TECHNICAL MEMORANDUM

127

Forecast skill and predictability

T.N. Palmer and S. Tibaldi

Research Department

September 1986

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

E.C.M.W.F. ECMWF/SAC(86)6

SCIENTIFIC ADVISORY COMMITTEE SHINFIELD

14th SESSION 28 July 1986

Subject: Forecast skill and predictability

1. INTRODUCTION

Despite impressive improvements in numerical weather prediction over the last few years, forecast models still show considerable day-to-day variability in predictive skill, particularly in the medium range. An example is shown in Fig 1. Whilst it is interesting to understand why such variability occurs, the ability to predict when forecasts are likely to be particularly skilful, or particularly unskilful, is, a fortiori, of great importance. Stimulated by recent discussion with Member States, a project to investigate the extent to which forecast skill is related to the large-scale flow, and to find possible predictors of forecast skill, has recently begun at ECMWF and some early results are recorded below.

Broadly speaking, one can distinguish two sources of forecast error. Firstly, loss of forecast skill can be associated with analysis errors, or errors in model formulation. These are related to deficiencies in the way we choose to observe and simulate the atmosphere; they are essentially 'man-made'. Secondly, there are errors associated with amplification of unavoidable uncertainties in the analysis or model formulation due to the intrinsic instabilities and nonlinearity of the atmosphere. This process cannot be ascribed to man-made errors. For example, no matter how good an analysis system might be, the final product will always represent one of a statistical ensemble of equally probable initial states consistent with the data.

Some sources of variability of forecast error can be ascribed to one or other of these categories. For example, day-to-day changes in data coverage will result in day-to-day variability in forecast skill. Changes in model formulation will result in much lower frequency variability in forecast skill. Furthermore, since stability properties of the atmosphere depend on the

strength of parameters describing the flow itself, loss of skill due to amplification of small initial uncertainties will vary with the initial and forecast flow patterns. However, variability in skill will also be influenced by interactions between the two categories of forecast error. For example, the propagation and amplification of analysis errors depend strongly on the structure of the large-scale flow. Similarly, the way in which systematic model errors influence forecast skill, will depend on the detailed structure of the flow. (For example, one might argue that systematic errors in the model's treatment of tropical convection will most strongly influence the extratropics when the large-scale flow is conducive to the meridional propagation of Rossby-wave activity.)

These remarks serve to illustrate the fact that whilst it may be possible to find relationships between the large-scale flow and forecast skill, the reasons for such relationships may be not be unique. In this text, we shall report the results of some statistical tests to try to discover significant relationships; however, at the present stage of the investigation, we can only speculate on possible reasons for such relationships.

Our ultimate goal in this study is to try to find a set of predictors that can be used in an operational sense to 'forecast the forecast skill', and we have examined a variety of such candidates. These include not only functions of the large-scale flow itself, but also variables expressing the skill of previous forecasts, and the consistency between forecasts initialised one day apart. These latter variables have often been cited as possible predictors of forecast skill.

The outline of the paper is as follows. In section 2 we give a brief survey of previous work in this area, and in section 3 we outline some of the work that has been done at ECMWF. Section 3 is divided into three parts. In 3.1 we describe the data analysis required for the subsequent investigations. In 3.2 we give results from correlations using forecast skill and forecast spread. In 3.3, results of correlations between forecast skill and the large-scale flow are discussed, both for hemispheric and limited area scores. Some concluding remarks are given in section 4.

2. SURVEY OF PREVIOUS WORK

Although numerical forecasts have been in operational use for several decades, comparatively little work has been done to relate quantitatively objective forecast skill to the skill of previous forecasts, forecast consistency or synoptic flow type. Subjective relationships between such quantities have been devised by bench forecasters to provide day-to-day guidance but not much systematic material has been available to substantiate these relationships.

Some early work in this field was presented at the 1982 Seminar/Workshop on "Interpretation of numerical weather prediction products" by Grønas (1983) and Åkesson (1983). Grønas found that, during persistent Euro-Atlantic blocking and deep cut-off low situations, medium-range forecasts show above average anomaly correlation coefficient in the European area, while during more zonal flow regimes, the model's performance is poorer over Europe. Conversely, if onset of Euro-Atlantic blocking takes place after day 3 to day 5, the model often fails to predict it.

Results consistent with these ideas and with the downstream development theory of Simmons and Hoskins (1979) were produced by Klinker (personal communication), elaborating on the work of Hollingsworh et al (1985), whereby the skill of day 7 forecasts (over the Northern Hemisphere) was shown to be related to activity over the Northern Pacific, though only during winters not dominated by highly meridional (blocked) circulation over the Atlantic region such as 1980-81 and 84-85 (see Fig.2).

Åkesson, from his preliminary investigations, could not draw definite conclusions about the usefulness of consistency between recent forecasts and forecast skill (see also Section 3.2). He also investigated relationships between day 5 forecast error patterns over Europe and synoptic characteristics of the initial conditions, and found some indications that strong negative errors over Europe were often associated with low index (meridional-type) flow situations over the American quadrant.

Grønas, in a later study (Grønas, 1985) also investigated forecast skill variability over Europe, using spatially truncated fields (T10) during both winter and summer periods. He defined subjective binary indicators of Euro-Atlantic blocking and of Pacific activity and found correlations around .5 between a combination of such indicators and objective forecast skill over Europe. Consistency between adjacent forecasts alone gave lower correlations with forecast skill than when combined with these synoptic subjective indicators. In the latter case the correlation reached values in excess of .6 for day 5 forecasts. He concluded that the relationship between forecast skill and flow type should be further investigated.

Recently Branstator (1986) has studied NMC Northern Hemispheric forecast skill in winter periods 1974 to 1985 and found significant correlations between forecast skill and both RMS amplitude of the anomaly and the persistence of large-scale flow. Some relevant results along these lines are to be found in Section 3.2. Also recently, Dalcher et al (1985) reported disappointing results in trying to forecast hemispheric forecast skill using the Lagged-Average Forecasting (LAF) technique (skill-spread correlation), while much more encouraging results were found using a similar technique on regional (limited-area) verifications (Kalnay and Dalcher, 1986 summarized in Kalnay et al, 1986). Instead of using the LAF technique, they constructed an ensemble of forecasts from alternative analyses produced in an Observing System Experiment by removing different components of the global observing system (e.g. satellite temperature soundings, radiosondes, cloud winds, and so on). In this study they found that, although the predictability of the hemispheric forecast skill was low, the skill over Europe and North America (again expressed as anomaly correlation coeffficients of geopotential height) could be much better forecast in terms of average forecast departure. They concluded that, within the limitations of their sample, this method showed the potential of providing a case-to-case a priori estimate of skill.

3. SOME CURRENT RESEARCH AT ECMWF

3.1 Data

The basic data used in this study are taken from six years of ECMWF forecasts. For each (extended) winter season from 1980/81 to 1985/86, the northern hemisphere 500mb geopotential height field of 100 day 1-to-10 forecasts and verifying analyses, from 1 December , were extracted from the archives. The data for the first five years were then concatenated, and the correlation and regression studies described below were performed on these data. The 100 forecasts and verifying analyses from the sixth available winter, 1985/86, are to be used as an independent sample with which to test the reliability of the statistical results.

Measures of forecast skill and forecast consistency were then calculated for the hemisphere north of 20N, and for a number of limited areas (defined below). Skill scores included both Root Mean Square error (RMS) and Anomaly Correlation Coefficient (ACC); for the latter, the 500-day mean of verifying analyses in the concatenated data was subtracted from each field. Forecast consistency was estimated by calculating the RMS difference and ACC between forecasts initialised one day apart. The RMS of the deviation of each forecast and analysis field from the 500-day observed 'climate' (the 'amplitude of the anomaly') was also calculated.

In order to study possible relationships between forecast skill and configurations of the large-scale flow, it is necessary to project the 500mb height data onto a suitable set of basis functions. In principal it would be possible to use, for example, a spherical harmonic basis. However, in an attempt to minimise the number of basis functions required to describe any significant relationships between large-scale flow variability and skill variability, it was decided to project the data onto a set of empirical orthogonal eigenfunctions (EOFs) defined from pentad mean fields from 32 years of wintertime analyses (1952-84, taken from NMC and ECMWF archives). These EOFs were calculated separately on the zonal mean and on deviations from zonal symmetry. The first of these EOFs corresponds to variation in the hemispheric mean height. Then five 'zonal' EOFs were retained, explaining 99.96% of the total variance of the zonal mean fields, and 17 eddy EOFs, explaining 86.9% of the variance corresponding to fluctuations in the

zonally-varying component of the flow. Fig 3 shows the first three zonal EOFs, and Fig 4 the first six eddy EOFs. It can be seen in Fig 4 that the first eddy EOF corresponds to the stationary wave pattern. Higher eddy EOFs have some similarities with the teleconnection patterns of atmospheric low-frequency (e.g. Wallace and Gutzler, 1981). Further properties of these EOFs, including their probability distribution functions, are given by Molteni and Sutera, 1986.

Finally, for each forecast field, so-called anomaly EOF coefficients were also calculated, where the contributions from the annual cycle and systematic forecast errors have been subtracted. This was done by calculating, for every calendar date and every forecast time, an average (a "climate") of each EOF coefficient over the five years. A fifteen-day running mean was applied to these averaged EOF coefficients, and each forecast and analysis field was then expressed as a difference from the appropriate value in this climate.

3.2 Skill-skill and skill-spread correlations

We show here the results of investigations on whether the skill of previous forecasts and/or the spread between adjacent forecasts contain any information about the objective skill of an operational winter forecast. However, before doing so we will briefly discuss how the RMS error and ACC scores compare with each other when they are used as measures of forecast skill and forecast departures (spread between adjacent forecasts).

Fig.5 shows the correlation coefficient between the 500 day time series of ACC and RMS error (lower curve) and ACC and RMS spread (higher curve). Both coefficients increase with forecast time, but the two quantities seem to be slightly more mutually consistent when used as measures of spread rather than skill. They have comparatively low correlation during the early part of the forecast time (60 to 70%) which confirms that the two measures have different characteristics. This also suggests that, in looking for spread-skill relationship, one is likely to find higher correlations between the two quantities where they are measured by the same objective indicator. We will see that this is indeed the case.

We now turn our attention to skill-skill and skill-spread correlations. The large number of such correlations that can be computed poses a problem of data

display, and we will therefore concentrate on day 3 and day 6 forecasts only. Following Branstator (1986), we use the convention of calling 'prognostic', those correlations that could be used to forecast, by statistical means, the skill of an operational forecast at forecast prediction time (i.e. all those that are based on forecast and analysis fields available at the time the forecast is completed); and 'diagnostic', those correlations between forecast skill and 'future' forecast skills and spreads (unavailable for operational use). Figs. 6 and 7 attempt to condense such information for day 3 and day 6 forecast skill respectively. The a) panels display correlations between skill and skill or spread both measured in terms of ACC and b) panels display the same quantities but when RMS departure is used in place of ACC to measure both skill and spread. The panels are arranged as follows. At the intersections between horizontal and sloping lines (the vertices of the rhomboids) the reported skill-skill correlations have been entered, whilst in the centre of diamonds the corresponding skill-spread correlations are shown. Horizontal lines represent single ten-day forecasts (from day 0 to day 10) and (initial-condition) time progresses from bottom left to top right. For example, Fig.6a shows that the skill (measured in ACC) of the day 3 forecast starting from 'today's' analysis, is correlated 86% with the day 4 forecast skill of the same forecast, but only 69% with the day 4 forecast skill from 'yesterday's' analysis, 52% with the day 4 forecasts skill from day before yesterday's analysis and so on. The correlation between the day 3 forecast skill and itself (100%) is not reported and in its place there is an arrowhead, indicating that the entire panel refers to day 3 forecast skill. If one wanted, however, to use such statistical relationships to predict the day 3 forecast skill, one would have to use forecasts that could be verified today. This means that only skill-skill correlations that lie to the left (or exactly on) the vertical line marked A-B could be used. In the case of day 3, Fig.6a, the highest prognostic correlation is with the day 1 forecast skill (39%).

Regarding skill-spread correlations, the number in the centre of a diamond in Fig.6a is the correlation between the day 3 forecast skill and the day N/day N+1 forecast spread (both measured in ACC), where N is the top left corner and N+1 is the bottom right corner of the diamond. For example the day 3 forecast skill is correlated 80% with the day 2/day 3 forecast spread (verifying at the same time), 72% with the day 1/day 2 spread verifying 24 hours earlier and

only 43% with the day 3/day 4 spread verifying on the same day. Here prognostic correlations are those below the horizontal line C-D and diagnostic correlations lie above this line. It should be noticed that the day 0/day 1 spread is equivalent to the day 1 skill (by definition). It should also be pointed out that only correlations ⩾30% have been entered in the panels of Figs. 6 and 7.

