TECHNICAL MEMORANDUM

139

Predictability studies in the medium and extended range

T.N. Palmer and S. Tibaldi

Research Department

September 1987

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

ECMWF

SCIENTIFIC ADVISORY COMMITTEE

15th SESSION

ECMWF/SAC(87)8

SHINFIELD

23 JULY 1987

Subject: Predictability Studies in the Medium and Extended Range

1. INTRODUCTION

Despite impressive improvements in numerical weather prediction over the last few years, the ECMWF forecast model shows considerable variability in predictive skill on many different timescales. An example of such variability (taken from Simmons, 1986) is illustrated in Fig 1, showing the day 3, 7 and 10 hemispheric skill scores for the operational model for November 1983. It can be seen that the later the model validity time, the larger the daily variability about the monthly mean score. As suggested by a straightforward extrapolation of these results, the mean skill of dynamical models in the extended range is typically very small, though case-to-case variability is sufficiently large that a few may be of genuine practical use (Mansfield, 1986, Molteni et al, 1987, Miyakoda et al, 1987, Tibaldi et al, 1987).

It is apparent, therefore, that a scheme to predict forecast skill will have substantial benefit in the medium range, and is an essential requirement for dynamical extended range forecasting. In order to quantify, in a simple way, the possible impact of such a scheme, we show in Fig. 2 the mean hemispheric RMS error of the ECMWF model averaged over the last 7 winters (solid line). If it was possible to discriminate between forecasts of above average and below average skill, and forecasts of below average skill were rejected, then the increase in mean skill of retained forecasts, shown as a dashed curve in Fig 2, is equivalent to enhanced predictability time of 2 days at day 10. Moreover, if a scheme could be devised which predicted when a forecast was at least one standard deviation better than average, and only these were retained, the mean skill (dotted curve) would show an increase in predictability time of 3 days at day 10. To put these numbers into context, Miller et al (1987) found that the orographic gravity wave drag parametrization which has a profound impact on model climate, improves predictability time by only about ½-1 day at day 10.

The top and bottom diagrams of Fig 11 show the same correlation coefficients except that the bottom diagram has been produced by first filtering all time series concerned (skill and spread) with the 5 day running-mean filter, to remove the high-frequency (day-to-day) variability of skill and spread, highlighting the low-frequency variations of both quantities (see above).

The large differences between the correlation coefficients shown in the two panels of Fig 11 underline the great difference in predictability of the high-frequency and the low-frequency variations of the forecast skill. It is also particularly interesting to note that the filter has a comparatively smaller effect on the correlation between the skill of today's forecast and the diagnostic spread between today's and tomorrow's forecast (.50 → .61). It is easier to interpret this if we think of the comparison between today's forecast and tomorrow's forecast as a sort of 'perfect mode' predictability experiment. If we think of the difference between today's and tomorrow's forecast as the growth, propagation and dispersion of the day 1 forecast error of today's forecast (as Lorenz did in his predictability work on the ECMWF forecasts; Lorenz, 1982) we can qualitatively understand why the spread between these two forecasts "knows" more about the skill of today's forecast than the spread between today's and yesterday's forecast (this spread, in turn, "knows" about yesterday's forecast skill, not today's). We should also remember that short range (day 1) forecast errors are our best proxy for analysis errors. We can then conclude that the comparatively smaller effect that the (5-day running mean) filter has on the correlation between skill and diagnostic spread suggests that the high-frequency variability of the forecast skill is more connected to analysis errors (that presumably have large day-to-day variability) than with any measure of the instability properties of the large-scale flow. Such instability properties might (and will indeed be shown later in the report to be) more correlated with the low-frequency variability of the skill, as skill of previous forecasts and prognostic forecast spread also are.

Let us now turn to spread-skill and skill-skill relationships on limited areas. Fig 12 has exactly the same layout as Fig 11, but refers to areas 1 (60N-30N, 0-30E), Western Europe; and 8 (60N-30N, 150W-120W), Eastern Pacific, both using the 5 day mean filtered data. As can easily be seen, the level of spread-skill correlation, although somewhat lower, is comparable to the hemispheric values. The skill-skill correlations, however, show a markedly different behaviour in the two areas. Let us consider, for example, the correlation between day 1 and day 6 forecast skill, with both forecasts starting from the same initial conditions (that is the same forecast run). This correlation (the only other example of 'diagnostic' correlation, as opposed to 'prognostic', shown in both Figs 11 and 12) measures essentially how good the low-frequency variability of the medium-range forecast skill is as an indicator of the low-frequency variability of the short-range forecast skill. The answer is quite different in different parts of the Northern Hemisphere: .38 correlation in Western Europe and -.03 (essentially nil) over the Eastern Pacific.

Given such a different result it is interesting to show the behaviour of such correlation as a function of the limited area (and therefore of longitude). Fig 13 shows this for filtered and unfiltered data. Two things appear clearly: the three limited areas for which such correlation is highest are limited areas 1, 2 and 9; and the correlation itself is made evident in these areas only after the 5-day mean filter has been applied. Fig 4 showed the location of such limited areas, together with the climatology of the high frequency eddy activity during the Northern Hemispheric winter. We therefore deduce that the short-range forecast skill is an indicator of medium-range forecast skill only in those areas of the Northern Hemisphere where the high-frequency (mainly baroclinic) eddy activity is lowest.

If we now turn to the spread-skill relationship and to its longitudinal dependence (Fig 14) we can recognize again that the prognostic spread-skill relationship shows a marked sensitivity to the 5-day filter (it exists only between the low-frequency variability parts of both spread and skill). The diagnostic spread-skill relationship, on the contrary, in addition to being higher (as noted before for hemispheric values) is little sensitive to the use of the 5-day filter (again as noted before), suggesting the notion that the diagnostic spread carries information about the effects of growth,

propagation and dispersion of short range forecasts (and therefore analysis) errors. Neither prognostic nor diagnostic spread-skill relationship seem to be particularly sensitive to longitude (i.e. to the particular limited area) save for a weak longitudinal trend, with (again) somewhat lower values over the Central Pacific and somewhat higher values over continental Europe.

It should also be mentioned that spread-skill and skill-skill correlations were also computed between different limited areas and at different time lags. The results were rather disappointing and showed that, almost invariably, the highest correlations are to be found between quantities computed on the same limited area and at the same time. Therefore when skill spread correlations are computed for the (independent) winter 1986/7 they will be correlated using the same regions.

To summarise the results of the skill-skill and skill-spread relationships, we conclude that

a) regarding the low-frequency variability of the skill (and spread), the skill-skill correlation is fairly good hemispherically, but on limited areas the signal can be seen clearly only in those regions not dominated by baroclinic activity (areas 1, 2 and 9), while the skill-spread correlation is higher (both hemispherically and on limited areas) and shows little longitudinal dependence.

b) regarding the high-frequency (day-to-day) variability of the skill, there are indications that it might be connected to analysis errors - this hypothesis should be confirmed by analyzing the relationship between medium-range skill and very short-range forecast error patterns. Work is already under way along this line.

4.2 Forecasting the forecast skill for 1986/7 and operational implementation

In this section we present some graphs of attempts to 'predict' the skill of forecasts of the winter 1986/7 using the variety of predictors discussed above. We shall concentrate entirely on predicting the day 3, 6 and 9 RMS error for regions 1 (60N-30N, 0-30E) and 8 (60N-30N, 150W-120W). In each graph the dashed line shows the actual forecast skill with 5-day running mean, the solid line shows the predicted skill. The correlation between actual skill and regressed skill for each predictor set is given in Table 3-5.

One final predictor whose skill we have indicated in Tables 3-5 is the prognostic equivalent of the persistence error diagnostic discussed above. Specifically we have calculated the RMS difference between the day n forecast field and the initial analysis in the appropriate region ('RMS tendency') and correlated this against skill. Note that for day 3 this is a poor indicator of skill in region 8 - the model is good at forecasting changes in flow at this range. However, at day 9 it is a much better indicator - with a correlation of .42 in region 8. There are some encouraging preliminary results that the RMS tendency of ten-day mean forecast fields may be a good predictor in the extended range (see, for example, Molteni et al, 1987).

Fig 15 shows graphs of the predicted and actual scores for day 3 RMS over regions 1 and 8, for the three predictor sets. As indicated in Table 3 the highest level of skill is achieved by the spread indicator in region 1 and the EOF indicator in region 8. The day 1 forecast error is a good indicator of skill in region 1, less so in region 8, consistent with the discussion in Section 4.1. Note in Fig 15 that there is no obviously strong correlation between the predictor sets, indicating some degree of independence.

Fig. 16 shows graphs of predicted and actual scores for day 9. As shown in Table 5, the spread indicator is the most skilful in region 1 whereas RMS tendency is most skilful, and EOF and spread indicators equally skilful, in region 8.

The relatively poor performance of the EOF predictor in region 1 may indicate that the training data is insufficiently homogeneous to give reliable results. As discussed in Section 3, the factor structure constants for skill in region 1 suggest that forecasts which develop anomaly patterns consistent with the model systematic error may be relatively unskilful. However model systematic error changes as a result of improvements to model formulation. In particular, only during the last year (1985/6) of the training period was the model horizontal resolution the same as that for the independent test year (1986/7). It was therefore decided to run, in addition to the six year regression, a regression using just the data for 1985/6. Of course the penalty for using a more homogeneous dataset is its relative shortness.

Table 4 shows correlations between predicted and actual skill at day 6 for both regressions for N=23. It is clear that the predictions using 1985/6 data are more skilful than those using all six years for both regions. Whether this is a reliable result will depend on analysis of the forthcoming winter. It is also interesting to note that when a reduced (N=9) predictor set is used, the skill of the regression using 1985/6 data is reduced, both for regions 1 and 8. This suggests that problems of overprediction are not occurring with N=23.

Predictions for day 6 are shown in Fig 17. As above, for region 1 the correlation between predicted and actual skill using spread is superior to the EOF prediction using 1985/6 data, though for region 8 it is inferior. For region 1 it is interesting to note that both spread and flow pattern predictors capture the poor forecast skill around day 40. However, they subsequently disagree after day 40. The skill of a day 1 forecast initialised 1 day before the day 6 forecast, whilst correlating reasonably well with observed skill at day 6, has failed to predict the poor forecasts near day 40.

It would appear that all techniques show some promise as potential predictors, and the most obvious conclusion is that some combination of them may prove optimal. For operational implementation we shall attempt to do this through a probabilistic categorical approach. First of all the distribution of each chosen predictor (eg spread, day 1 forecast, and rotated EOF coefficient) will be divided into five equally likely categories (much above average, above average, average, below average, and much below average), based on data for 1985/6 and 1986/7, when the T106 model was in operation.

Secondly, contingency tables will be constructed for each predictor showing the probability that a prediction in category i corresponded to a forecast error in category j. The weight that one predictor is given relative to another can be determined essentially by the trace of the contingency matrix.

In order to consider how this procedure would give a probabilistic prediction of forecast skill, consider the following hypothetical example where we use two predictors A and B, whose contingency tables in terms of a tercile categorisation are

A

predictor category \ forecast error category			
	.30	.01	.01
	.02	.30	.02
	.01	.02	.30

B

predictor categogry \ forecast error category			
	.20	.08	.03
	.10	.20	.10
	.03	.05	.20

A is a better discriminator than B, and we will weight predictions from A by the factor

$$\frac{(.3 + .3 + .3) - .33}{1 - .33} = .85$$

and predictions from B by the factor

$$\frac{(.2 + .2 + .2) - .33}{1 - .33} = .40$$

(If we had decided to include a third predictor C whose contingency table had elements each equal to .11, i.e. no discrimination, its weight would be equal to zero). Now suppose, for a given operational forecast, predictor A gives category 3 (above average) and predictor B gives category 2 (average). The probability of the three categories of forecast error are

$$\frac{.85}{.85 + .40} \times \begin{bmatrix} .03 \\ .06 \\ .90 \end{bmatrix} + \frac{.40}{(.85 + .40)} \times \begin{bmatrix} .24 \\ .60 \\ .15 \end{bmatrix} = \begin{bmatrix} .10 \\ .23 \\ .66 \end{bmatrix}$$

ie a probability of 10% of a below average forecast, 23% of an average forecast, and 66% of an above average forecast.

This approach is at present under development, and, it is hoped, will be implemented for the winter season 1987/8.

5. A POSSIBLE PHYSICAL MECHANISM FOR MEDIUM AND EXTENDED-RANGE FORECAST SKILL VARIABILITY

In this section we shall discuss a possible mechanism that might explain the dependence of forecast skill on the modes of variability illustrated in Figs 6 and 7. Broadly speaking candidate mechanisms could be classified as those related to systematic deficiencies in model formulation, and those related to intrinsic loss of predictability of the atmospheric flow. Indeed the existence of model systematic errors can give rise to a 'trivial' relationship between forecast skill and variability in forecast flow. Consider, for example, a model with no predictive skill and a climate drift $\Delta C(t)$ where t is the forecast verification time. Clearly, over a large enough sample the RMS error of those forecasts which are close (in an RMS sense) to the observed climate will be smaller than those which are further away. Hence forecasts with anomaly - $\Delta C(t)$ (relative to the model climate) will over a large sample, have smaller RMS error than those with anomaly + $\Delta C(t)$.

