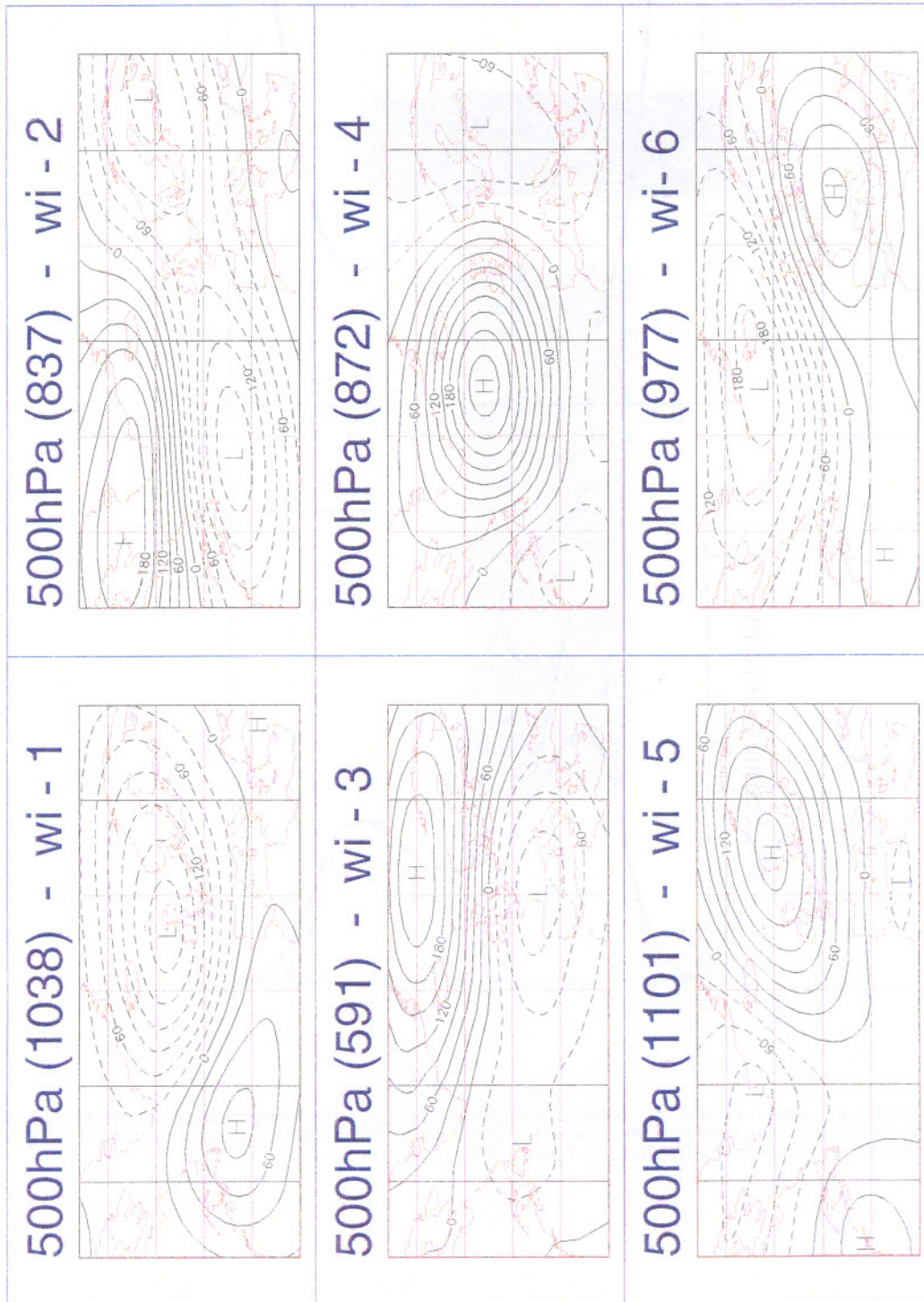


Figure 25: Six winter regimes defined from ERA-15 data by cluster analysis over the Euro-Atlantic sector. Fields shown are anomalies of 500hPa height. (From Chessa, 2000)

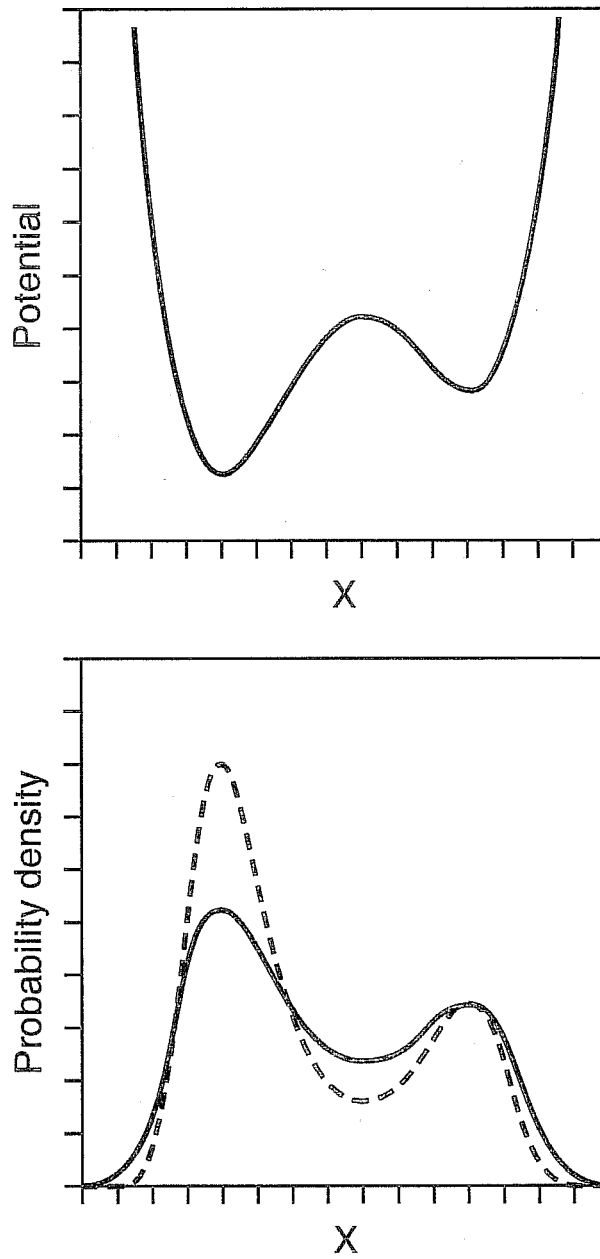


Figure 26: The effect of a reduction in transient stochastic forcing on the pdf (bottom) of a stochastically-forced system constrained to a twin-well potential surface (top), as estimated from a Fokker-Planck equation. The dashed line shows a pdf in which the stochastic forcing has been reduced by a factor of two. (From Molteni and Tibaldi, 1990)