Comparing Figs. 6 and 7 and panels a) and b) leads to a number of interesting points:

a) These results are quantitatively consistent with those presented by Grønas (1985) and Branstator (1986).

b) The highest prognostic skill-skill (ACC) correlation is 39% for day 3 and below 30% for day 6. This implies that the skill-skill relationship is very limited as a statistical predictive tool.

c) The highest prognostic skill-spread (ACC) correlation is 43% (day 3/day 4) for day 3 and 35% (day 6/day 7) for day 5. This indicates that forecast spread is a better predictor of forecast skill than the skill of previous forecasts available at forecast time.

d) ACC appears to be, at this stage, a more useful skill indicator than RMS error. Their different behaviour, in this respect, is consistent with the results displayed in Fig.5.

e) The correlation between the skill of a day N forecast (in our case N=3 and N=6) and the day N/day N+1 spread is consistently lower (by almost a factor of two) than the correlation with the day N-1/day N spread.

Fig.8 shows the correlation between the day N/day N+1 spread and the day N-1/day N spread and the day N skill for all forecast times up to 10 days when either RMS or ACC are used as measures of both skill and spread. If we concentrate on the two correlations between skill and day N/day N+1 spread and skill and day N-1/day N spread, both measured in RMS (the same argument holds for ACC), we see that they are different at short forecast ranges but tend to converge to similar (and low) values at 10 days. This implies that, even towards the end of the forecast, the day N forecast contains a measurably higher information content than the day N+1 forecast verifying at the same time.

The poor level of correlation found could be due to at least two factors:

i) the hemispheric scale of the fields employed. Other studies (Kalnay and Dalcher, 1986 and Molteni et al, 1986) would suggest that more local limited area (probably time-lagged) correlations might give better correlation levels.

ii) the size of the sample estimating the spread. In a LAF (or Monte Carlo) environment more than two (up to 10) samples are usually available to estimate the spread.

These possibilities will be investigated further.

There is one more point that emerges from the analysis of the results reported in Figs.6, 7 and 8 that is worth commenting on in more detail. This is point d) above; the apparent higher predictability of ACC as a forecast skill measure when compared to RMS error. This would also appear to be confirmed by other recent studied (e.g. Branstator, 1986 and Kalnay et al 1985).

Fig.9 shows, in addition to the graphs in Fig.8 similar correlations but where skill is measured in ACC and spread in RMS and vice versa. (For the latter the correlation coefficient is plotted with its sign changed to make the diagrams directly comparable). It appears immediately evident that of correlations (at short and medium range) are considerably smaller.

It is well known that the ACC is sensitive not only to phase differences between forecast and verifying analyses, but also on the amplitude of the verifying anomaly (e.g. Arpe, et al, 1985). It is possible that this produces a 'spurious' correlation between spread and skill obscuring the relationship that we are attempting to verify. Fig.10 shows the correlation between skill (measured either by RMS or by ACC) and the amplitude of the forecast (or observed) anomaly. When the forecast errors are small, i.e. at short range, the correlation between ACC skill and size of the anomaly (either forecast or observed) is indeed very high, while it is negligible if the skill is measured by the RMS error.

Fig.11 shows the same correlations but with spread instead of skill. We can verify that again the ACC measure of forecast spread is considerably correlated to the amplitude of either the forecast or the observed anomaly, whilst this is not the case with RMS.

It is therefore concluded that most (if not all) of the apparent extra predictability that ACC has over RMS (that is, the difference between full lines A and B in Fig.9) is due to the relationship between both ACC skill and ACC spread and the amplitude of the forecast anomaly. This is especially evident at very short forecast times, when the observed and forecast anomaly fields are very similar to each other.

3.3 Relationship between forecast skill and large-scale flow

i) Hemispheric skill

In this subsection we consider results from a regression analysis of EOF coefficients against hemispheric skill scores. The type of question that we wish to ask is whether there are certain patterns in the large-scale flow, from model forecast fields or from initial analyses, or both, for which forecast skill is predictable. Consideration of both initial and forecast fields provides some information on the evolution of the forecast. We have therefore studied linear regressions of forecast skill against forecast EOF coefficients only; against EOF coefficients of the initial analysis only; and against forecast and initial EOF coefficients together. As in section 3.2, results are for the 500-day concatenated dataset. Fig.12 a) to c) shows graphs of the correlation between the regressed EOF coefficient and both the Fisher z-transform (e.g. Morrison, 1983) of the ACC (ZAC coefficient) defined as

$$z = \frac{1}{2} \ln \left(\frac{1+\rho}{1-\rho}\right) \qquad \text{where } z \text{ is ZAC and } \rho \text{ is ACC}$$

and the RMS skill scores. In Fig.12a) the model climate and annual cycle have been subtracted from the EOF coefficients (see section 2). It can be seen that correlations for the ZAC coefficient, fi(t), are higher than for RMS error. The curve fi(t) has a maximum at day 2, whilst for the RMS measure, correlations are largest at the end of the forecast, where they are comparable in magnitude with correlations with the ZAC score. (For fields near the beginning of the forecast period, regression against the ZAC coefficient gives, not surprisingly, a higher correlation than against the ACC.) If 'climate' is not subtracted from the EOF coefficients, then the correlations are a little higher (up to 0.06) than shown in Fig.12, indicating that there are correlations between EOFs and skill scores due purely to the annual cycle.

It is interesting to note that the shape of the curve fi(t) in Fig.12a) is somewhat complicated. The maximum at day 2 is followed by a minimum at day 5 followed by a second relative maximum at day 8. Some insight into possible reasons for this can be found by considering corresponding correlation curves for ZAC scores against regressed EOFs for the forecast data and initial data separately. These are shown in Fig.12b), where it can be seen that the correlations, f(t), for forecast EOF coefficients only and correlations, i(t), for initial data only decrease essentially monotonically, (it is quite possible that the small departures from monotonicity are due to sampling

problems.) The function $FI(t) = \sqrt{f^2(t)+i^2(t)}$ (precisely monotonically decreasing), shown in Fig.12c), would give the correlation coefficient between ZAC scores and regressed forecast and initial EOF coefficients, if the forecast EOF coefficients were independent of the initial EOF coefficients. Because of the temporal auto correlation of atmospheric variables, this is not the case. The ratio $R(t)= fi(t)/FI(t)$, shown in Fig.12c), gives a measure of the degree of redundancy between data from the forecast EOF coefficients and the the initial data EOF coefficients. It can be seen that in the early stages of the forecast there is a degree of overlap between both sets of EOF coefficients. However at day 10 the ratio equals 1, indicating complete independence. Hence, the non-monotonic behaviour of the curve fi(t) can be understood if we consider it to be a product of the monotonically decreasing function FI(t) and the monotonically increasing function R(t).

The regression analysis picks out from the forecast EOF coefficients, a pattern of 500mb geopotential height that is representative of the synoptic situations connected to the most 'forecastable' forecast skill. These patterns are shown, with climate subtracted, in Fig.13a) for day 2 forecasts. If the day 2 forecast anomaly pattern is highly positively correlated with the pattern in Fig.13a), then the ACC will be higher than normal; conversely, if it is highly negatively correlated, the ACC will be lower than normal. (If the magnitude of a forecast anomaly is equal to that in Fig.13a), and is perfectly correlated with it, then the expected anomalous ZAC skill score, normalised by its standard deviation, will be equal to the correlation coefficient of the regression analysis.) The so-called factor structure constants of the regression analysis give the correlation between this pattern and the original EOF patterns. For the day 2 regression with ZAC coefficients, the two highest factor structure constants are for the first zonal EOF, which gives the mean strength of the hemispheric westerlies, and the first eddy EOF, which gives the strength of the stationary waves. It easily seen that the pattern of 500 mb height in Fig.13a) corresponds to weaker than normal zonal winds, particularly over the oceans, and weaker than normal stationary waves. Hence we can expect that forecast fields with weak stationary waves and weak zonal flow will be relatively skilful. The pattern of 500mb height anomaly in the initial data that correlates most strongly with the day 2 ZAC coefficients is shown in Fig.13b). The redundancy in using both initial and forecast data for a day 2 regression is apparent.

The corresponding day 2 forecast pattern that correlates most strongly with the day 2 RMS errors, is shown in Fig.13c). The magnitude of the factor structure constants for the first zonal and first eddy EOFs are not as large as for the ACC regression; however, one can see that there is considerable similarity in the two patterns. The sign of the amplitude of the patterns are opposite, of course, since a skilful forecast should have high ACC but low RMS error. The consistency between results for the RMS regression and the ACC coefficient regression is encouraging.

In attempting a physical explanation of the above results, we note that the patterns in Fig.13 correspond to a weakening of the climatological westerlies across the oceanic storm tracks. These storm tracks are associated with regions of genesis and amplification of the most unstable quasi-geostrophic modes of the extratropical atmosphere, the baroclinic waves. In the presence of such waves, small analysis errors or uncertainties will amplify rapidly. Whilst our single level height data cannot determine changes in baroclinity, it is likely that, assuming climatological wind shears, the patterns in Fig.13 do correspond to weakened baroclinity in the storm track region, and hence less rapid amplification of analysis errors.

Shown in Fig.14a is a scatter diagram of the day 2 ZAC coefficients and the amplitude of the regressed EOF coefficient, relative to its mean, of the day 2 forecast fields. The scatter in the diagram reminds us that the regression explains only about 40% of the variance (see Fig.12). In Fig.14b, c we show the subset of points for the years 1980/81 and 1984/85, the first and last years of the concatenated dataset. Comparison of these two diagrams indicates how much changes in model formulation and analysis procedures have improved day 2 forecast skill. We note also apparent indications of interannual variability in the amplitude of the regressed EOF coefficient. However, for both years, a trend between skill score and regressed EOF amplitude can be discerned. This indicates that at least some of the relationship between EOF coefficient and skill scores is associated with correlations between interannual variability in the strength of the large-scale flow independent of changes in model formulation. (The diagrams suggest that a higher regression correlation might be obtained if the ACC scores were first normalised and standardised to values for the version of the model to which they belong.) In Fig.14d), we show a similar scatter diagram for the independent 100 days for 1985/86 using the regression weights from the

500 day concatenated dataset. The relationship between the day 2 skill and regressed EOF coefficient is apparent.

In Fig.15a,b we show the pattern of the day 9 forecast field and the corresponding initial analysis that correlate most strongly with the day 9 ZAC scores. The forecast field pattern is very similar to the day 2 pattern, corresponding to a weak zonal flow and weak stationary waves. The initial data, pattern is somewhat different, however. Over the Pacific and north America, it is almost anti-correlated with the forecast field pattern, and across the Atlantic it is much less zonal.

Hence, if the model develops during the course of the integration stronger than normal westerly flow, relative to the model climate, then the forecast will tend to be unskilful. Put this way the result seems to accord with experience; development of excessive westerly flow is likely to be associated with deteriorating forecast skill.

The discussion above represents a preliminary assessment of possible relationships between hemispheric skill and large-scale flow patterns; research is continuing in this area.

ii) Limited-area skill

In this sub-section we shall outline some results from regression analyses between forecast EOF patterns, and skill scores in pre-defined regional areas.

Before doing so, however, it is worth recalling some idealised barotropic model experiments reported by Simmons et al (1983). Fig.16 shows the amplification of two localised initial perturbations, on a longitudinally varying basic state representative of climatology. The positions of the two initial perturbations are quite different to one another. On day 2 the influence of each perturbation can be seen as a wavetrain propagating downstream from the initial disturbance. By day 10, however, the perturbation streamfunction has amplified in a geographically fixed location over the east Pacific, independent of the position of the initial disturbance.

These results, as pointed out for example by Grønas (1985), may have some application to studies of the behaviour of forecast errors in the short and medium range. The results at day 2 cannot be quantitatively correct, since,

as we have discussed above, the amplitude of downstream propagating errors will depend on whether they propagate into disturbances that are growing due to baroclinic instability. Hence we can expect forecast errors in some spatially localised region of the atmosphere to be correlated with the phase of transient synoptic scale systems.

Since the group velocity of the error disturbance is generally greater than the phase speed of the baroclinic waves, we can expect the day 2 RMS forecast error in a spatially isolated region of one of the oceanic storm tracks to be correlated with a geopotential height pattern with a low centre either over or just upstream of the isolated region. However, results from the barotropic model indicate that forecast errors in the medium range should correlate more with the (equivalent barotropic) long wave structure in the atmosphere, than with more local transient disturbances.

These remarks (made with hindsight!) describe quite well some of the results of the regression analysis of skill scores with forecast EOFs. They also accord with the findings of Wallace et al (1983). Specifically, 12 limited areas were defined from 60N-30N and sectors of 30 degrees of longitude. The regression analysis was performed against both RMS and ZAC scores. In general, higher correlations were obtained when correlating RMS scores. Since the results form the barotropic model above suggest that the strongest differences between day 2 and day 9 forecasts occur over the Pacific, we shall first show results from this region.

Fig.17, then, shows the anomalous pattern of forecast field which correlates most strongly with the day 2 RMS error in three of the limited areas, 180-150W, 150W-120W, and 120W-90W. The first two of these coincide with regions where the day 10 response in Simmons et al's model is strongest. The number at the top left hand side of each diagram gives the correlation coefficient for the regression analysis (using both forecast and initial EOF coefficients). It can be seen that the pattern is indeed mobile with a negative centre positioned within , or just upstream of, the pre-defined area. For the day 9 forecast (Fig.18) on the other hand, the pattern is stationary, independent of the position of the pre-defined area, and is strongly correlated with the Pacific/ North American teleconnection pattern over that

region (Fig.19a). It is interesting to note that, for the day 9 forecast, the correlation coefficient is smallest for the pattern that is least spatially localised over the Pacific/North American region.