Fig 18 shows the ECMWF model mean (or systematic) day 9 wintertime error, from 1980/1 to 1984/5. If the PNA-like mode of variability shown in Fig 8 correlated strongly with this mean error pattern, there would be strong prima facie evidence that variability in the impact of model systematic errors was the dominant mechanism explaining the results correlating PNA mode and forecast skill. However, it can be seen that the principal anomaly centres of the PNA mode are not in phase with the centres of maximum systematic error. On the other hand, as discussed above, there is some correspondence between Fig 18 and the mode shown in Fig 7, correlating with European scores.

In order to study the stability characteristics of flow patterns associated with either skilful or unskilful forecasts, it is worth recalling that in the medium and extended range, forecast errors are dominated by large-scale quasi-stationary patterns (e.g Wallace et al, 1983). Simmons et al (1983) have shown that the climatological zonally-varying flow is barotropically unstable to just such large-scale quasi-stationary modes. Therefore, in order to test the importance of variability in the intrinsic predictability of forecast flows, the barotropic stability of basic state flows which differ from climatology by either the subtraction or addition of the anomaly field shown in Fig 7 was studied. These two basic states are shown in Fig 19.

(More precisely these represent 300mb streamfunction fields formed by adding to a climatological 300mb streamfunction field, either minus (Fig 19a) or plus (Fig 19b) the geostrophic streamfunction anomaly formed from Fig 2, scaled by the density factor 5/3). One can see that with the mode subtracted, (basic state B-), the east Asian jet extends across the Pacific and is weakly diffluent over the east Pacific. The jet is then steered north by the enhanced Rockies ridge. With the mode added (basic state B+), the jet is more strongly diffluent over the central Pacific.

In order to present results in a way that corresponds most closely with the hypothesis that the variability in skill is related to the rate at which initial errors grow, we show in Fig 20 results from two initial value problems, where in both integrations an identical disturbance was introduced at 30N, 120E. (Similar types of initial value problems were reported by Simmons et al, 1983. As discussed by these authors, the quasi-stationary growth of an initial disturbance can be interpreted in terms of its projection onto the unstable normal modes of the basic state flow. Readers unfamiliar with this type of calculation, and its interpretation are therefore referred to Simmons et al for details.)

Fig 20 shows the model streamfunction response at days 2, 4 ,9 and 30 for the two basic states. At day 2, the initial disturbance is seen propagating downstream and the effects of the difference in basic states is small. At day 4 the disturbance has propagated further downstream, and, for example the effect of the enhanced Rockies ridge in the basic state B- can be seen steering the disturbance to the north. However, the major difference is the amplitude of the disturbance centre over the central Pacific. By day 9, this difference is more clearly marked; the downstream propagating wavetrain has largely dissipated and a quasi-stationary unstable mode is growing in the central Pacific. There is almost a factor of three difference in the amplitude of the central Pacific response at day 9 between B- and B+. This difference continues to grow throughout the thirty day integration; by day 30 there is essentially an order of magnitude difference in the response between B- and B+.

A detailed analysis of the normal mode structure of B- and B+ will appear elsewhere; however, relative to B-, the amplitude of the most unstable normal mode is well localised in the Pacific jet diffluence region. Small scale perturbations, as they propagate through this region, project significantly onto this normal mode. With basic state B+ the amplitude of the most unstable normal mode is more uniformly distributed around the hemisphere, and localised perturbations project relatively weakly onto this mode.

In addition to these results, it should be remarked that Simmons (1987) has studied the normal mode structure of the monthly mean flow for January 1981, and finds significantly smaller e-folding times for the most unstable mode compared with that for a climatological flow.

Of course, it is not possible to claim any quantitative significance to these results. For one thing, large-scale baroclinic instability would alter the growth rates of the modes (Frederiksen, 1986). However, the results are sufficiently striking that it appears quite possible that the stability characteristics of the two 'composite' basic states could explain the forecast skill variability results above. Obervational evidence for the instability of the PNA mode is presented in Palmer (1987).

It is possible that this technique could be used as a dynamical method for forecasting the skill of a medium or extended range forecast. For example the ten day mean field of a given forecast could be used to define a basic state, and the day 1 forecast error could be used as the 'initial' perturbation. Such a technique would combine dynamically the correlations between forecast flow patterns and short range forecast error with skill. Work to explore this possibility is underway.

6. CONCLUDING REMARKS

Three types of question motivate predictability research at ECMWF; firstly, to what extend does interannual variability in the atmospheric general circulation influence the skill of the forecast, secondly can the observed variability in forecast skill be predicted, and thirdly, what are the dynamical mechanisms that give rise to variability in forecast skill?

With regard to the first question we noted that in regions of large low-frequency atmospheric variability (only) there were substantial year-to-year correlations between day 9 forecast error and day 9 persistence error. Furthermore, differences in model error between years with similar persistence error but different model formulation were substantially smaller than differences in model error between years with similar model formulation but different persistence error. We can therefore conclude that interannual variability in the atmospheric general circulation is dominant in accounting for year-to-year variability in regional forecast skill.

To determine possible predictors for a scheme to predict forecast skill, we have investigated descriptions of forecast and initial flow patterns correlated with forecast skill, consistency between consecutive forecasts, and the skill of a short range forecast verifying on the initialisation date of the forecast whose skill we wish to predict. Each of these predictors appeared to show useful skill when applied to the winter 1986/7, and a probabilistic approach to combining these predictors was described. It is hoped that this approach will be used to provide real-time estimates of forecast skill for the coming winter.

It was noted that the flow pattern predicted by the regression analysis to be most strongly correlated with day 9 forecast skill over the Pacific region was similar to the Pacific/North American (PNA) mode of atmospheric low-frequency variability. It was noted that variability in extended range forecast skill also appeared to be associated with the signed amplitude of the PNA mode. Using a barotropic model with two basic states with equal and opposite PNA mode anomalies, it was shown that initial perturbations can grow much more strongly on basic states with one sign of the PNA mode than the other. This suggests that a possible mechanism to explain some of the results described

above is in terms of (relatively slow) barotropic instability of the large scale flow, and its dependence on the signed amplitude of the PNA mode. It is possible that such barotropic model integrations could provide a dynamical tool for predicting, operationally, forecast skill.

Research into these three areas is continuing.

Acknowledgement

Frequent and constructive discussions with F Molteni at various stages during this work are gratefully acknowledged.

Table 1a

Region	no time filter	with 5-day mean filter
1	.53	.72
2	.45	.64
3	.41	.61
4	.41	.65
5	.43	.71
6	.40	.66
7	.43	.64
8	.51	.69
9	.44	.69
10	.38	.55
11	.38	.59
12	.52	.71

Table 1b

Region 1

	day 3	day 6	day 9
46 EOFs	.78	.75	.72
18 EOFs	.68	.66	.64

Region 8

46 EOFs	.66	.65	.69
18 EOFs	.48	.50	.54

Table 1 Correlation between regressed scores and actual scores using forecast and initial data from 1980/86. Day 6 RMS a) for 12 equal area regions (see Fig 4. b) For region 1 and 8; days 3, 6 and 9 using N=23 and N=9.

Initial date	Anomaly correlation	PNA index
01 January 1977	.81	150
01 January 1978	.68	120
01 January 1979	.42	20
16 January 1979	.57	-30
01 January 1980	.52	-10
01 January 1981	.71	210
01 January 1982	.49	-60
01 January 1983	.83	190

Table 2 Initial dates, 30 day mean (bias adjusted) anomaly correlations, and amplitude of the observed monthly mean PNA index for 8 extended range forecasts, as reported by Miyakoda et al (1987).

Predictor Set	Correlation	
	Region 1	Region 8
EOF regression using 1980/6 training data	.16	.42
day 3/day 4 prognostic spread	.57	.31
day 1 prognostic forecast error	.42	.16
RMS tendency	.14	-.28

Table 3 Correlation between actual day 3 RMS errors over regions 1 and 8 (see Fig 4), for the winter 1986/7 using a variety of different predictor sets (5-day mean filter applied).

References

Arpe, K, A Hollingsworth, M S Tracton, A C Lorenc, S Uppala and P Kållberg, 1985: The response of numerical weather prediction systems to FGGE level IIb data. Part II: Forecast verifications and implications for predictability. Quart J R Met Soc, 111, 67-101.

Branstator, G, 1987: The variability in skill of 72-hour global-scale NMC forecasts. Mon Wea Rev.

Frederiksen, J S, 1986: Instability theory and nonlinear evolution of blocks and mature anomalies. Adv in Geophys, 29, 277-304.

Kalnay, E and A Dalcher, 1987: Forecasting forecast skill. Mon Wea Rev, 115, 349-356.

Lau, N G, G H White and R L Jenne, 1981: Circulation statistics for the Extratropical Northern Hemisphere Based on NMC analyses. NCAR Tech.Note NCAR/TN-171+STR.

Lorenz, E, 1982: Atmospheric predictability experiments with a large numerical model. Tellus, 34, 505-513.

Mansfield, D A, 1986: The skill of dynamical long-range forecasts, including the effect of sea surface temperature anomalies. Quart J Roy Met Soc 112, 1145-1176.

Miller, M J, T N Palmer and R Swinbank, 1987: Orographic gravity wave drag: its parametrization and influence in general circulation and numerical weather prediction mode.

Miyakoda, K, J Sirutis and J Ploshay 1987: One month forecast experiments - without anomaly boundary forcings. Mon Wea Rev, 114, 2363-2401.

Molteni, F, 1987: Empirical Orthogonal Function Analysis of the Zonal and eddy components of 500 mb height fields in the northern extratropics. ECMWF Tech.Rep. No.61.

Molteni, F, U Cubasch and S Tibaldi, 1987: 30- and 60-day forecast experiments with the ECMWF spectral models. Proceedings of the ECMWF Workshop on Predictability in the Medium and Extended Range. 17-19 March 1986. ECMWF, Reading, UK.

Morrison, D F, 1983: Applied linear statistical methods. Prentice-Hall, Englewood Cliffs, N J, 562pp.

Palmer, T N, 1987: Medium and extended range predictability, and stability of the PNA mode. Quart J Roy Met Soc. submitted.

Palmer, T N and S Tibaldi, 1986: Forecast skill and predictability. ECMWF Tech.Memo No.127.

Simmons, A J, 1987: Barotropic instability, and anomalies of the extratropical northern winter circulation. Pontificae Academiae Scientiarum Scripta Varia 'Drought, El Nino and Teleconnections'. Published by Pontificia Academia Scientarium. Vatican City.

Simmons, A J, J M Wallace and G Branstator, 1983: Barotropic wave propagation and instability and atmospheric teleconnection patterns. J Atmos Sci, 40, 1363-1392.

Tibaldi, S, C Brankovic and U Cubasch, 1987: 30-day integrations using the operational ECMWF spectral model. ECMWF Tech.Memo.No.138.

Wallace, J M and D S Gutzler, 1981: Teleconnections in the geopotential height field during the northern hemisphere winter. Mon Wea Rev, 109, 784-812.

Wallace, J M, S Tibaldi and A J Simmons, 1983: Reduction of systematic forecast errors in the ECMWF model through the introduction of an envelope orography. Quart J R Met Soc, 109, 683-718.

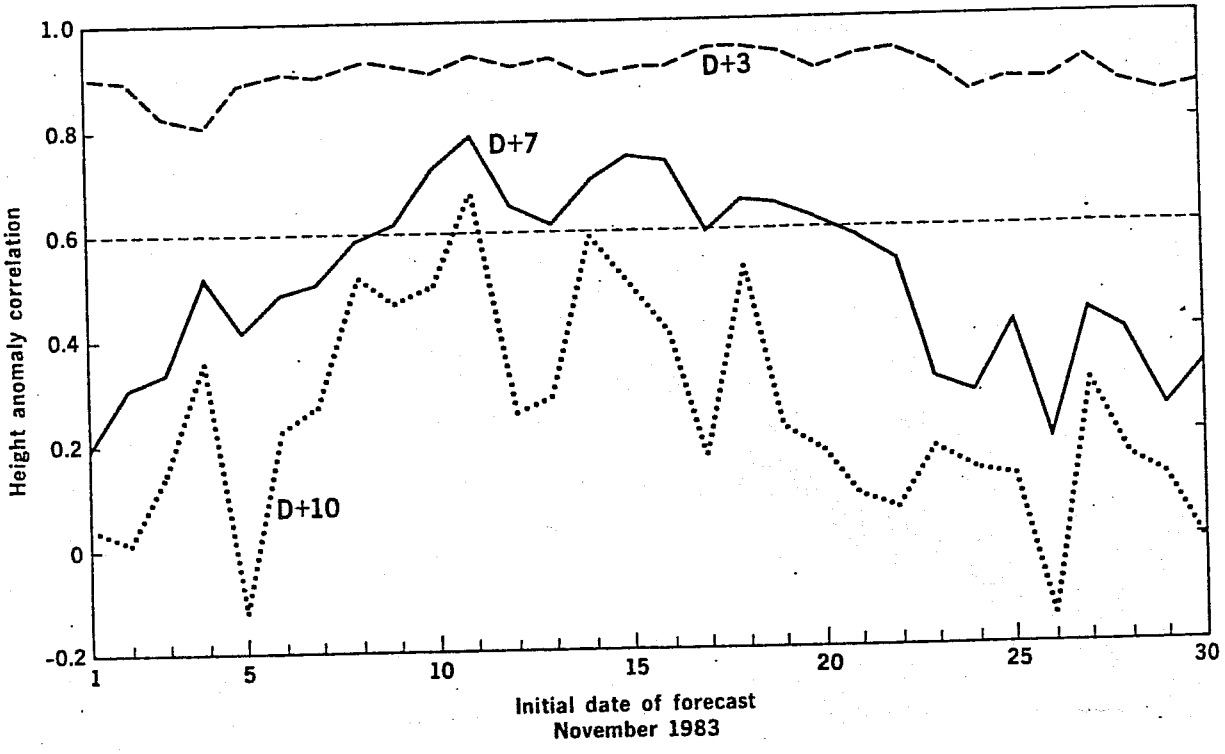


Fig 1 Anomaly correlations of height for 1000-200 mb and the extratropical Northern Hemisphere for 3-, 7-, and 10-day forecasts performed from initial dates within the month of November 1983.