Simmons et al (1983) interpreted their results by showing that the climatological flow was barotropically unstable, and that the day 10 perturbation streamfunction can be considered as a superposition of the unstable modes. In turn therefore, the results from these limited area analyses suggests that, if the magnitude of day 9 forecast errors is related to the barotropic stability of the large scale flow, then flows with a large negative PNA index should be more unstable than flows with a large positive PNA index. We hope to test this speculation in the future.

Results from regressions with limited areas over Europe do not appear to show such distinct differences between day 2 and day 9 results, consistent with Simmons et al. In Figs.20 and 21, for example, we show patterns of the day 2 and day 9 forecast field that correlate most strongly with RMS forecast errors in the regions 30W-0, and 0-30E. In all four cases there is a negative centre over the appropriate limited area. For the area 30W-0, there is an interesting transition from day 2 to day 9, towards a highly localised disturbance. The regression correlation is noticeably higher for the more localised pattern. Wallace and Blackmon (1983) have documented teleconnection patterns with base points in this localised area. These are shown in Fig.19c) and d). When the data is band-pass filtered, emphasising fluctuations in the range 2.5-6 days, the teleconnections extend across the Atlantic and into Europe. However, when the data is low-pass filtered, emphasising periods longer than 10 days, the teleconnection pattern becomes highly localised. Hence, as for the Pacific region, the evidence seems to suggest that, in the medium range, forecast errors are related to the low-frequency variability in the atmospheric flow. For the area 0-30E, however, both the day 2 and day 9 forecast patterns appear very similar. As Wallace and Gutzler (1981) discuss, low-frequency teleconnection patterns in the European/Asian area are not as spatially constant as over the Pacific and north America. Nevertheless, the patterns in Figs.20b) and 21b) do have some correspondence with their Eurasian teleconnection pattern (Fig.19b)). Clearly further investigation is required to understand differences between results for the different regional areas.

4. CONCLUDING REMARKS

Due to the preliminary character of this report, only very tentative conclusions can be attempted at this stage. More appropriately, the possibilities for future work will be discussed.

It has been argued above that the choice of objective scores with which to quantify the skill of a forecast is, for a start, not an obvious one. Indeed it has been shown that ACC (and, equally, ZAC) suffers the drawback of being too sensitive to the absolute amplitude of the anomaly. RMS does not have the same problem, but its deficiencies are otherwise well known (i.e. giving unfair advantages to excessively smooth fields). While hemispheric skill-skill prognostic correlations are comparatively smaller, skill-spread correlations using ACC (but not RMS) show practically useful levels of correlations. Such results, however, are influenced by the abovementioned problems affecting the ACC skill score.

Independent results obtained at other centres would, however, suggest that skill-spread correlations computed on limited areas (regionally) and, possibly, at appropriate time lags might yield more encouraging results. Work is in progress along this line.

Regarding relationships between hemispheric skill and large-scale flow, it has been found that, towards the end of the forecast period, both the initial condition and the forecast fields contain independent information about the skill of the numerical prediction. The characteristics of the large-scale flow that seem to be most connected to forecast skill are those expressed by the first zonal EOF, which is proportional to the mean strength of the hemispheric westerlies, and the first eddy EOF, which is proportional to the amplitude of the stationary waves. Anomalously weak (strong) zonal flow and/or stationary waves situations seem to be associated with relatively skilful (unskilful) forecasts.

Results of regression analysis between hemispheric patterns and limited area (regional) skill scores indicate that day 2 forecast patterns that correlate most strongly with RMS errors in the Eastern Pacific/North American quadrant are mobile and tend to be linked in phase with the limited area where the skill score is measured. By day 9, however, the pattern appears to be more stationary, independent of the position of the predefined area and strongly

correlated with the PNA teleconnection pattern over that region. This is consistent with idealized barotropic model experiments. Similar, but weaker, correspondence can be found between the more localized patterns singled out by the forecast skill over the European region and Wallace and Gutzler's (1981) Atlantic and Euro/Asian teleconnection patterns.

Clearly much more work is needed both to enlarge the investigations on the 6 years ECMWF forecast dataset and to complete the interpretation of the results. Areas that need particular attention are skill-skill and skill-spread regional and time lagged correlations, forecast error teleconnection patterns, relationships between forecast skill and stability of the large-scale flow, synoptic activity-forecast skill lagged correlations and stratifications of all such relationships by weather regime, defined either by bimodality indicators of the type derived by Sutera (1986) or by means of EOF coefficients.

Work is also underway at ECMWF to investigate along the lines proposed by Roads (1985), possible relationships between flow energetics and forecast skill (Persson, 1986, work in progress).

References

Åkesson, O., 1983: Evaluation and use of ECMWF forecasts. Proceedings of the ECMWF Seminar/Workshop on Interpretation of Numerical Weather Prediction Products, 13-24 September 1982. ECMWF, Shinfield Park, Reading, UK.

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., 1986: The variability in skill of 72-hour global-scale NMC forecasts. To appear in Mon.Wea.Rev.

Dalcher, A., E. Kalnay, R. Livezey and R.N. Hoffmann, 1985: Medium Range Lagged Average Forecasts. Preprints from the Ninth AMS Conference on Probability and Statistics in Atmospheric Sciences. Virginia Beach, October 1985, 130-136.

Grønas, S., 1983: Systematic errors and forecast quality of ECMWF forecasts in different large-scale flow patterns. Proceedings of the ECMWF Seminar/Workshop on Interpretation of Numerical Weather Prediction Products. 13-24 September 1982. ECMWF, Shinfield Park, Reading, UK.

Grønas, S., 1985: A pilot study on the prediction of medium range forecast quality. ECMWF Tech.Memo.No.119, 23pp.

Hollingsworth, A., A.C. Lorenc, M.S. Tracton, K. Arpe, G. Cats, S. Uppala and P. Kållberg, 1985: The response of numerical weather prediction systems to FGGE level IIb data. Part I: Analyses. Quart.R.Met.Soc., 111, 1-66.

Kalnay, E., R. Livezy, J. Pfaendtner and Y. Sud, 1986: GCM studies on medium and long range forecasting at GLA.

Kalnay, E. and A. Dalcher, 1986, Forecasting forecast skill. Submitted to Mon.Wea.Rev.

Molteni, F., A. Sutera and N. Tronci, 1986a: Probability density distribution of the 500 mb EOFs in a large dataset. Submitted to Mon.Wea.Rev.

Molteni, F., U. Cubasch and S. Tibaldi, 1986b: Monthly 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, Shinfield Park, Reading, UK.

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

Roads, J.O., 1985: Temporal variations in predictability. J.Atmos.Sci., 42, 884-903.

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

Sutera, A., 1986: Probability density distribution of large scale atmospheric flow. Advances in Geophysics, 29, 319-338.

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. and M. Blackmon, 1983: Observations of low-frequency atmospheric variability. Large-scale dynamical processes in the atmosphere. Eds. B.Hoskins and R.Pearce. Academic Press, 397pp.

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.

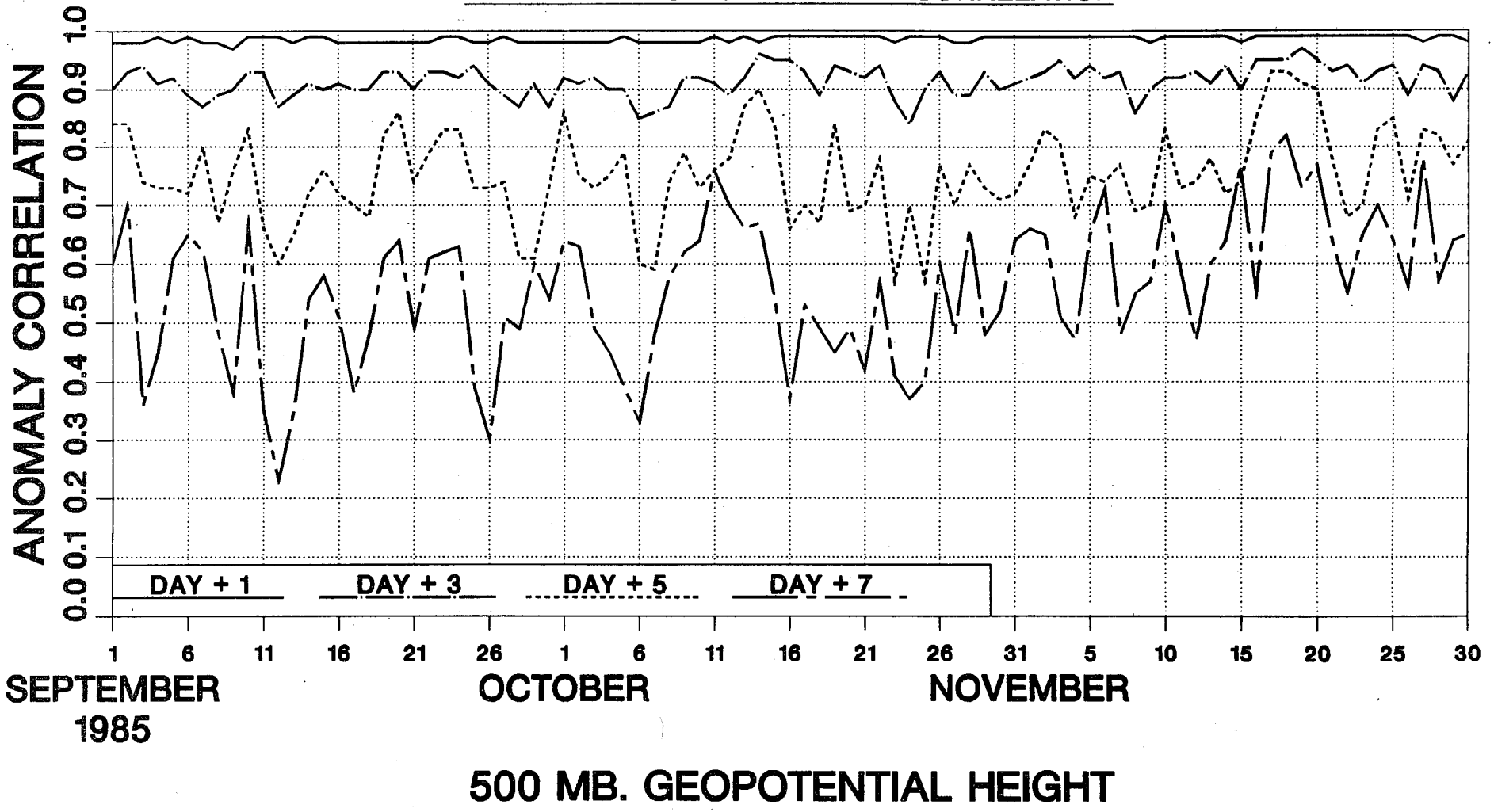


Fig.1 Daily variations of anomaly correlation of NH 500 mb geopotential height for day 1, day 3, day 5 and day 7 ECMWF forecasts. September to November 1985.

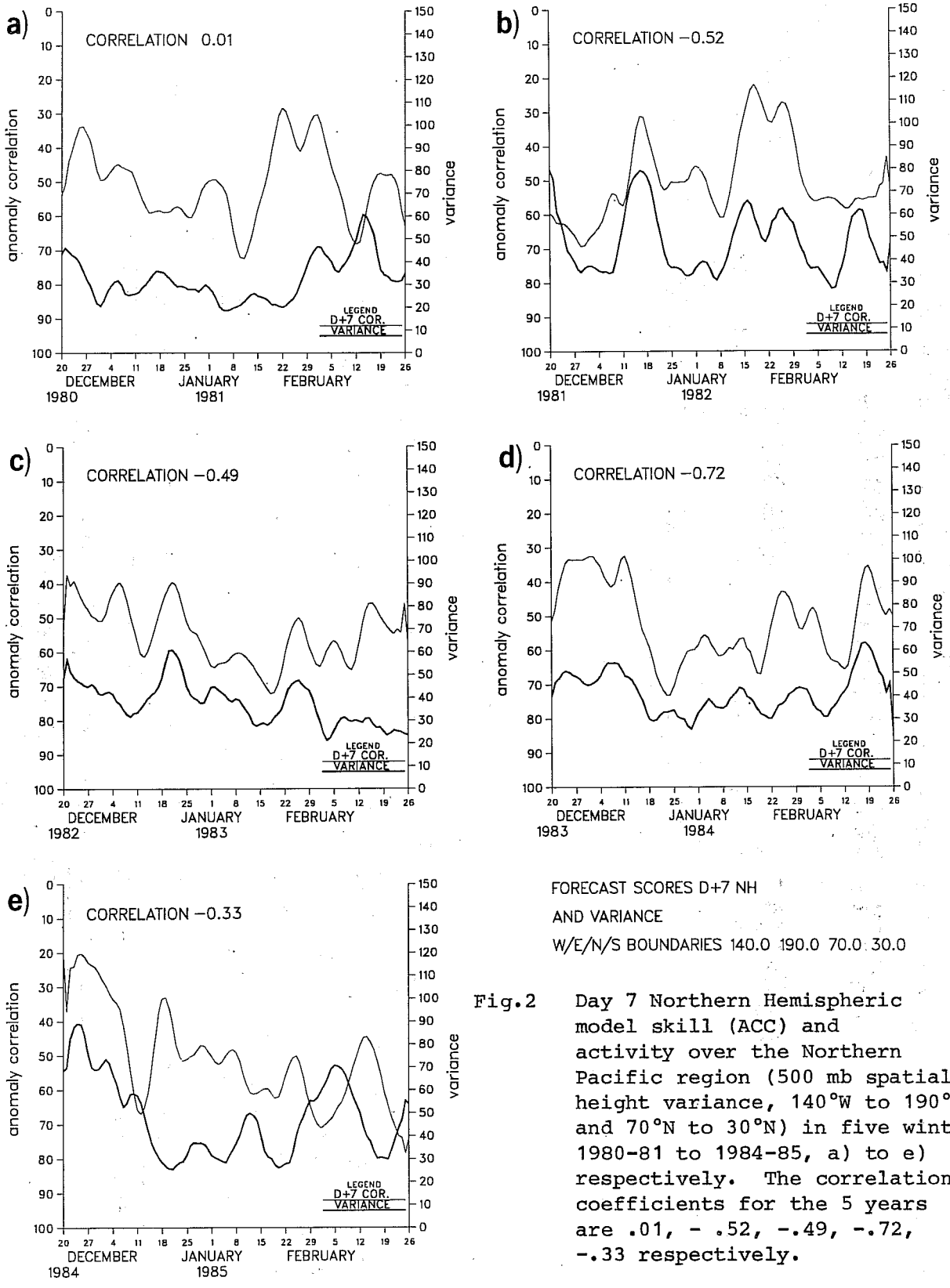


Fig.2 Day 7 Northern Hemispheric model skill (ACC) and activity over the Northern Pacific region (500 mb spatial height variance, 140°W to 190°W and 70°N to 30°N) in five winters, 1980-81 to 1984-85, a) to e) respectively. The correlation coefficients for the 5 years are .01, - .52, -.49, -.72, -.33 respectively.

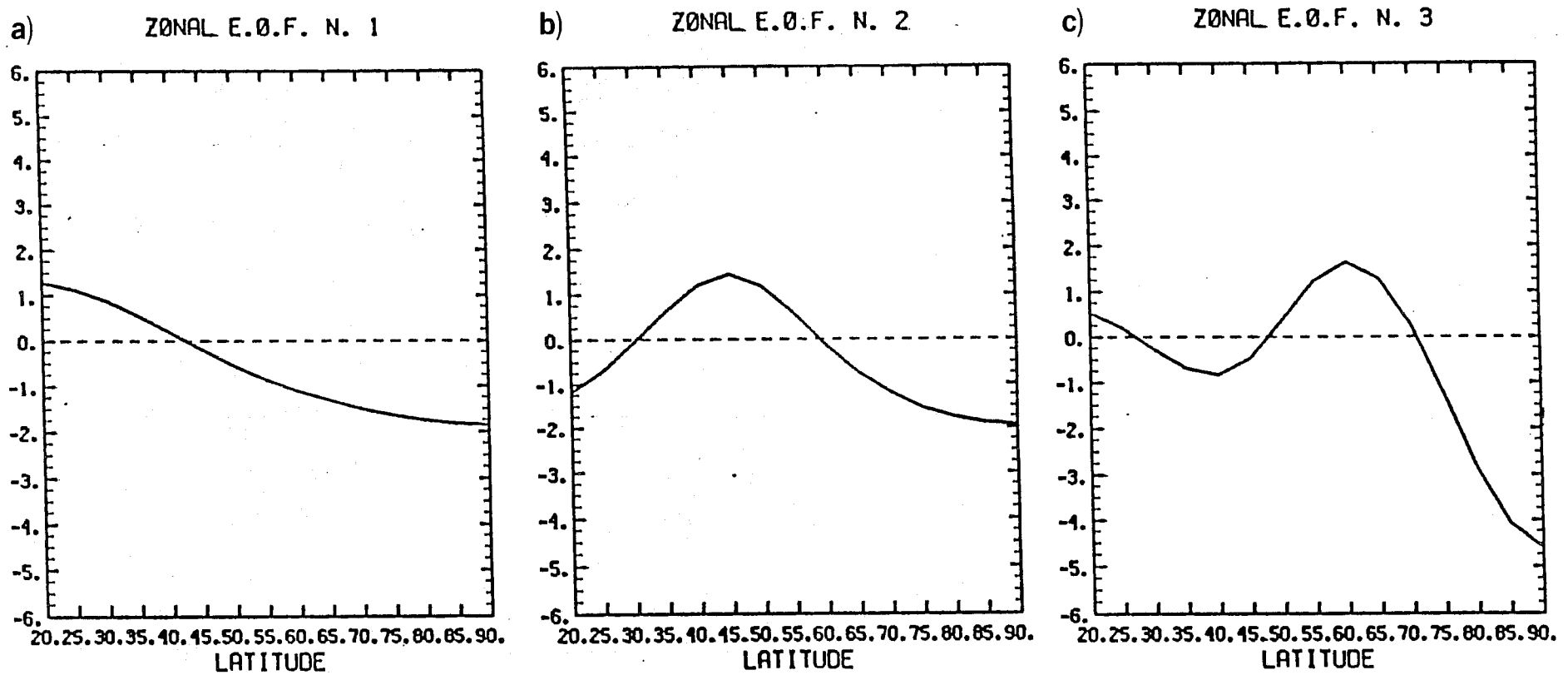


Fig.3 First three NH zonal EOFs, a) to c). Non-dimensional amplitude as a function of latitude.

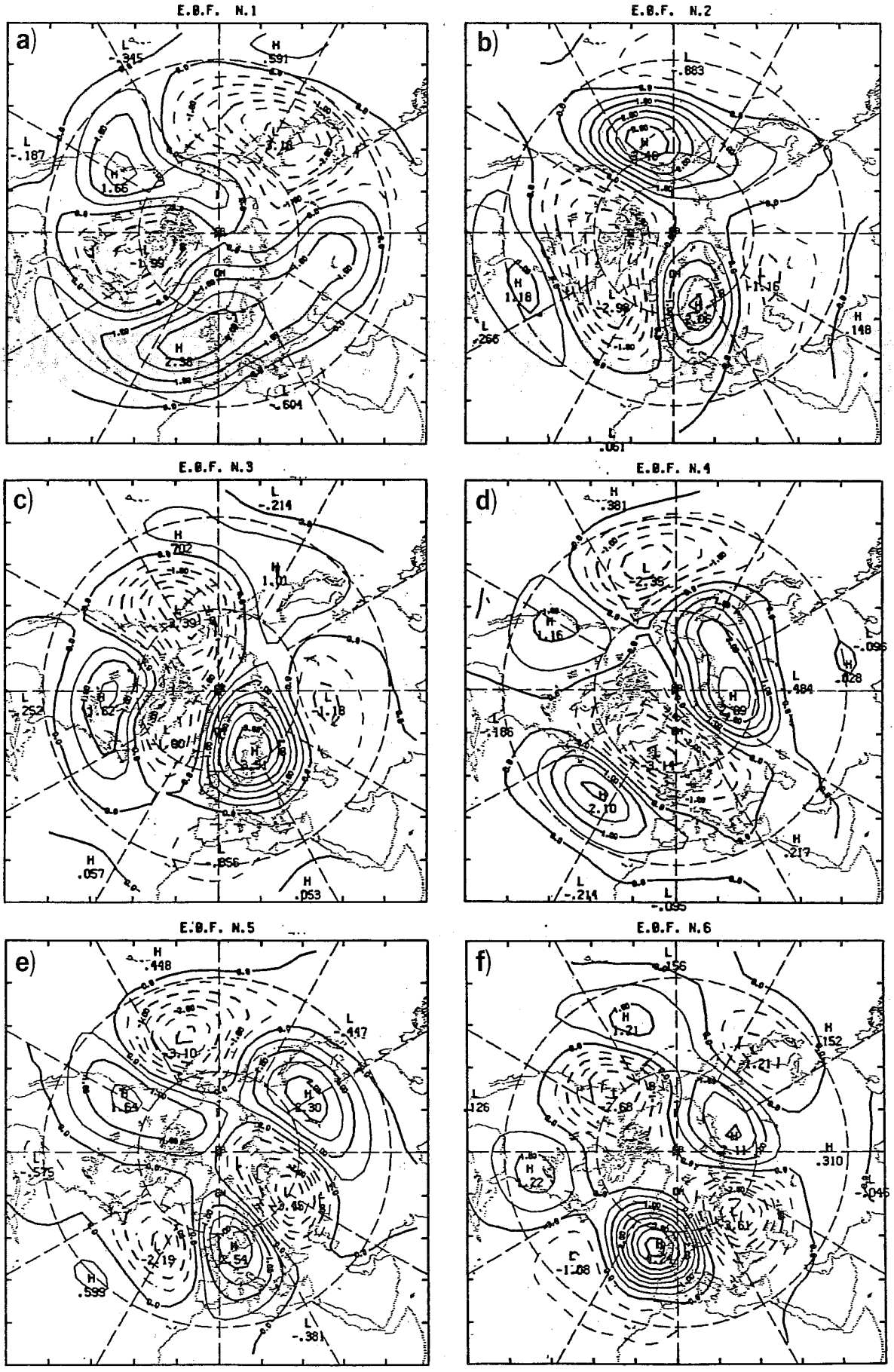


Fig.4 First six NH eddy EOFs, a) to f). Non-dimensional amplitude.

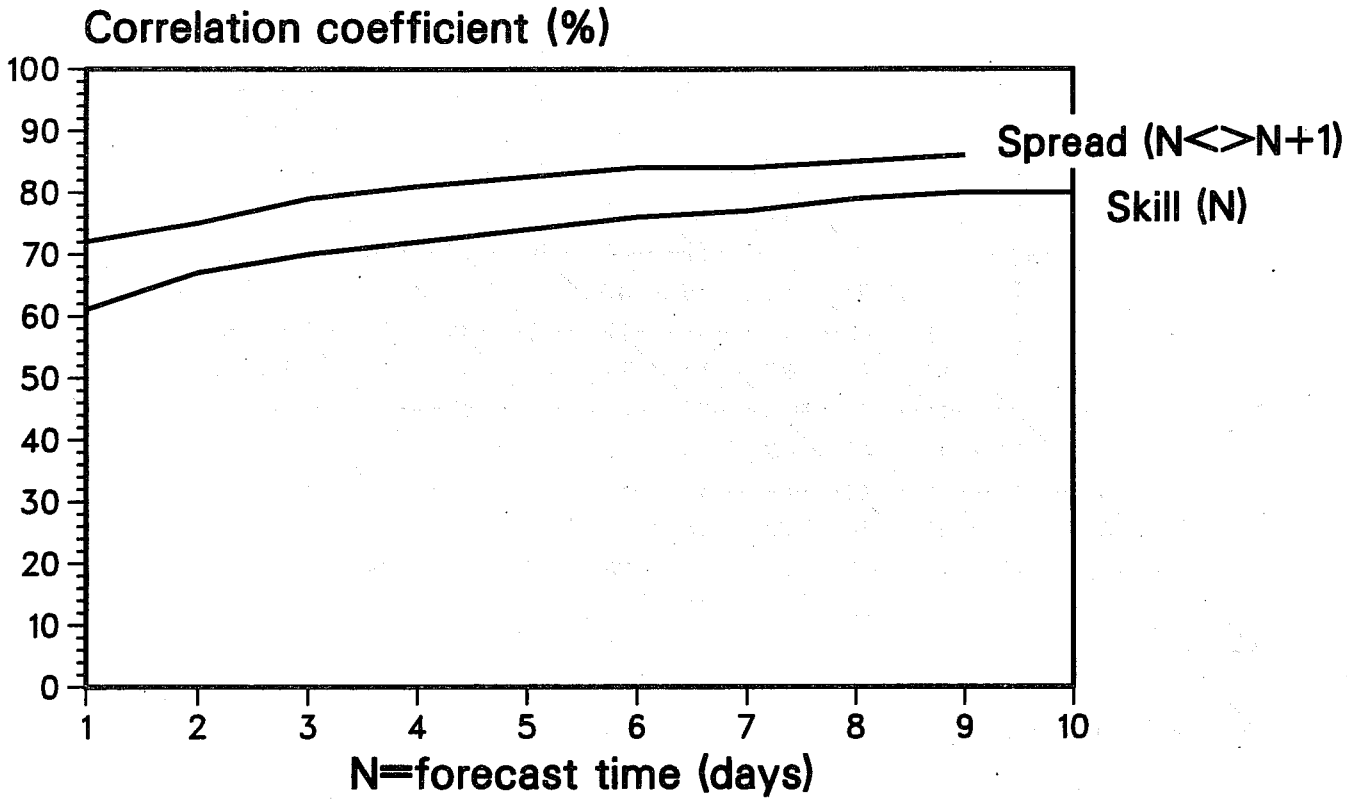


Fig.5 Correlation coefficient between ACC and RMS error (lower curve) and ACC and RMS spread (upper curve between day n and day n+1). 500 winter days 1980-81 to 1984-85, Northern Hemisphere.