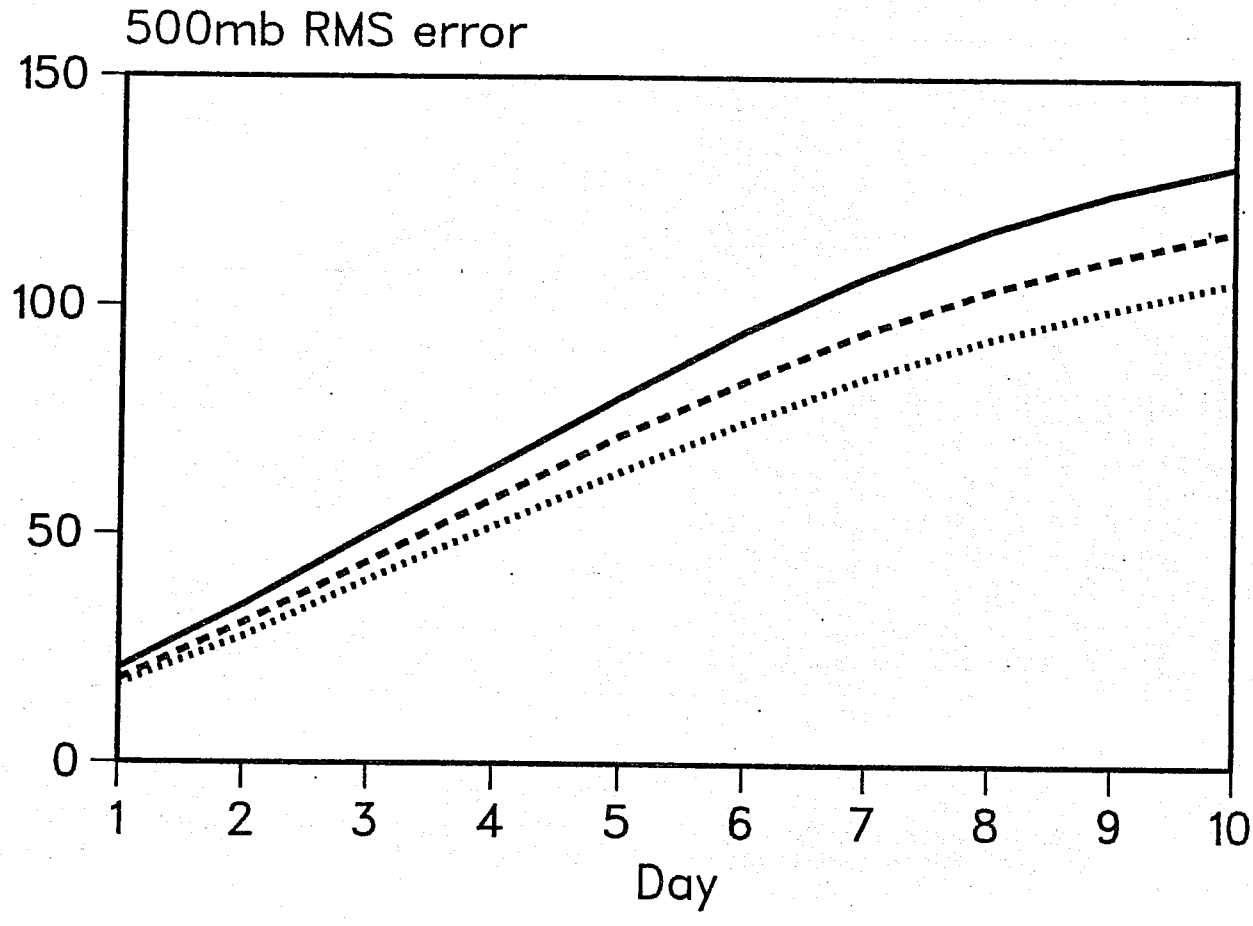


Fig 2 500 mb RMS error of model forecasts. Full line - averaged over 7 winter periods from 1980/1. Dashed line - RMS error of those forecast with above average skill. Dotted line - RMS error of those forecasts with skill one standard deviation above average.

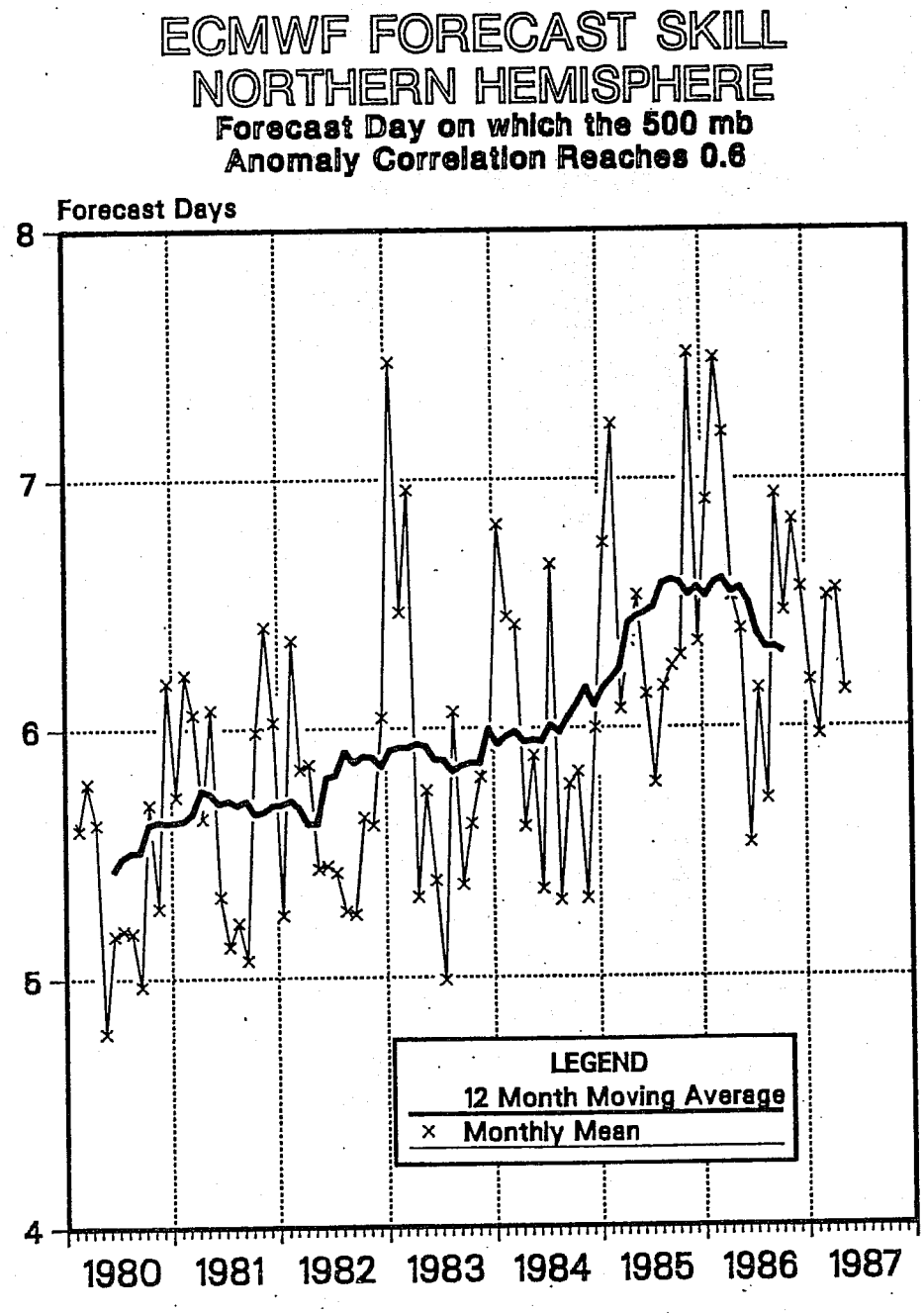


Fig 3 Monthly and 12-month moving average values of forecast day on which the hemispheric anomaly correlation of geopotential height (500 mb) falls to 0.6.

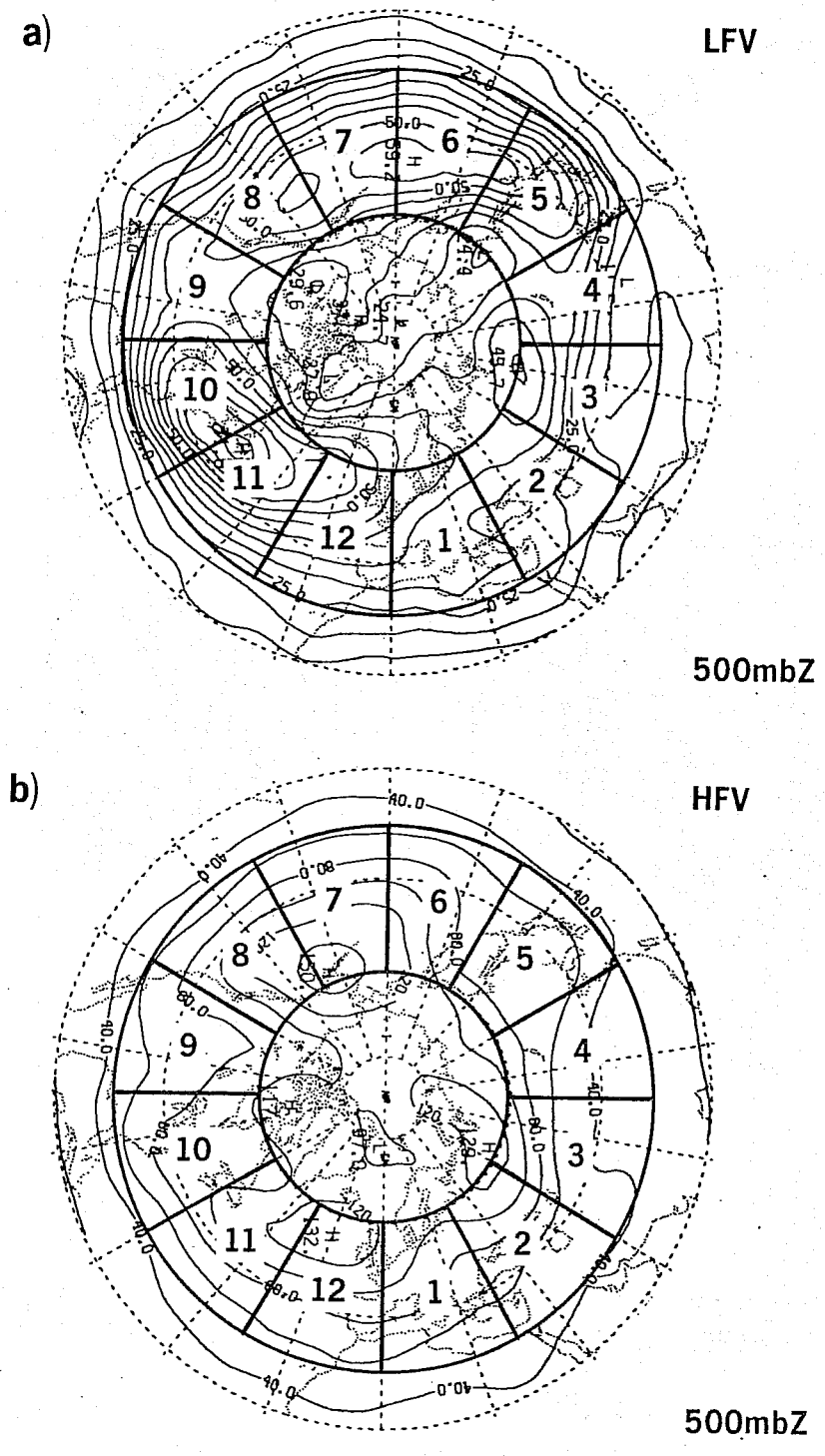


Fig 4 Northern Hemisphere wintertime standard deviation of 500 mb geopotential height ($\overline{Z^2}^{\frac{1}{2}}$). Contour interval 5m. Blackmon band-pass filter. Periods retained between 2.5 and 6 days (a) and larger than 10 days (b) - limited areas where skill scores have been computed are superimposed (1 to 12, 30° longitude width). (From Lau et al, 1981).