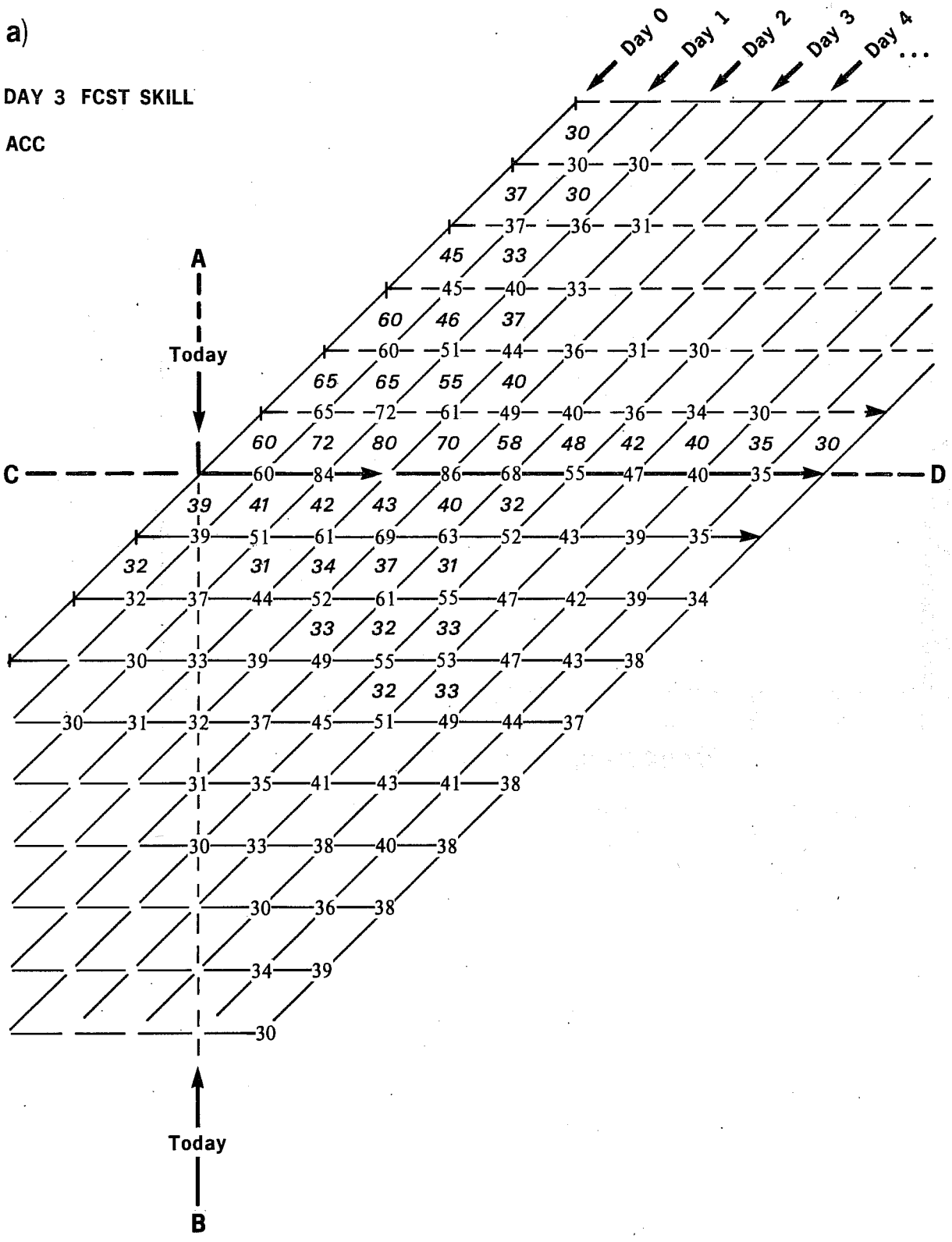


Fig.6 Skill-skill and skill-spread correlations for day 3 forecasts. 500 days concatenated dataset
a) both skill and spread measured in ACC
b) both skill and spread measured in RMS
For explanation see text.

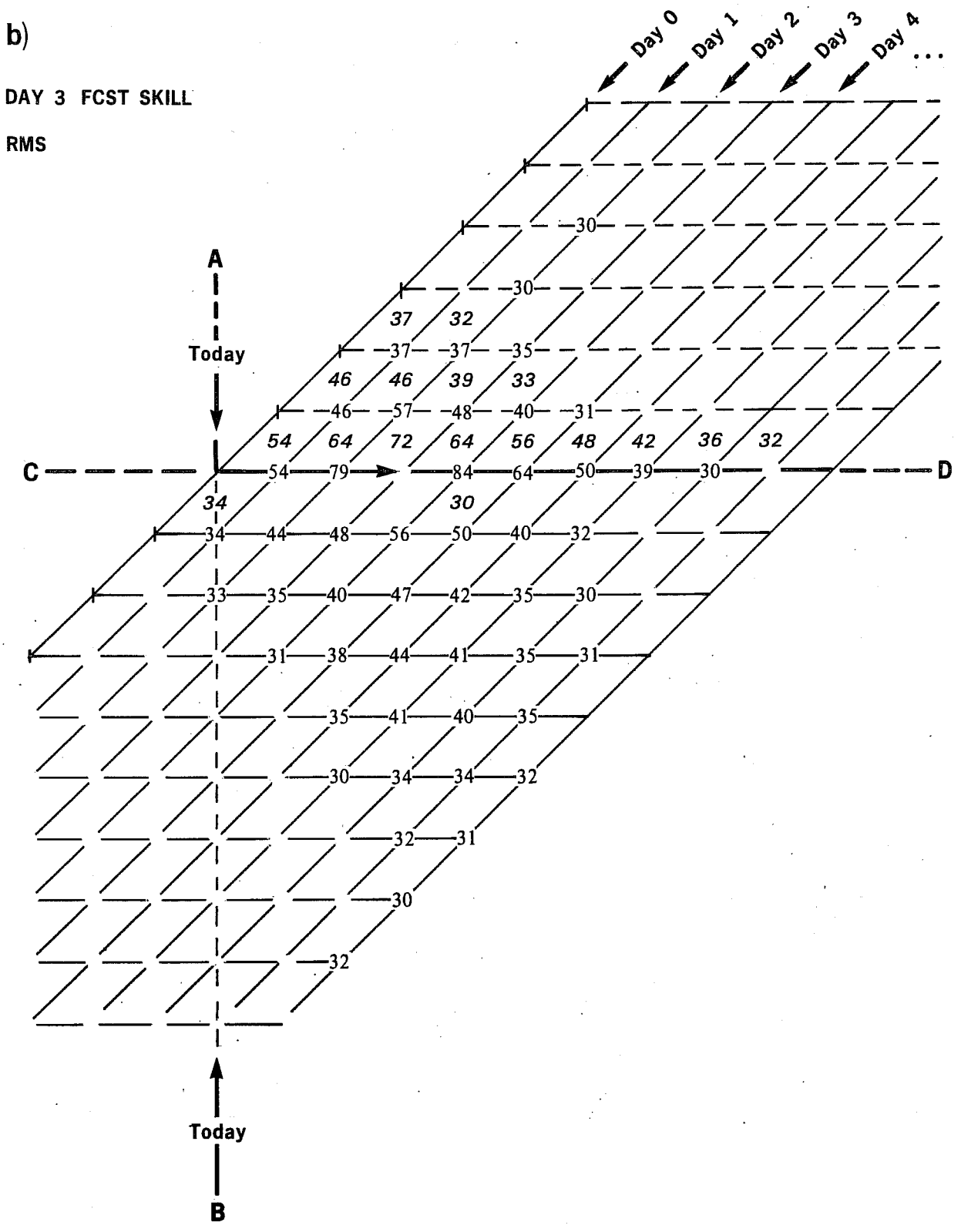


Fig.6 Continued

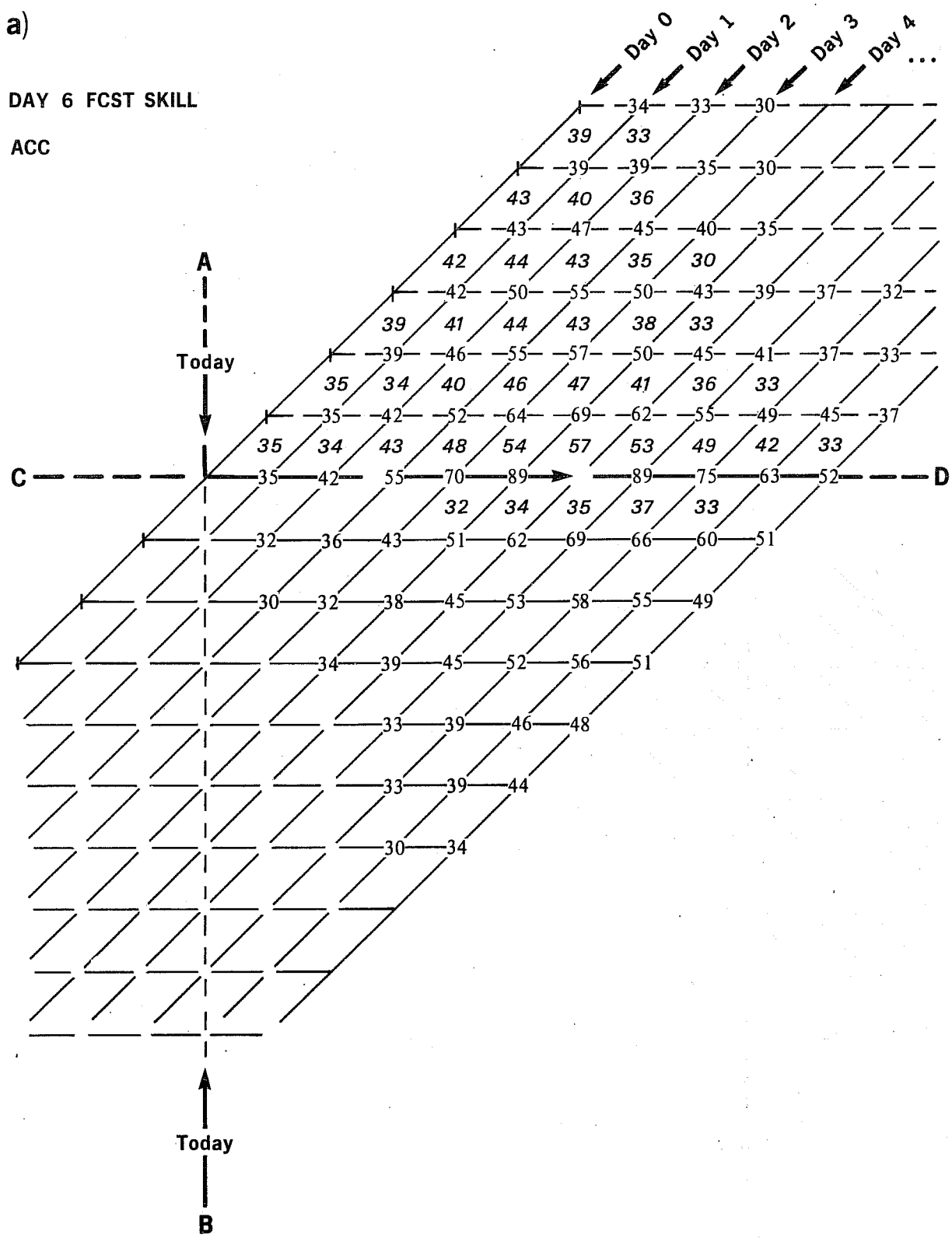


Fig.7 As Fig.6 but for day 6 forecasts.