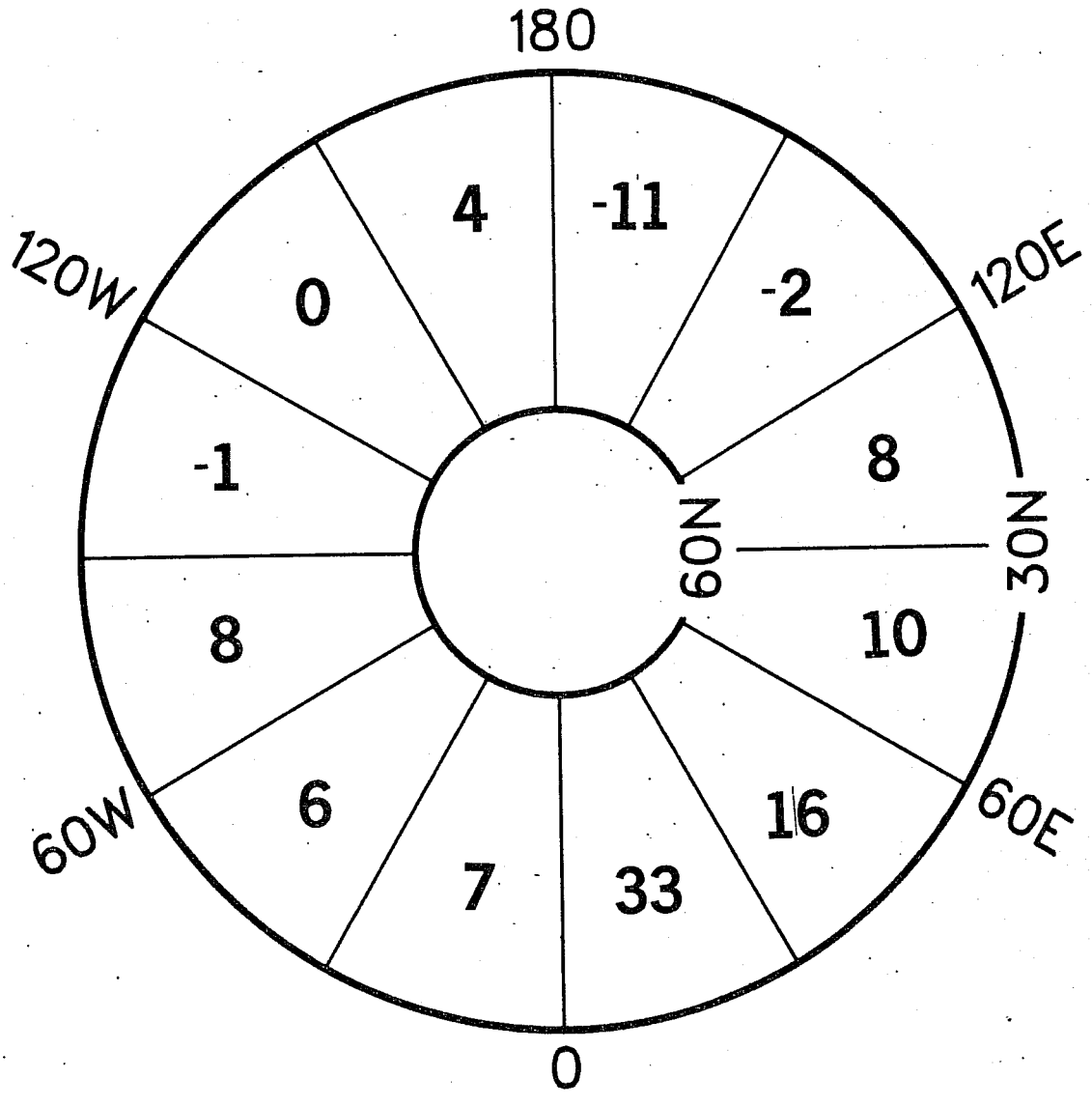


Fig 5 Difference in 500 mb RMS error in 12 regions between the winter 1986/7 and the winter 1985/6.

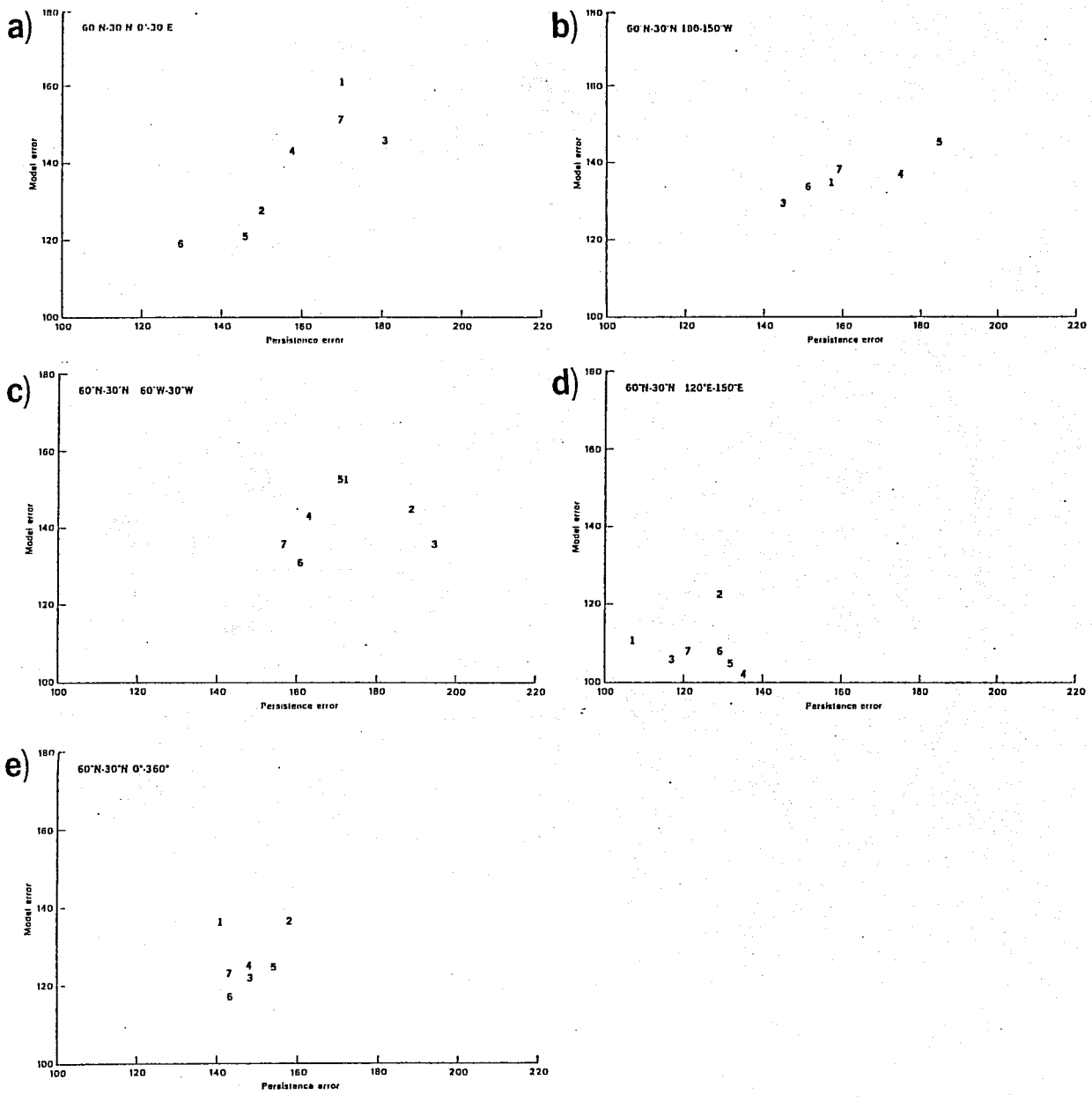


Fig 6 Scatter plot of winter mean day 9 500 mb height RMS error against the day 9 500 mb height RMS error of a persistence forecast. 1 = 1980/1, 2 = 1981/2...7 = 1986/7.

a) 60-30N, 0-30E
b) 60-30N, 180-150W
c) 60-30N, 60W-30W
d) 60-30N, 120E-150E
e) 60-30N, 0-360E

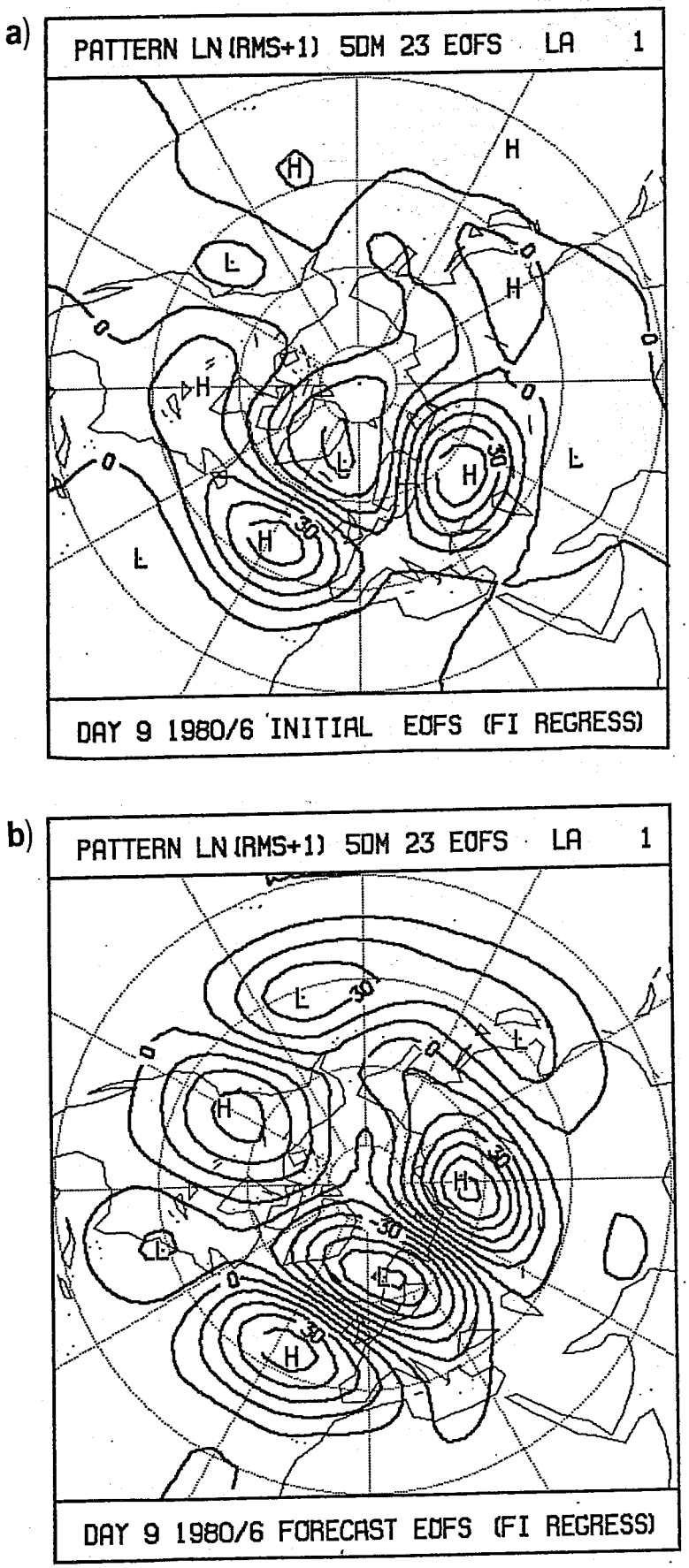
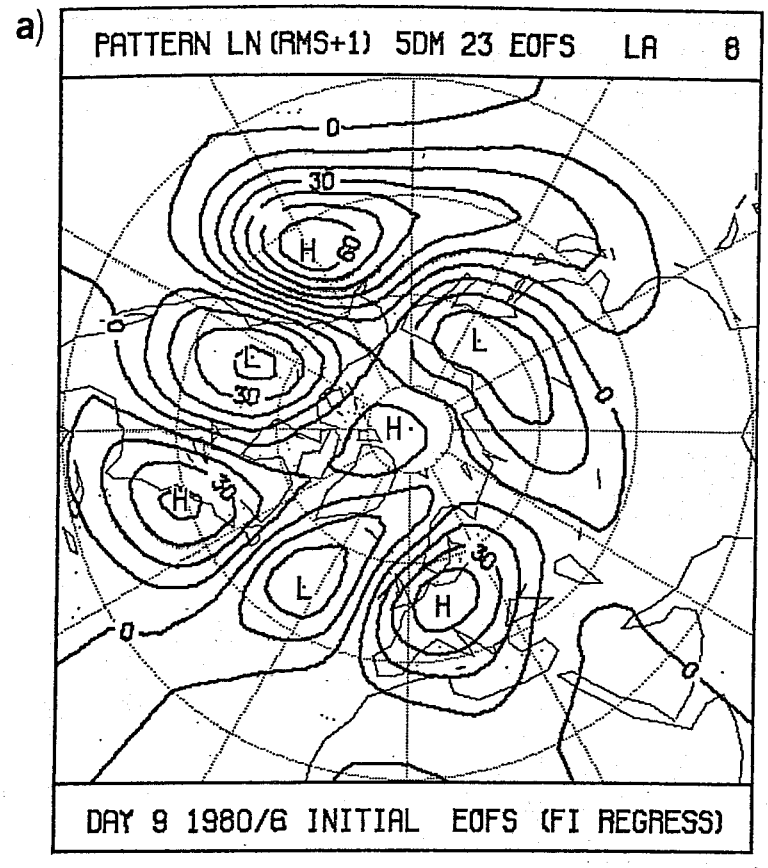


Fig 7 Factor structure constants giving pattern of 500 mb height anomaly of a) initial data b) day 9 forecast that correlates most strongly with the day 9 RMS error over the region 60-30N, 0-30E.