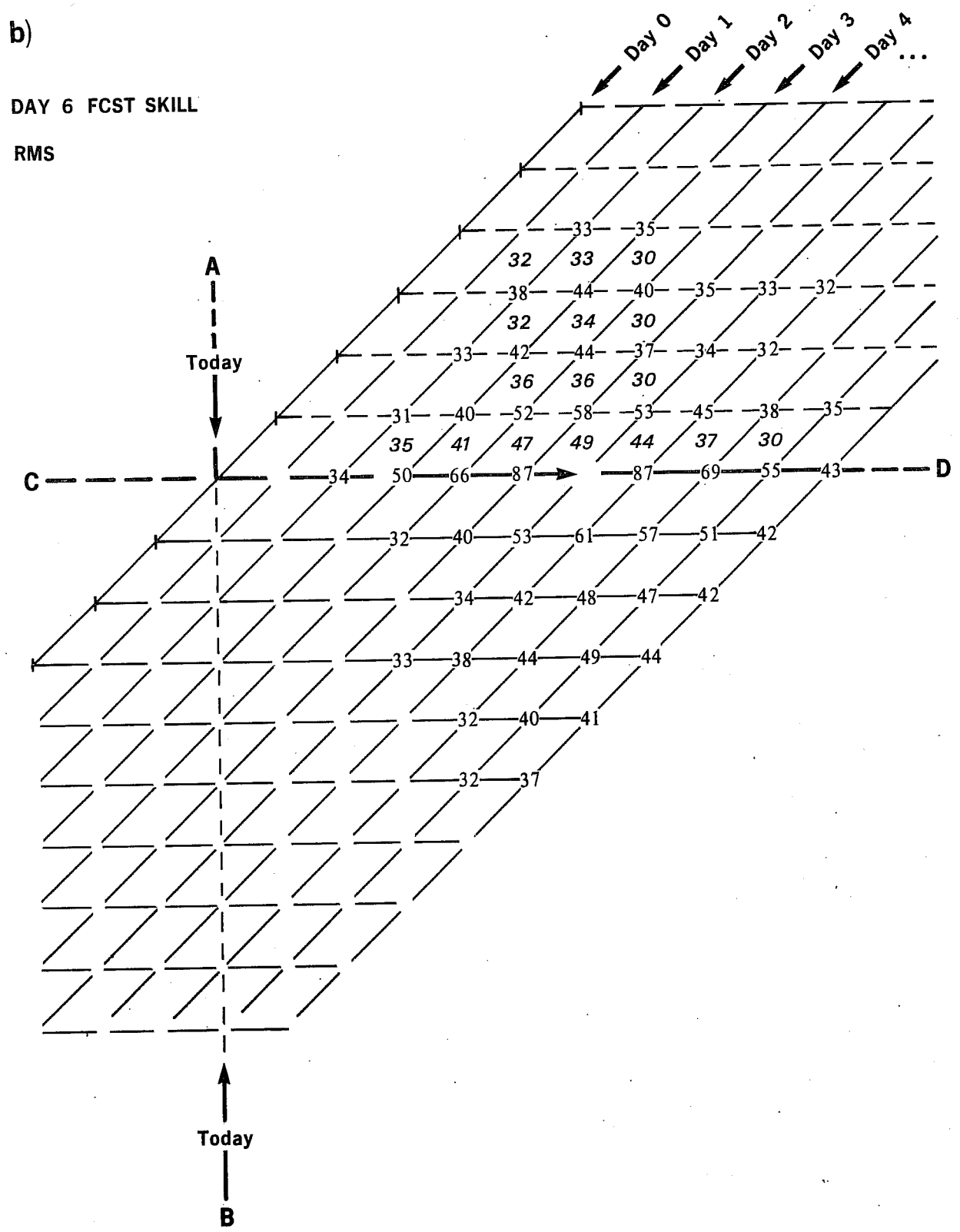


Fig.7 Continued

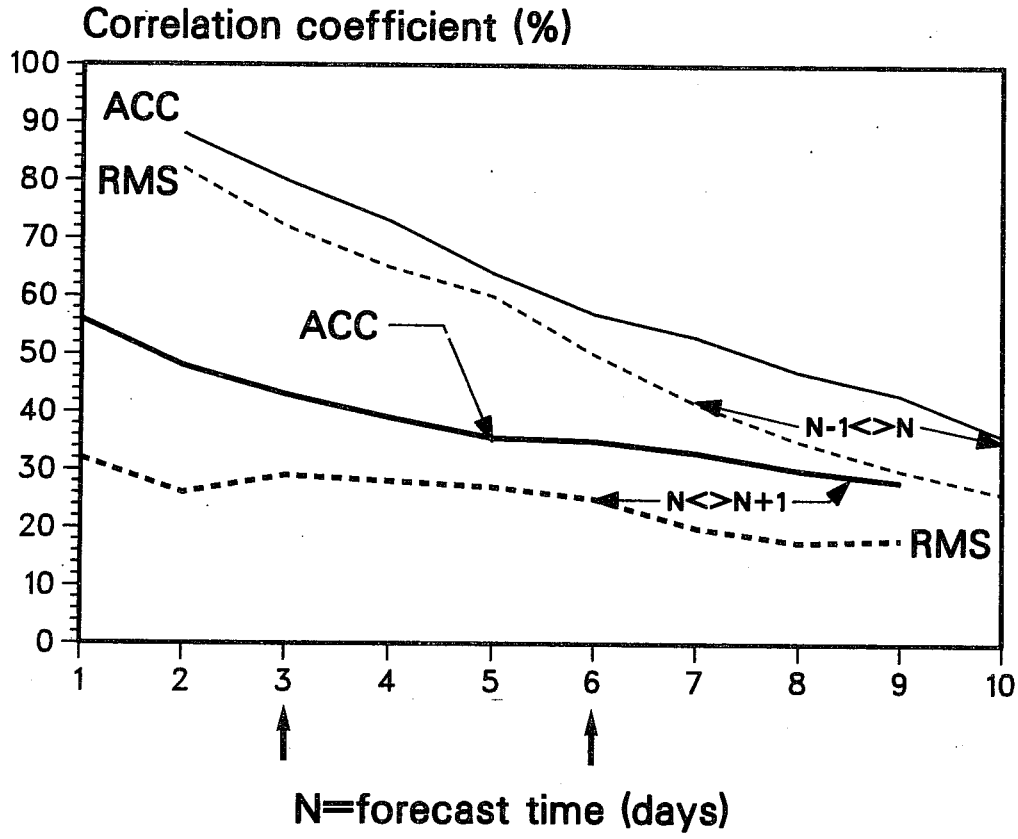


Fig.8 Correlation coefficient between the skill of a day N forecast and the spread between the day N and day N+1 forecasts (thick lines) and the day N-1 and day N forecasts (thin lines). Both skill and spread measures in ACC (full lines) or in RMS (dashed lines).

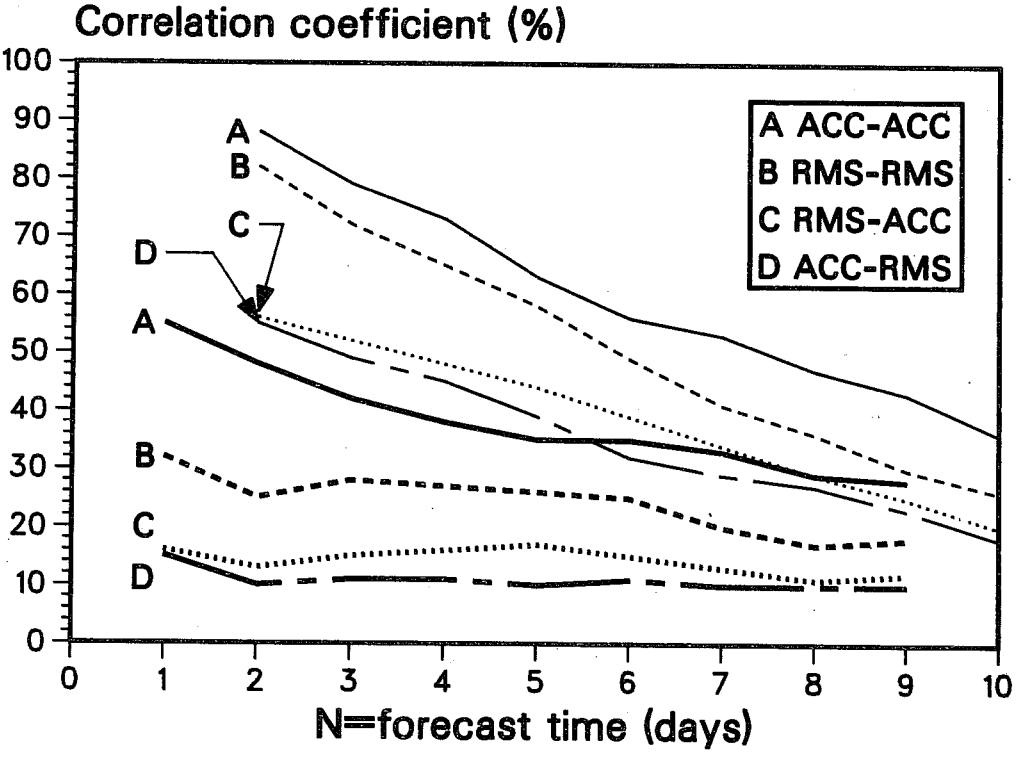


Fig.9 As Fig.8, but, in addition, skill in RMS and spread in ACC (dotted lines) and skill in ACC and spread in RMS (dash-dot lines).

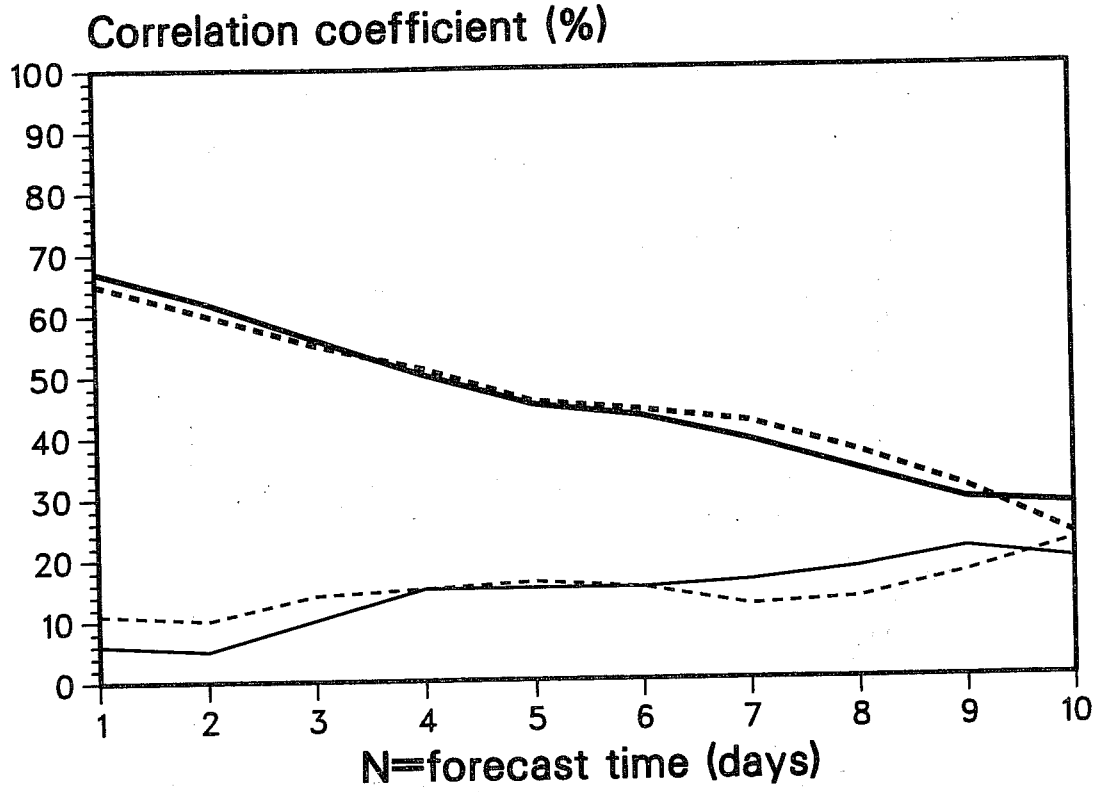


Fig.10 Correlation coefficient between the forecast skill (ACC, thick lines; RMS, thin lines) and RMS amplitude of the anomaly (observed anomaly, solid lines; forecast anomaly, dashed lines).

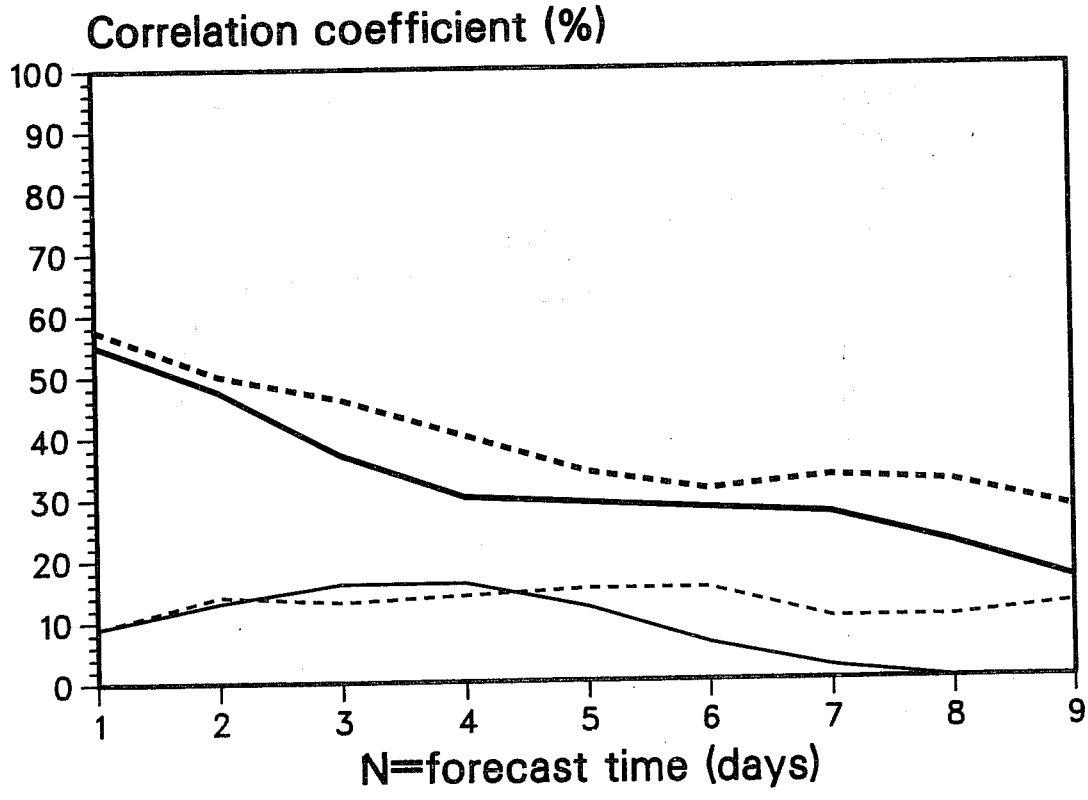


Fig.11 As Fig.10 but correlation between forecast spread (instead of skill) and RMS amplitude of the anomaly.