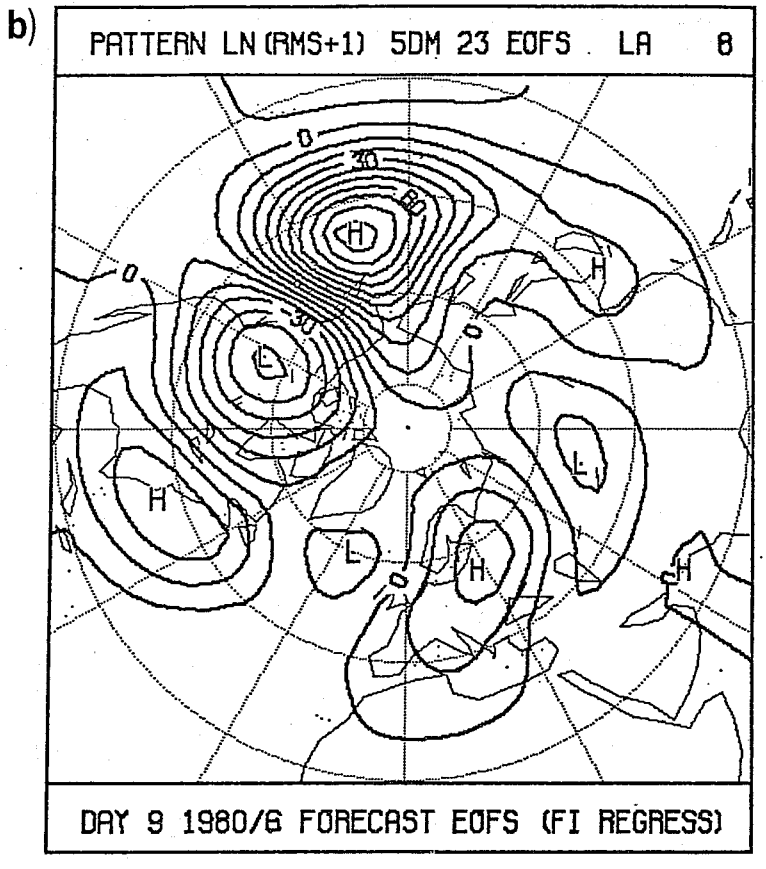


Fig 8 As Fig 7 but for region 60-30N, 150-120W.

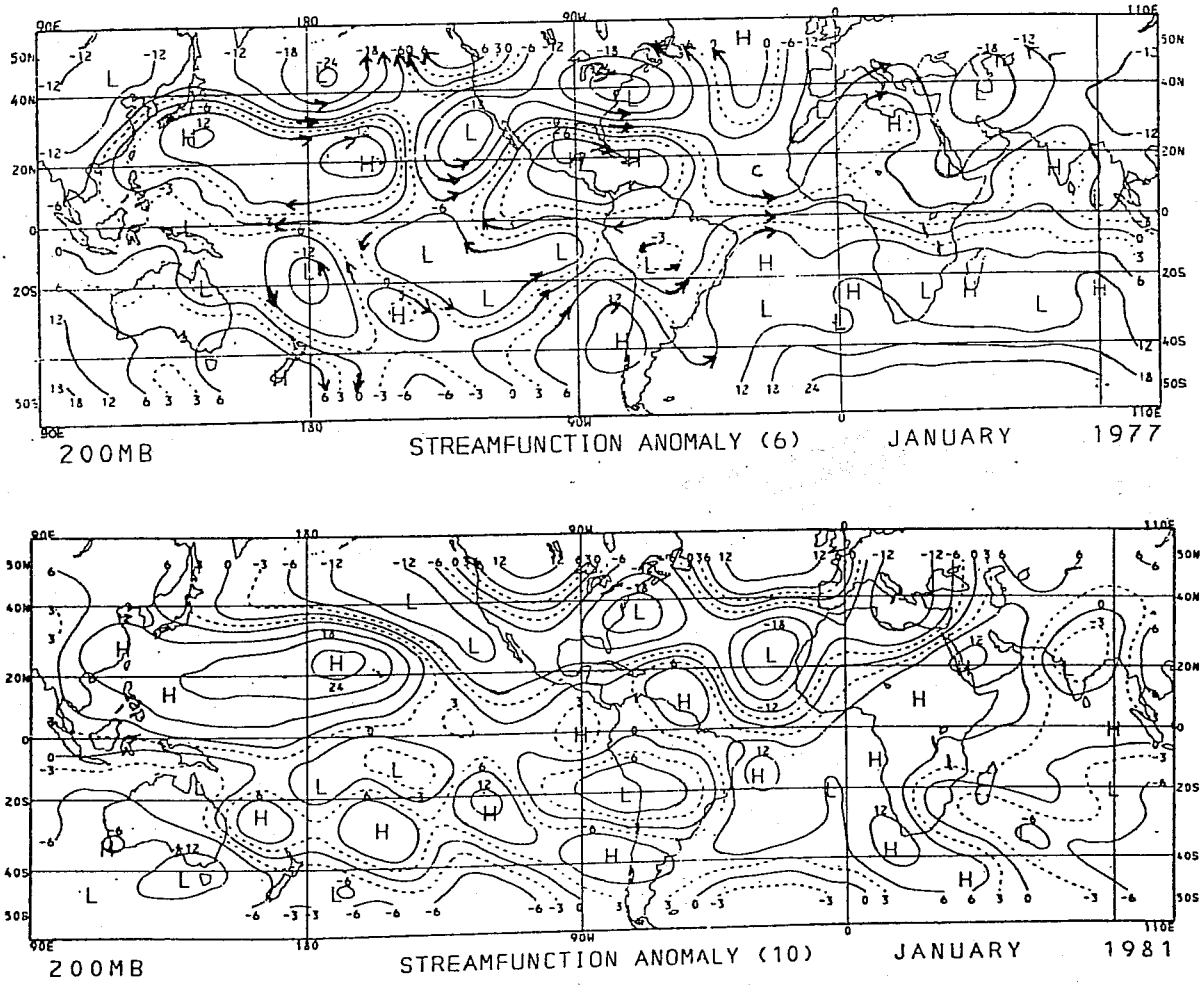


Fig 9 Observed 200 mb streamfunction anomaly (x 10^6 m^2s^{-1}) for January 1977 and January 1981. From Climate Analysis Center archives.

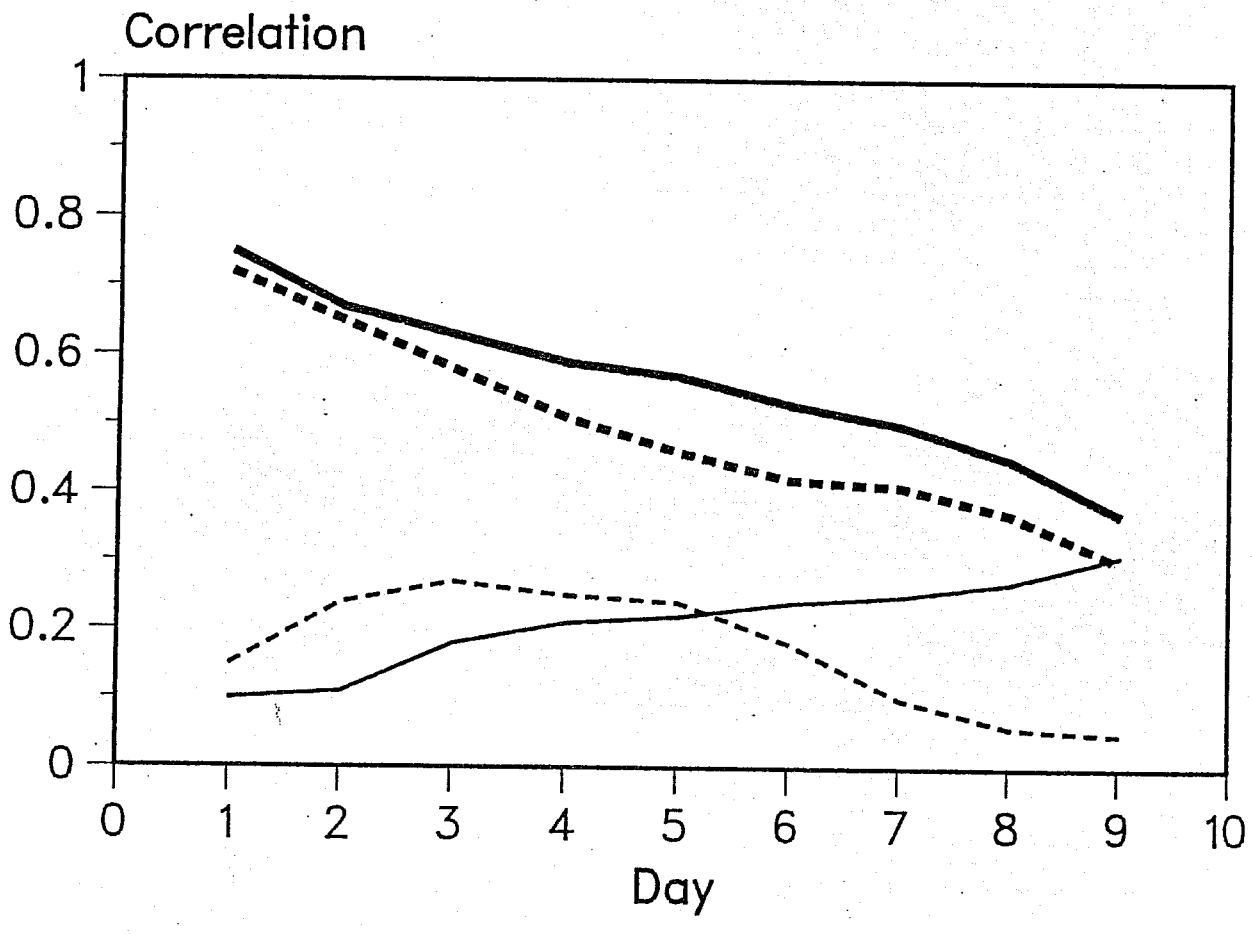


Fig 10 Correlation between forecast skill and RMS observed anomaly (full lines) and between forecast spread and RMS observed anomaly (dashed lines) as a function of forecast time.
Bold lines: ZAM is used to measure forecast skill and spread.
Thin lines: RMS is used to measure forecast skill and spread.

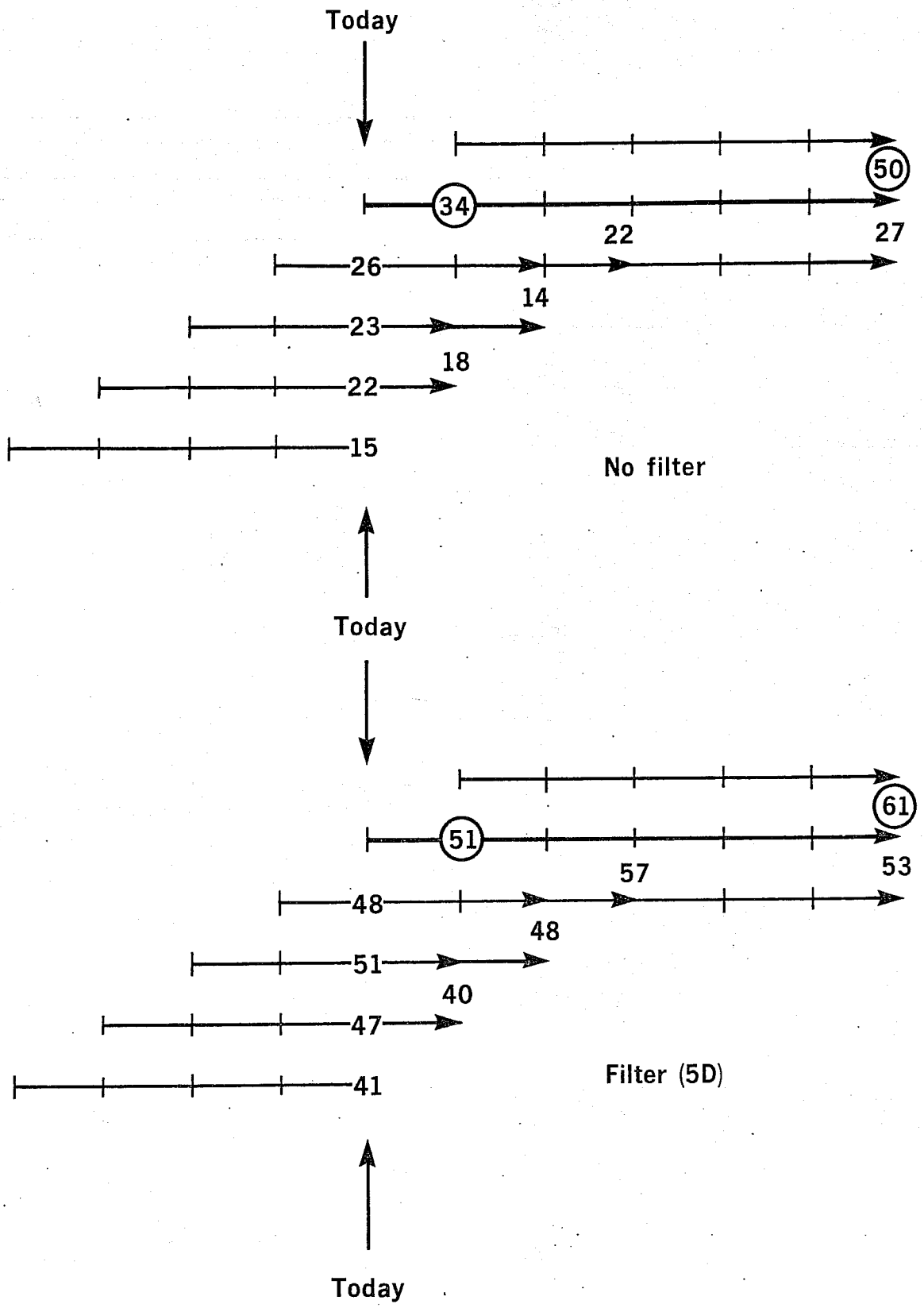


Fig 11 RMS forecast error - day 6 forecasts correlation with forecast spread and previous forecast skill. 600 winter days (1980-81 to 1985-86). Northern Hemisphere - No filter (top) - 5 day filter (bottom). Diagnostic correlations are encircled. For explanation see text.