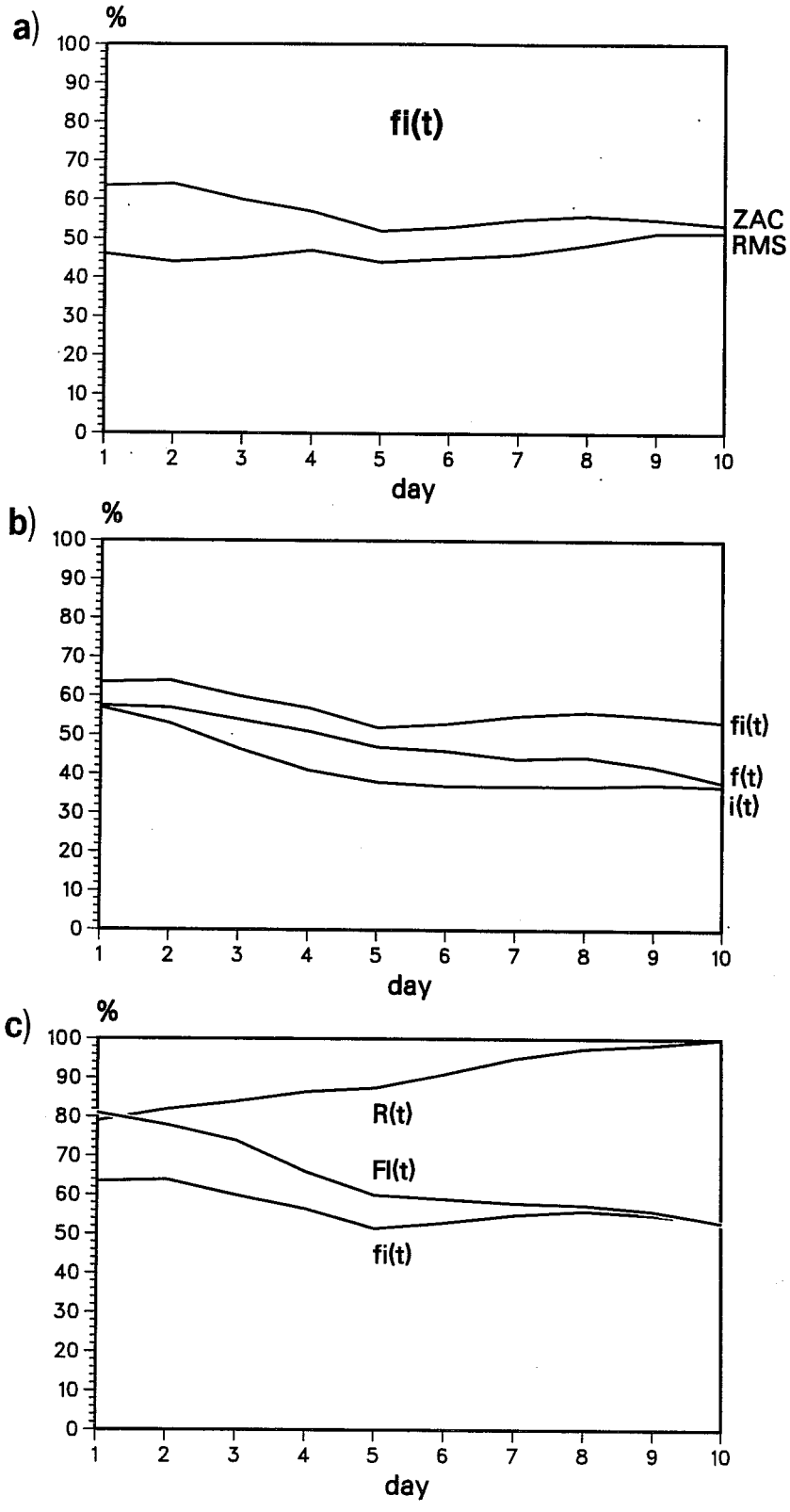


Fig.12 a) Correlation between regressed EOF coefficient, using both forecast and initial data, and hemispheric skill scores (ZAC and RMS) as a function of forecast time. b) Correlation between regressed EOF coefficient, and hemispheric ZAC scores, using forecast and initial data (fi(t)), forecast data only (f(t)), and initial data only (i(t)). c) A plot of the functions fi(t), $FI(t)= \sqrt{f^2(t)+i^2(t)}$, and the ratio R(t)= fi(t)/FI(t). See text for explanation.

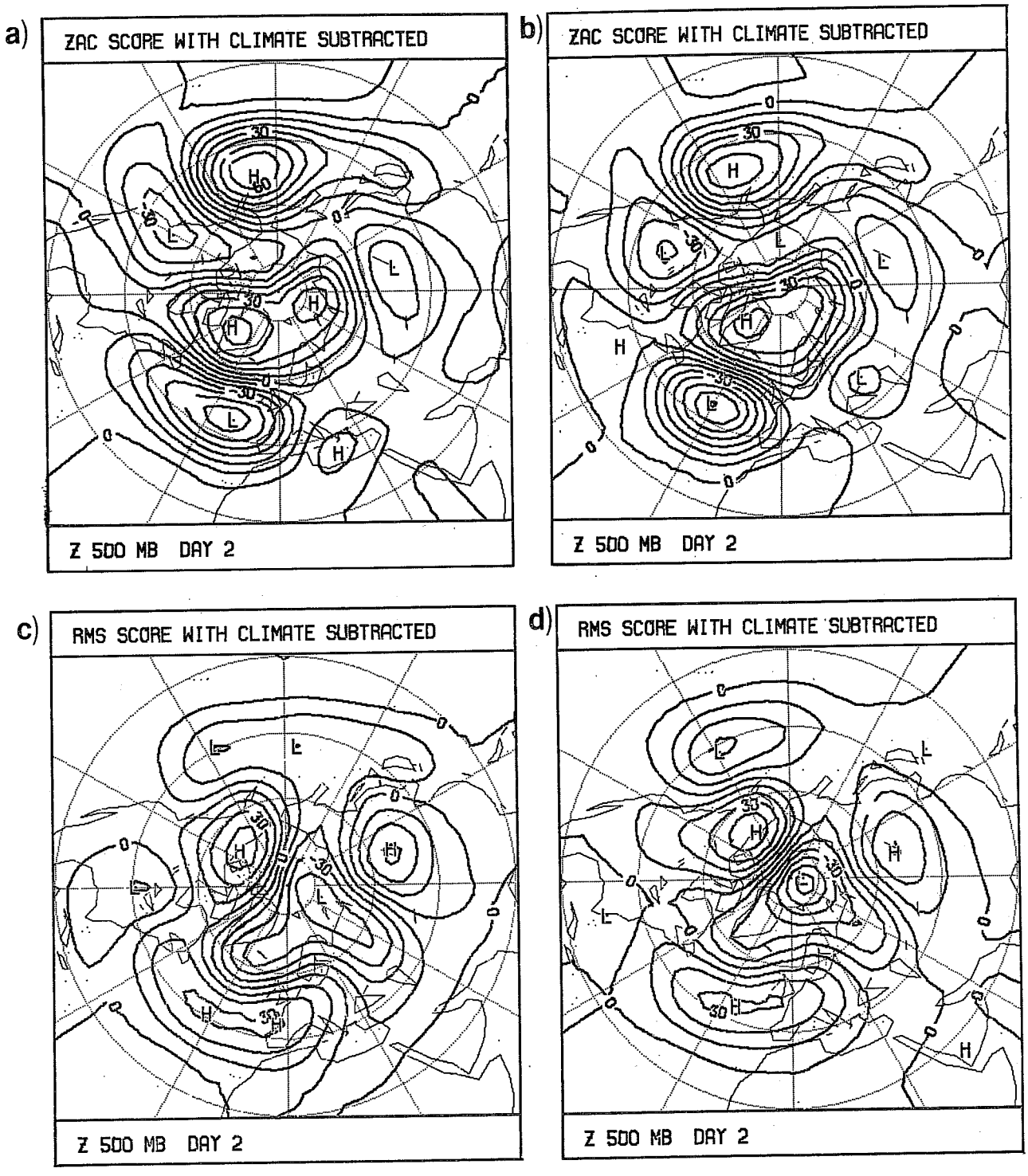


Fig.13 Pattern of 500mb height anomaly that correlates most strongly with hemispheric skill scores. a) day 2 forecast field and day 2 ZAC score b) initial data and day 2 ZAC score. c) day 2 forecast field and day 2 RMS score. d) initial data and day 2 RMS score.

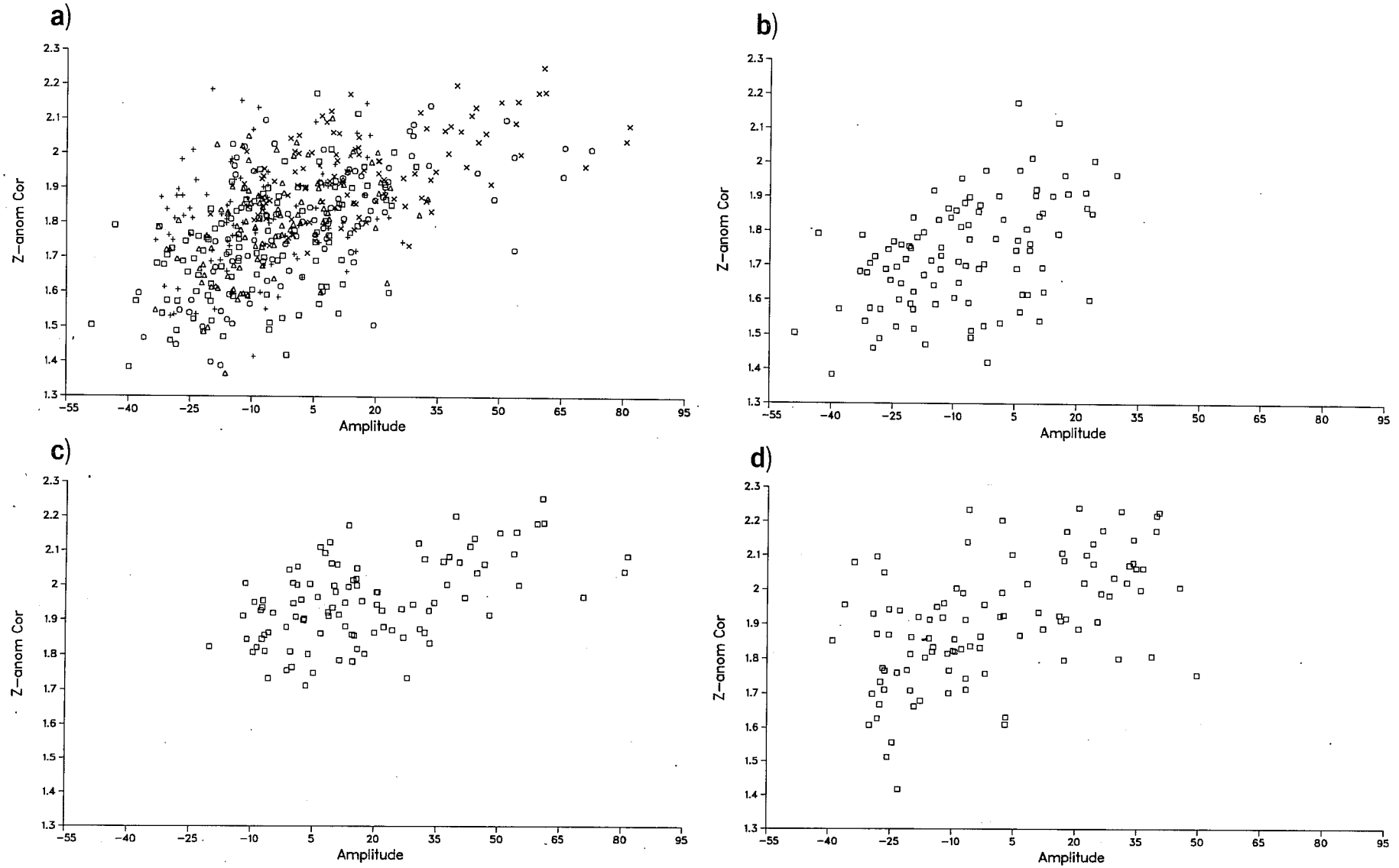
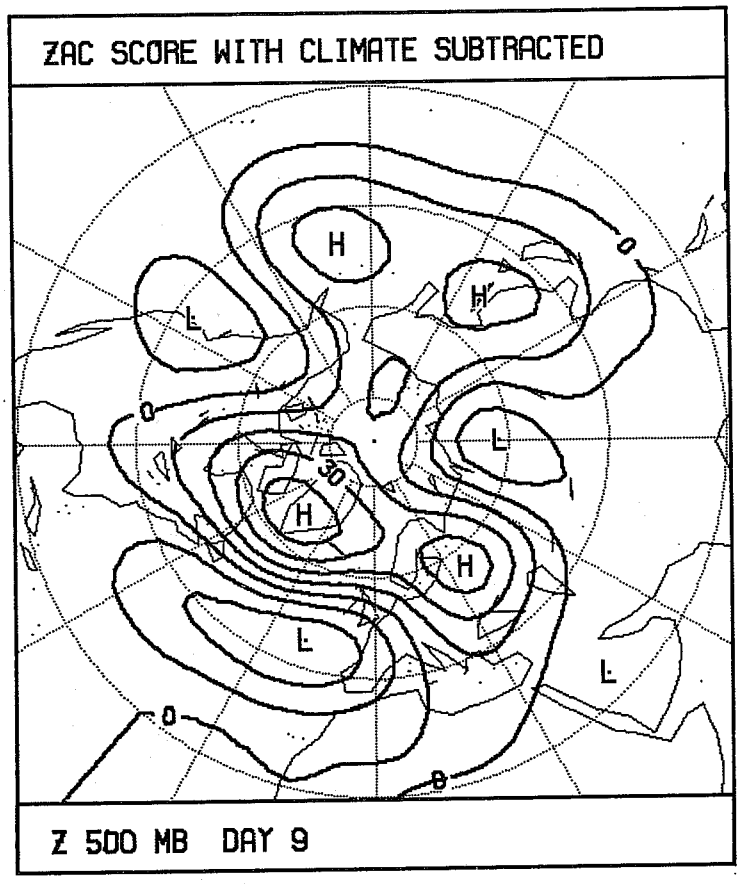