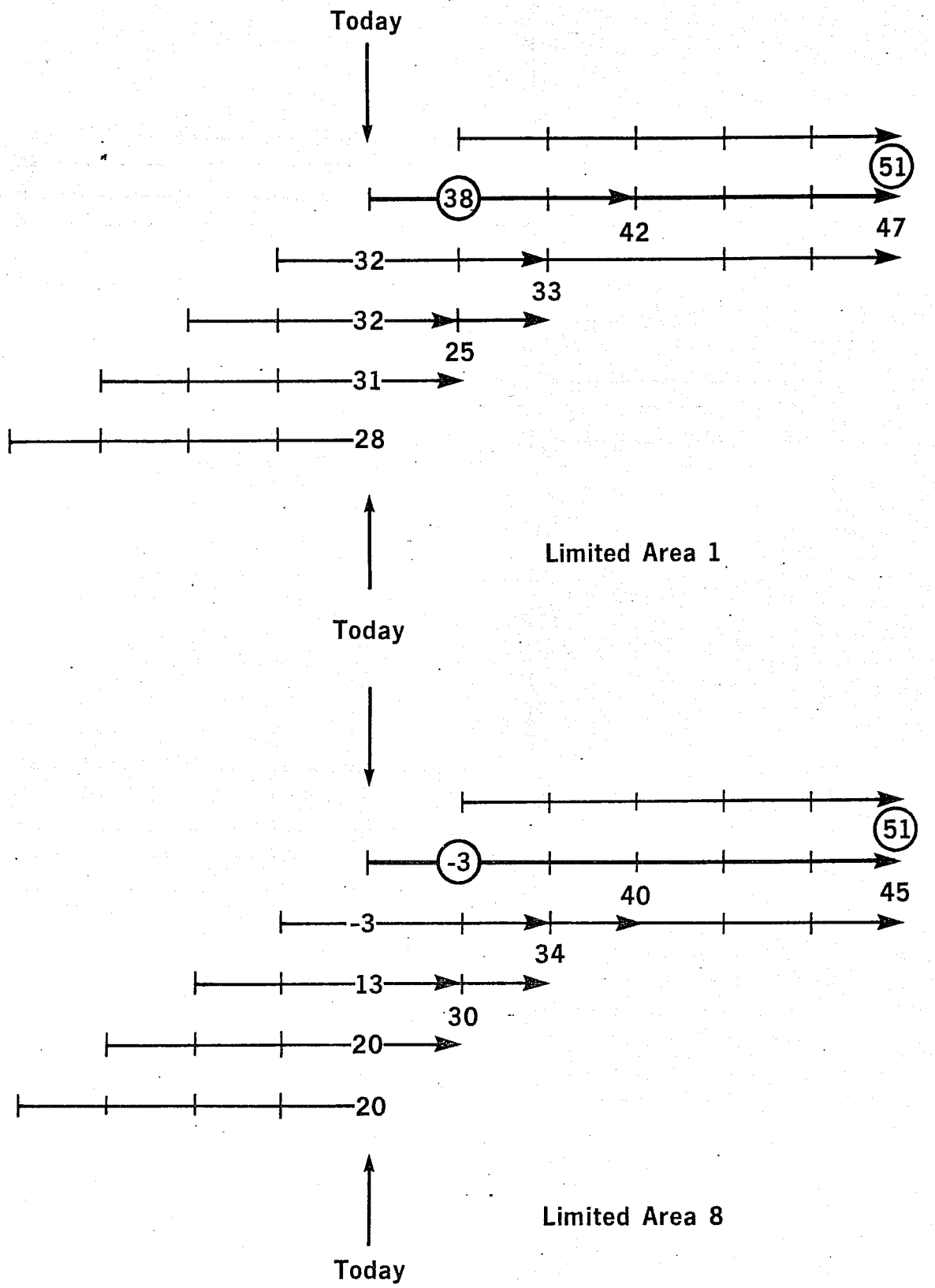


Fig 12 RMS forecast error - day 6 forecasts correlation with forecast spread and previous forecast skill. 600 winter days (1980-81 to 1985-86). Limited areas 1 (top) and 8 (bottom) - 5 day filter is on. Diagnostic correlations are encircled. For explanation see text.

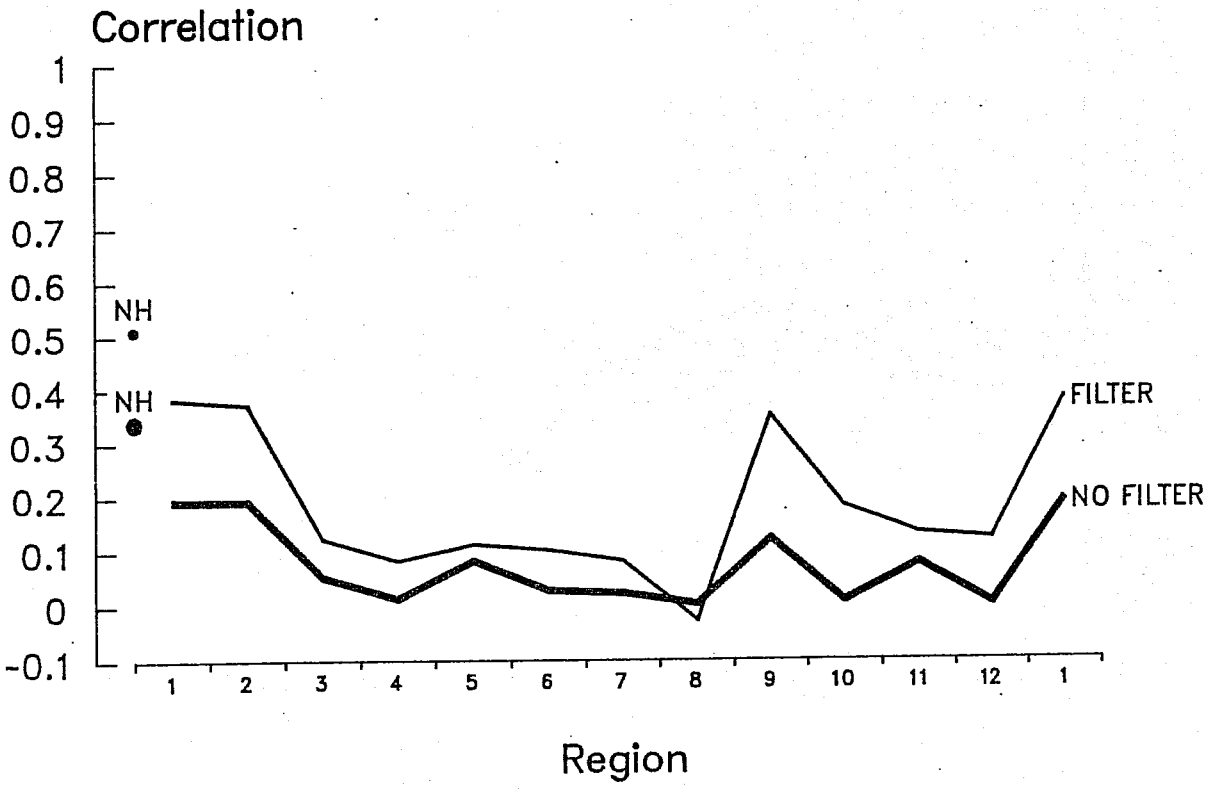


Fig. 13 Correlation between day 1 and day 6 forecast skill (RMS error) with and without 5-day mean filter as a function of longitude (limited area. The black dots on the left represent hemispheric values.

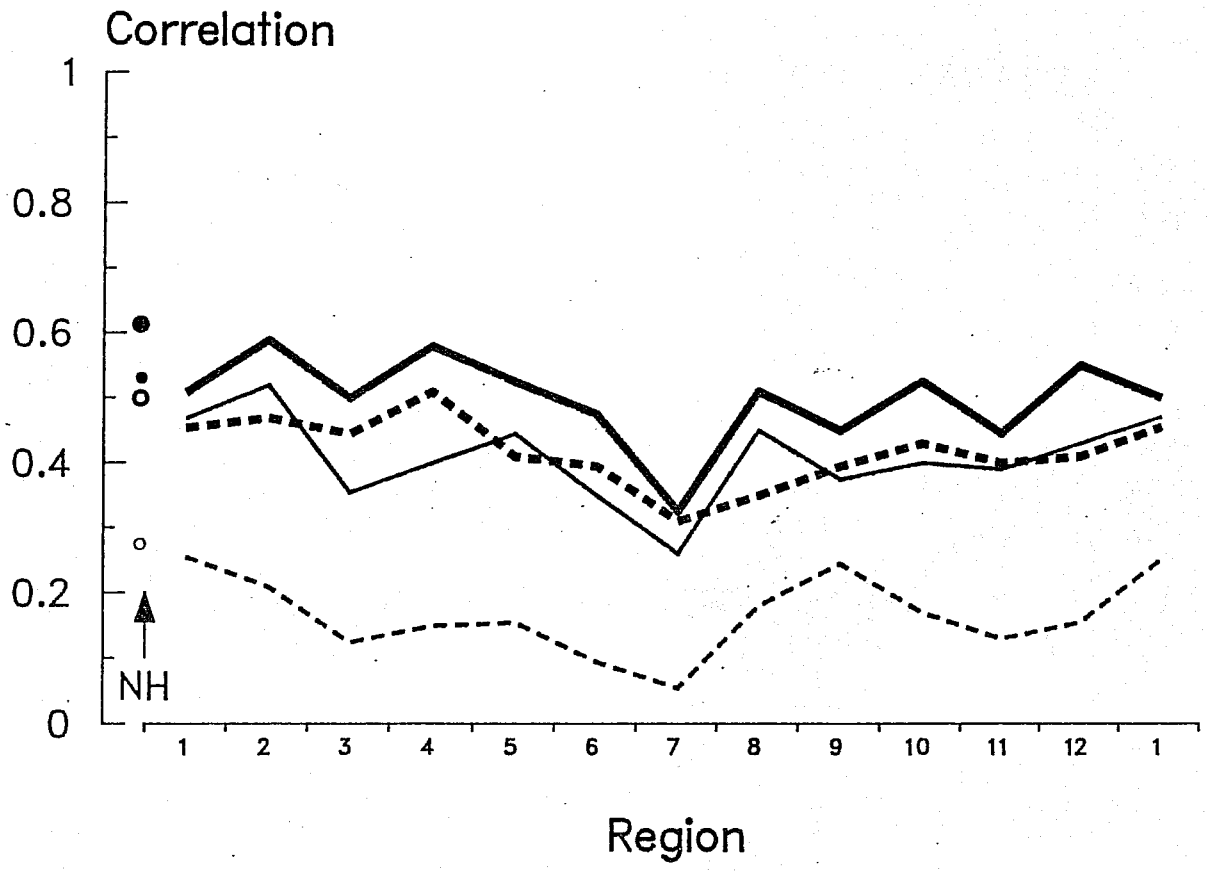


Fig 14 Correlation between day 6 forecast skill and both prognostic (day 6-7) and diagnostic (day 5-6) forecast spread. Bold lines and large dots: prognostic spread. Thin lines and smaller dots: diagnostic spread. Full lines and full dots: 5 day running mean filtered data. Dashed lines and circles: unfiltered data. The dots on the left represent hemispheric values.

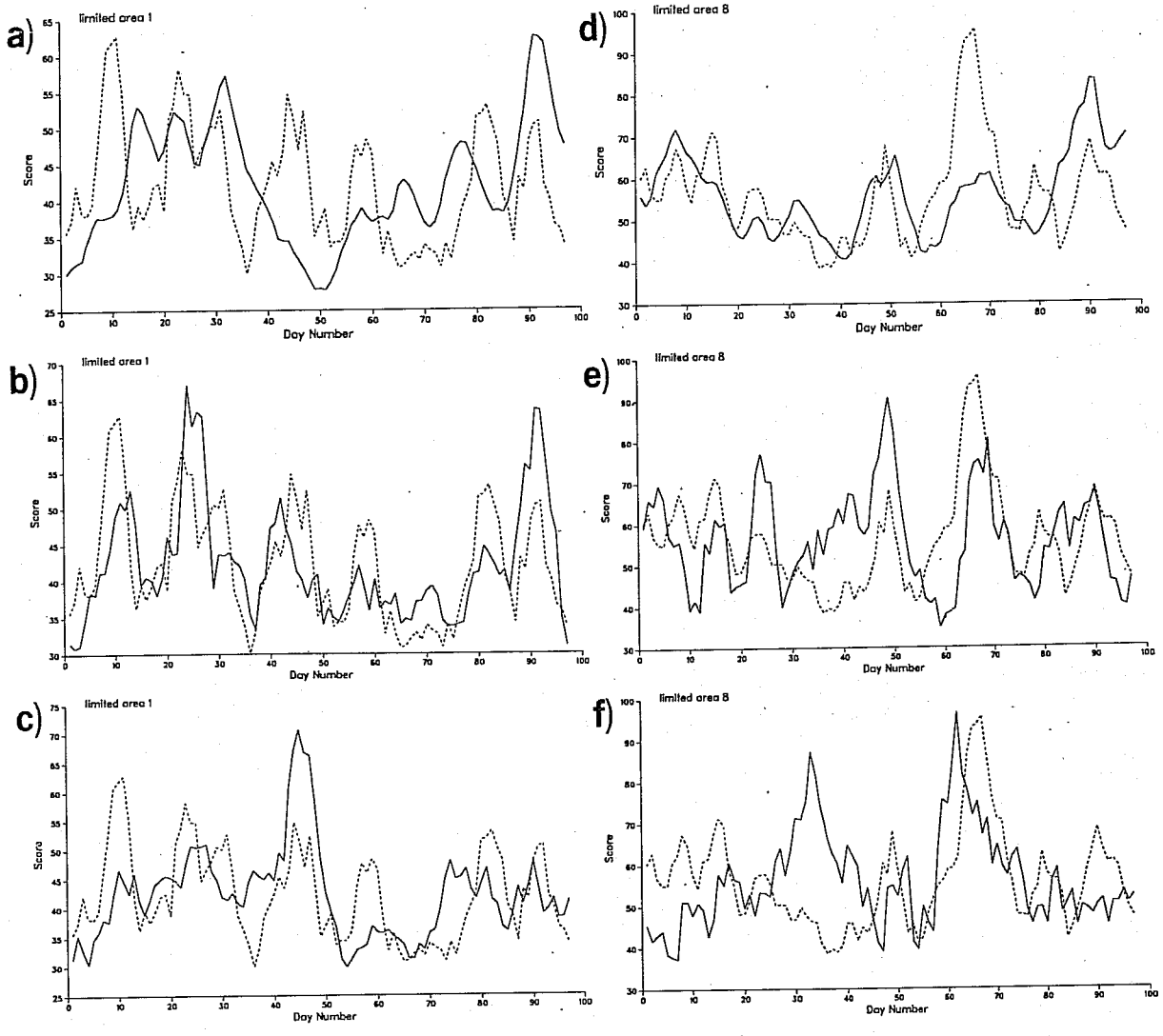


Fig 15 Prediction of day 3 forecast rms error for the winter 1986/7 (day 1 = Dec 1) using

a) skill in region 1, and EOF regression
b) skill in region 1, and day 3/day 4 spread
c) skill in region 1, and day 1 forecast error
d) skill in region 8, and EOF regression
e) skill in region 8, and day 3/day 4 spread
f) skill in region 8, and day 1 forecast error.

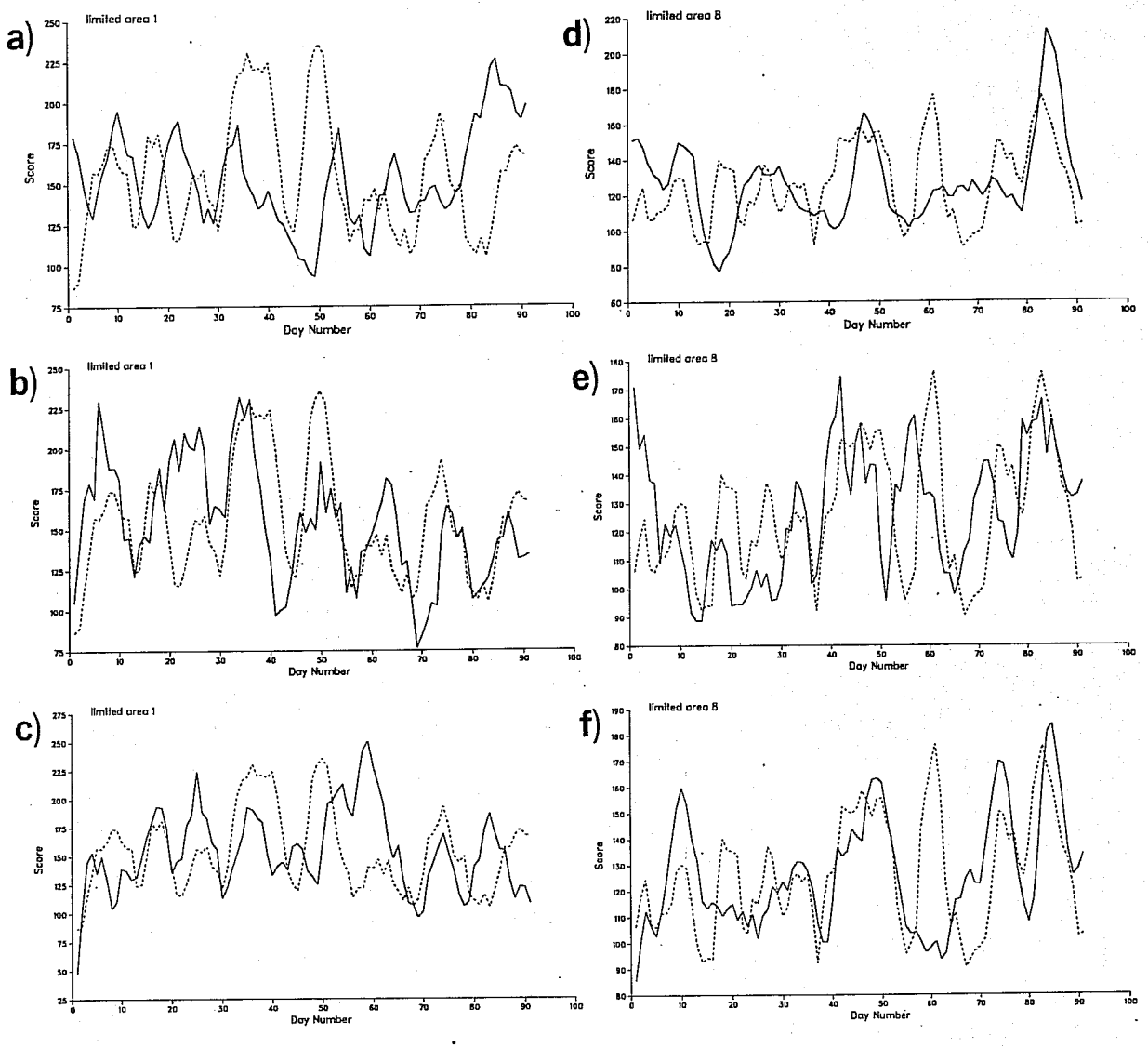


Fig 16 As Fig 15 but for day 9 (and day 9/day 10 spread) and c) and f) showing rms tendency diagnostic

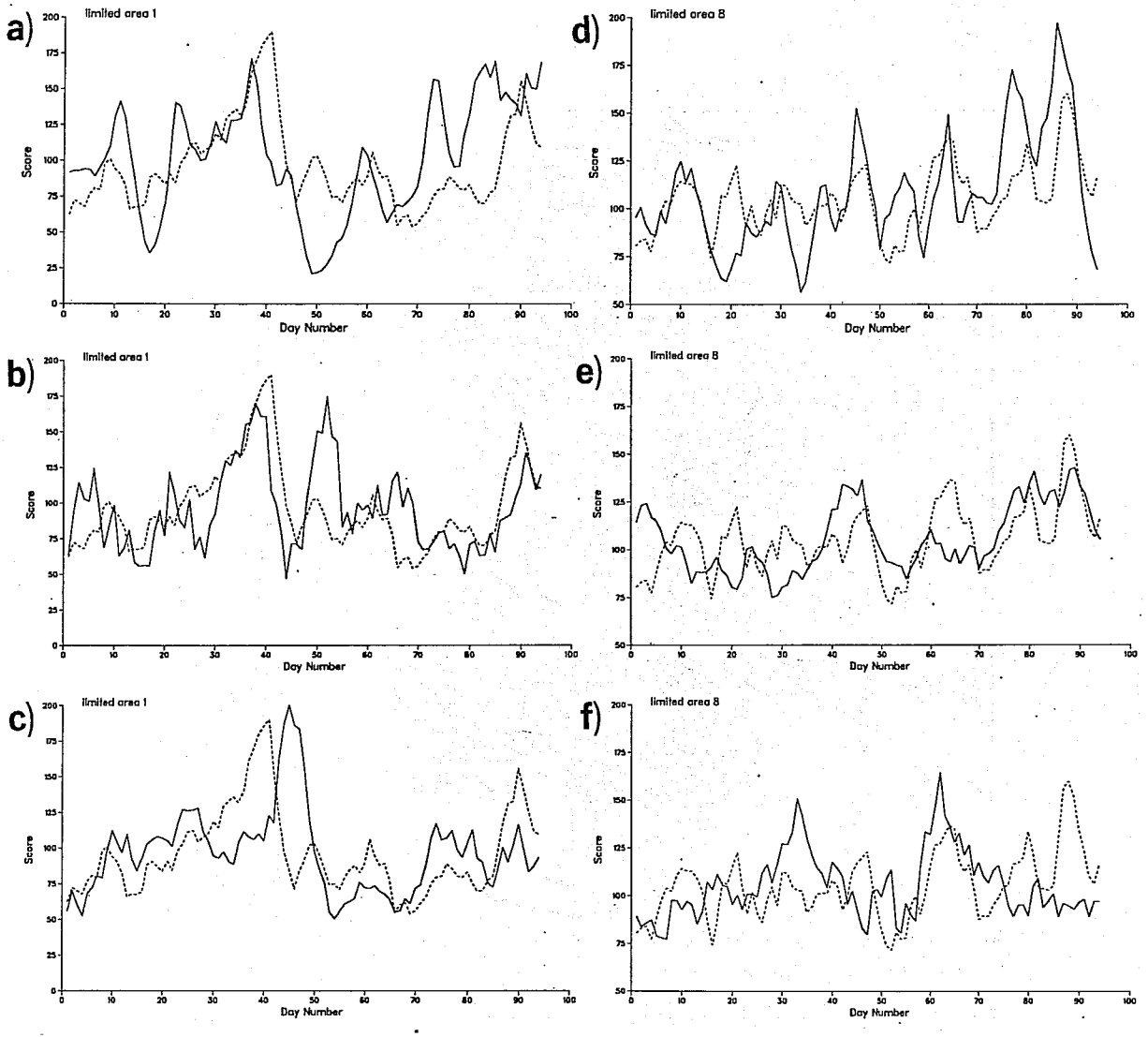


Fig 17 As Fig 15 but for day 6 (and day 6/day 7 spread) and EOF regression based on 1985/6 training data

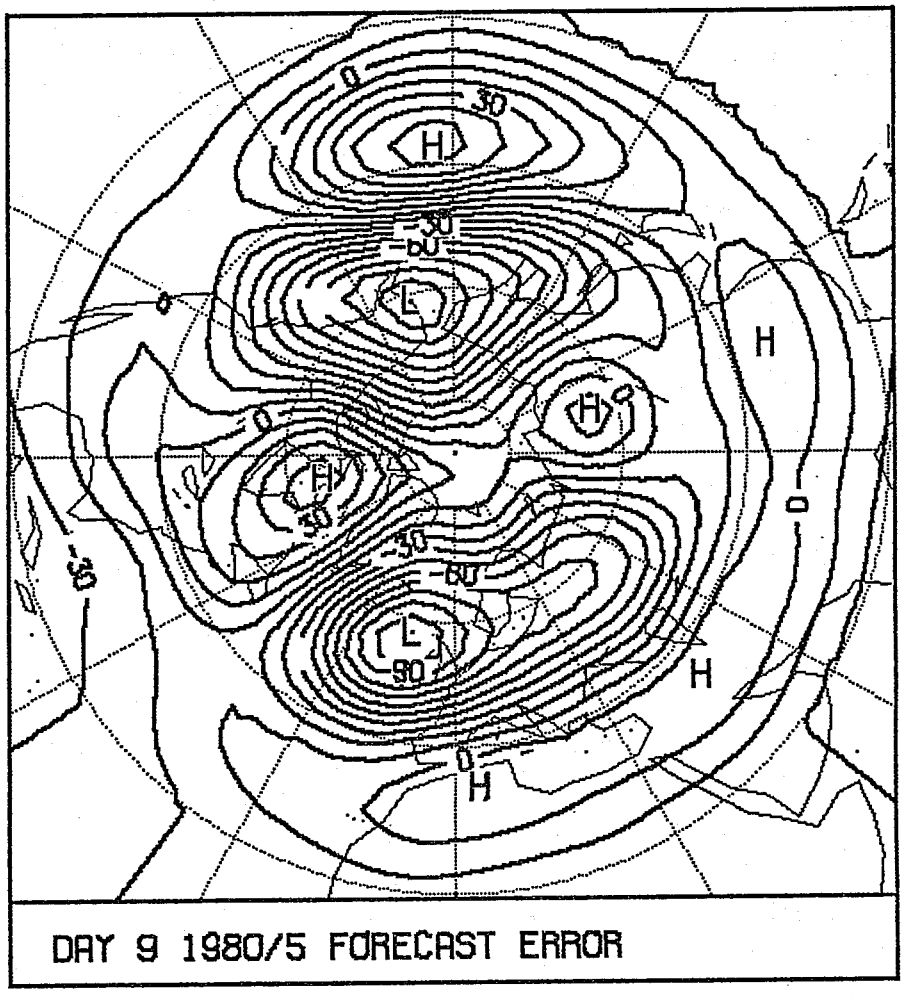
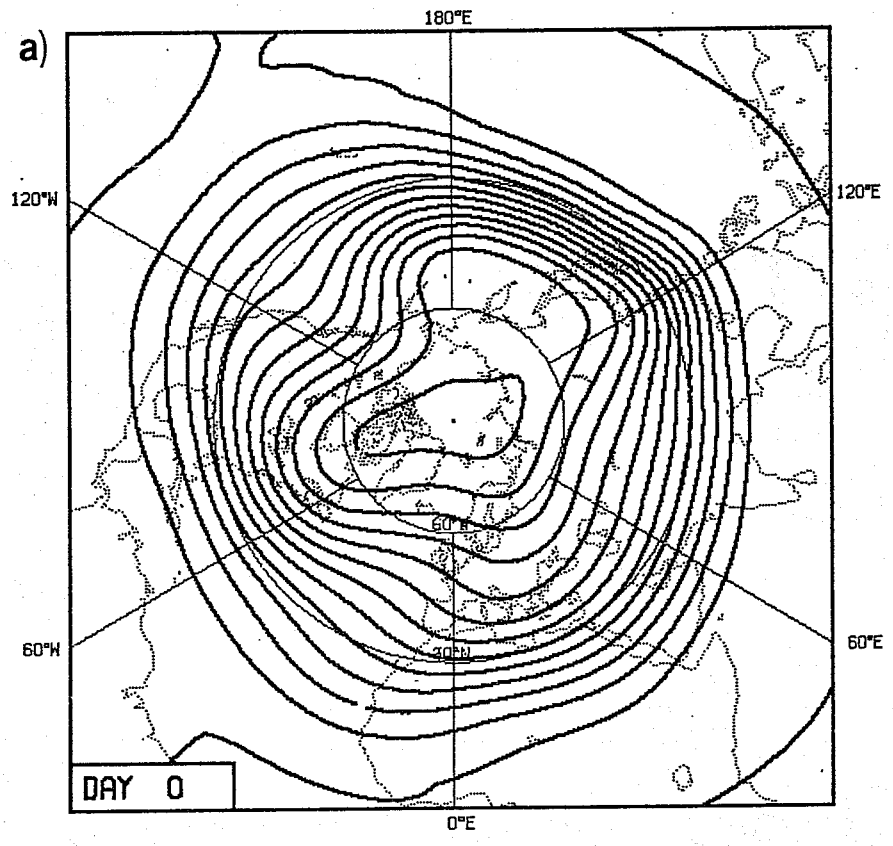


Fig 18 Mean error of the day 9 ECMWF winter forecasts from 1980/1 to 1984/5 sample.



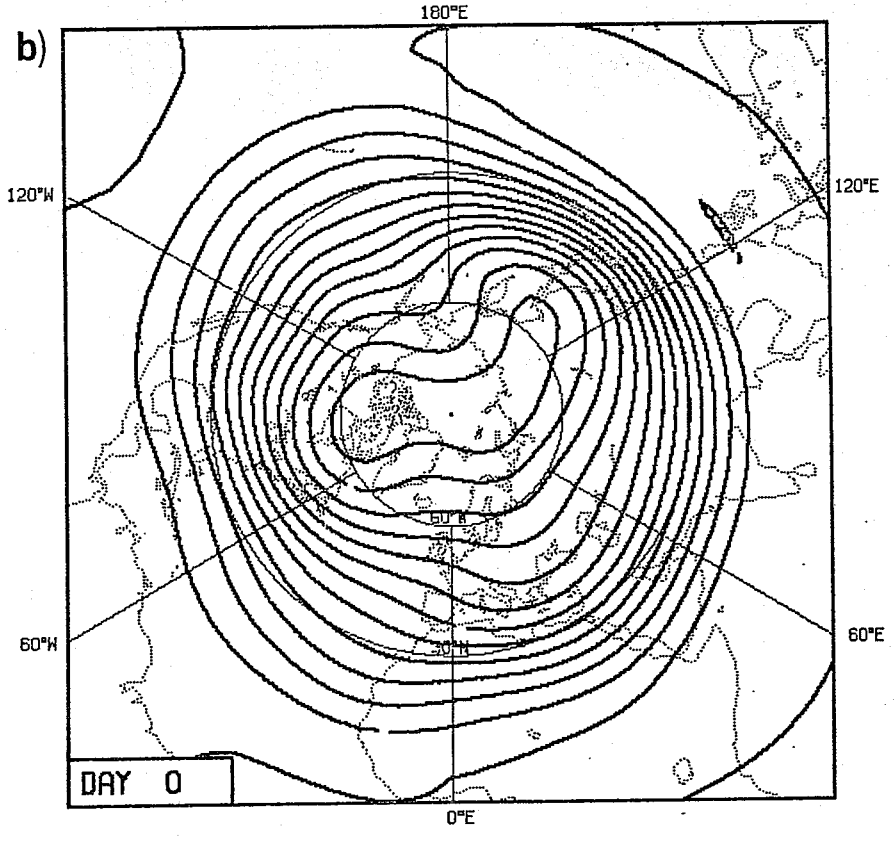


Fig 19 300 mb streamfunction basic states defined by subtracting (B-; Fig 19a) or adding (B+; Fig 19b) onto a 300 mb stream function climatology, the mode shown in Fig 7 associated with variability in forecast skill.

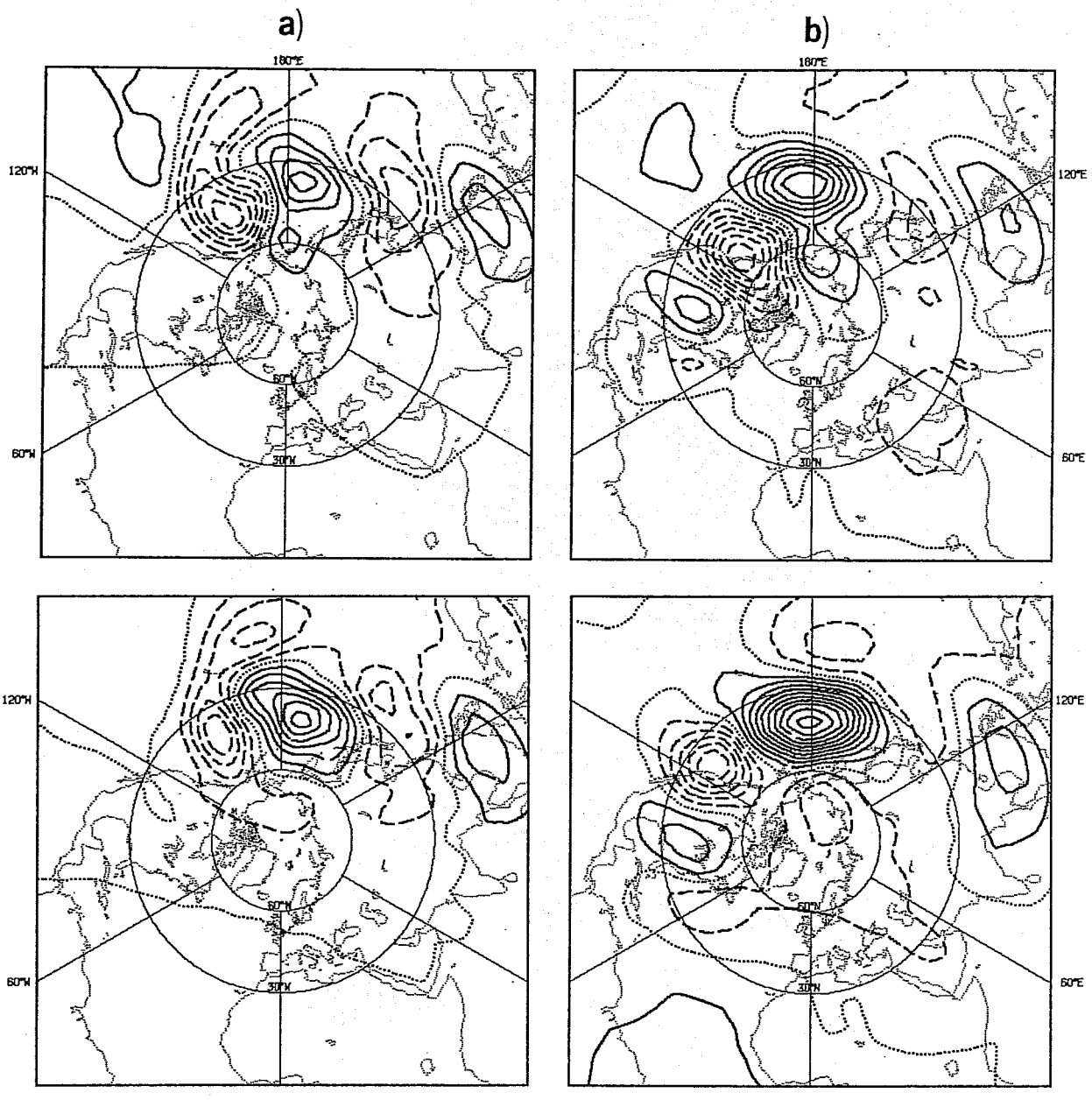


Fig 20 Perturbation streamfunction from integrations of a barotropic model with the two basic states shown in Fig 19, and identical perturbations positioned at 30N, 120E. The top diagrams are for basic state B-, the bottom for B+. a) Day 2, b) Day 4, c) Day 9, d) Day 30. For details see text and Simmons et al (1983).

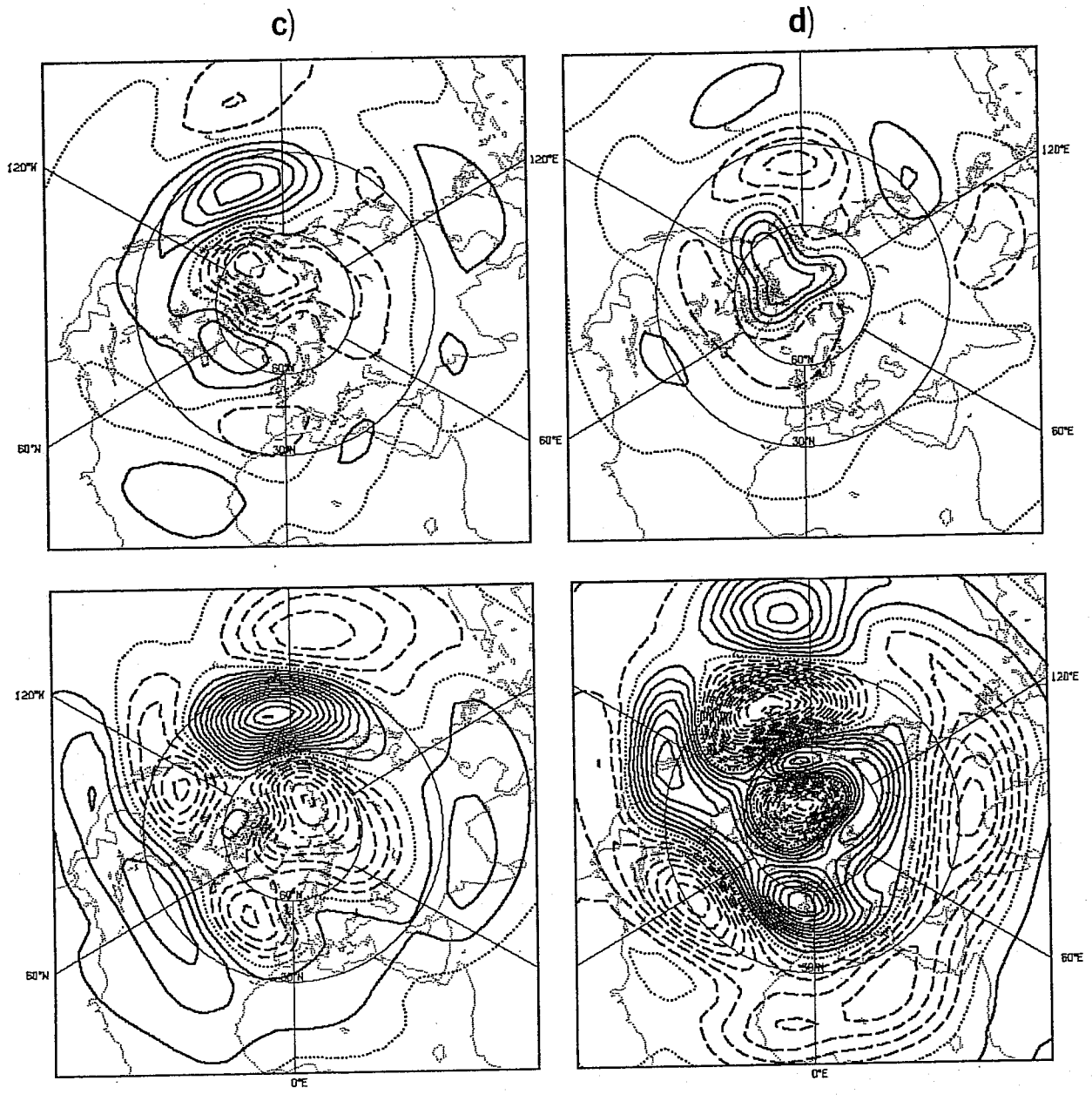


Fig 20 Continued...