Fig.14 Scatter diagram between regressed EOF coefficient for day 2 forecast field, and hemispheric ZAC score. a) all 500 days. b) data for 1980/81 only. c) data for 1984/85 only. d) data from the independent 100 days 1985/86, using the regression weights from the 500 days concatenated dataset.

a)



b)

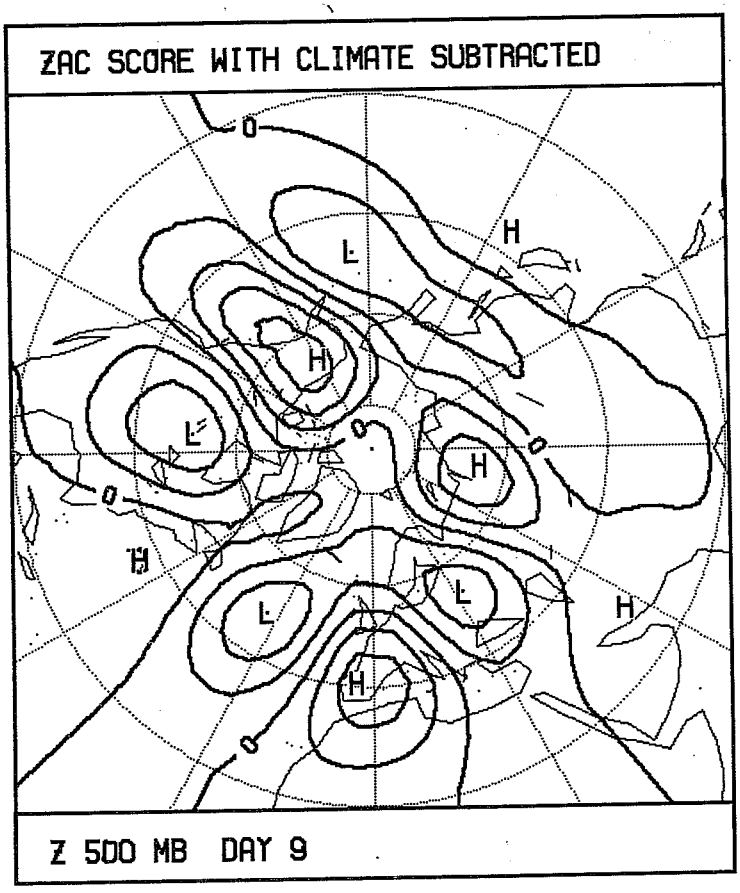


Fig.15 a) b). As Fig.13a), b) respectively, except for day 9 forecast fields.

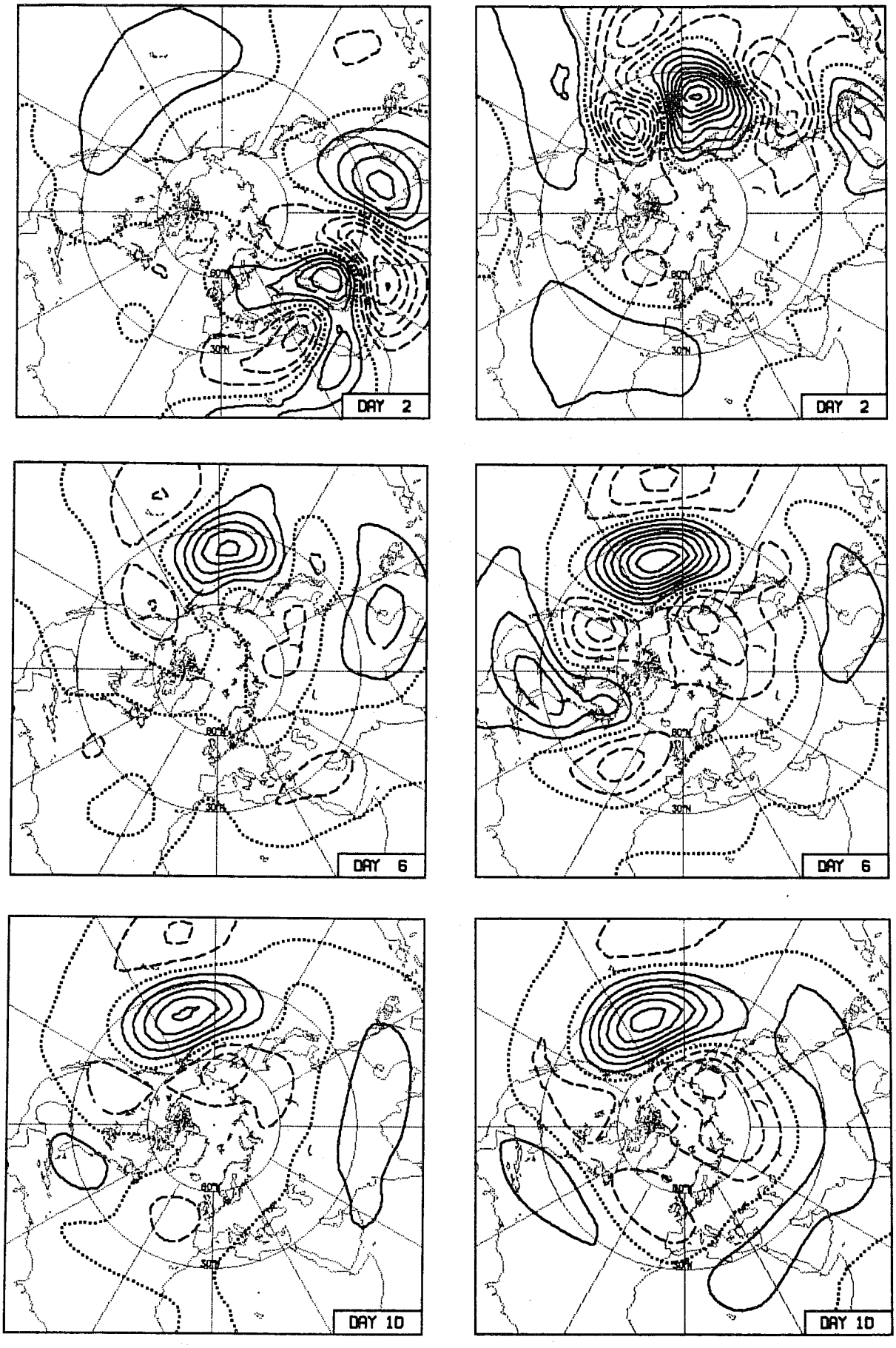


Fig.16 The perturbation streamfunction in a barotropic model at days 2,6,and 10 for an initial disturbance centred at 30N,0 (left hand side), and 30N,120E (right hand side). From Simmons et al (1983).

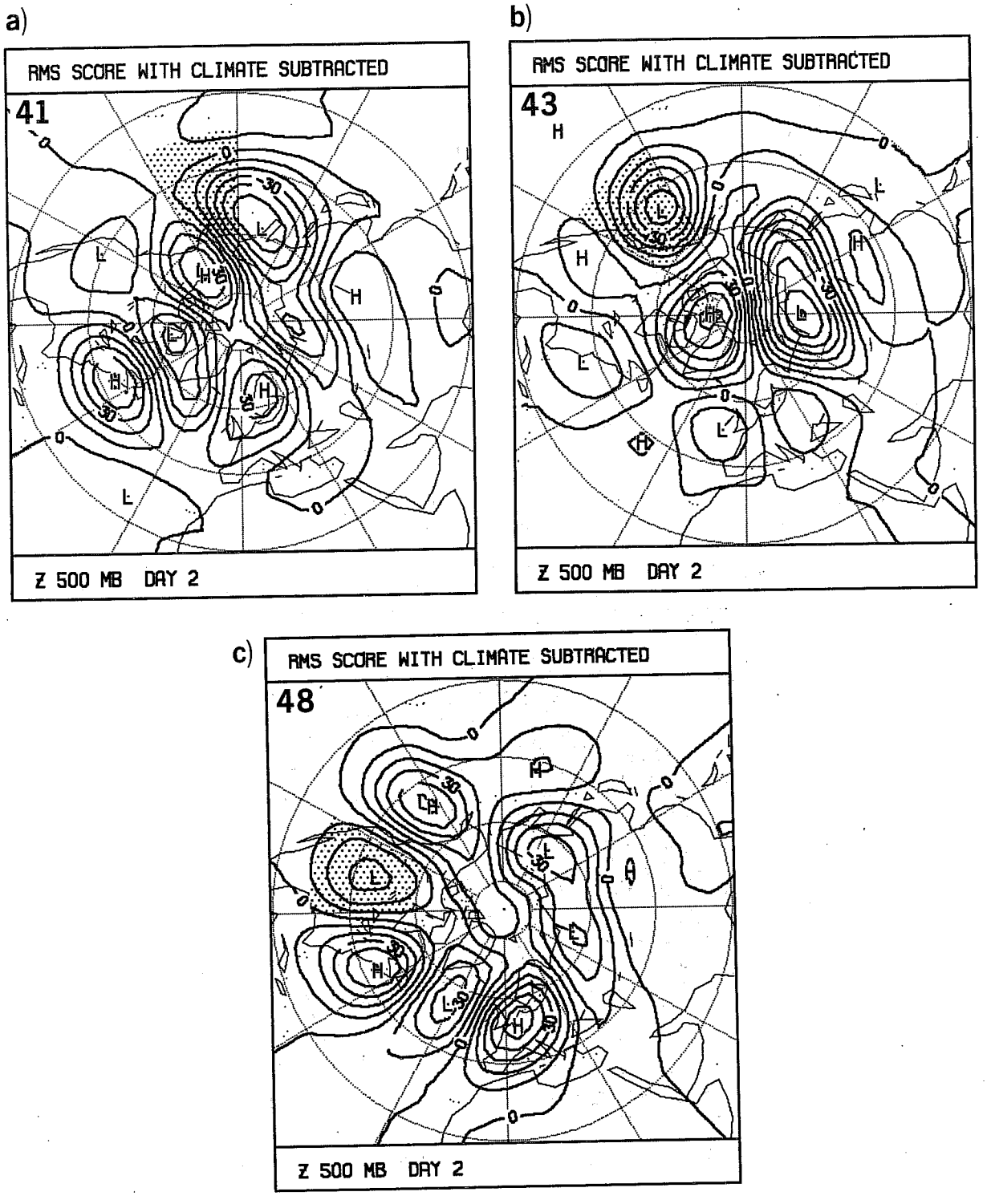


Fig.17 Pattern of 500mb height anomaly (m) for day 2 forecasts that most strongly correlates with RMS scores in the regional area 30N-60N and a) 180-150W, b)150W-120W, c)120W-90W.

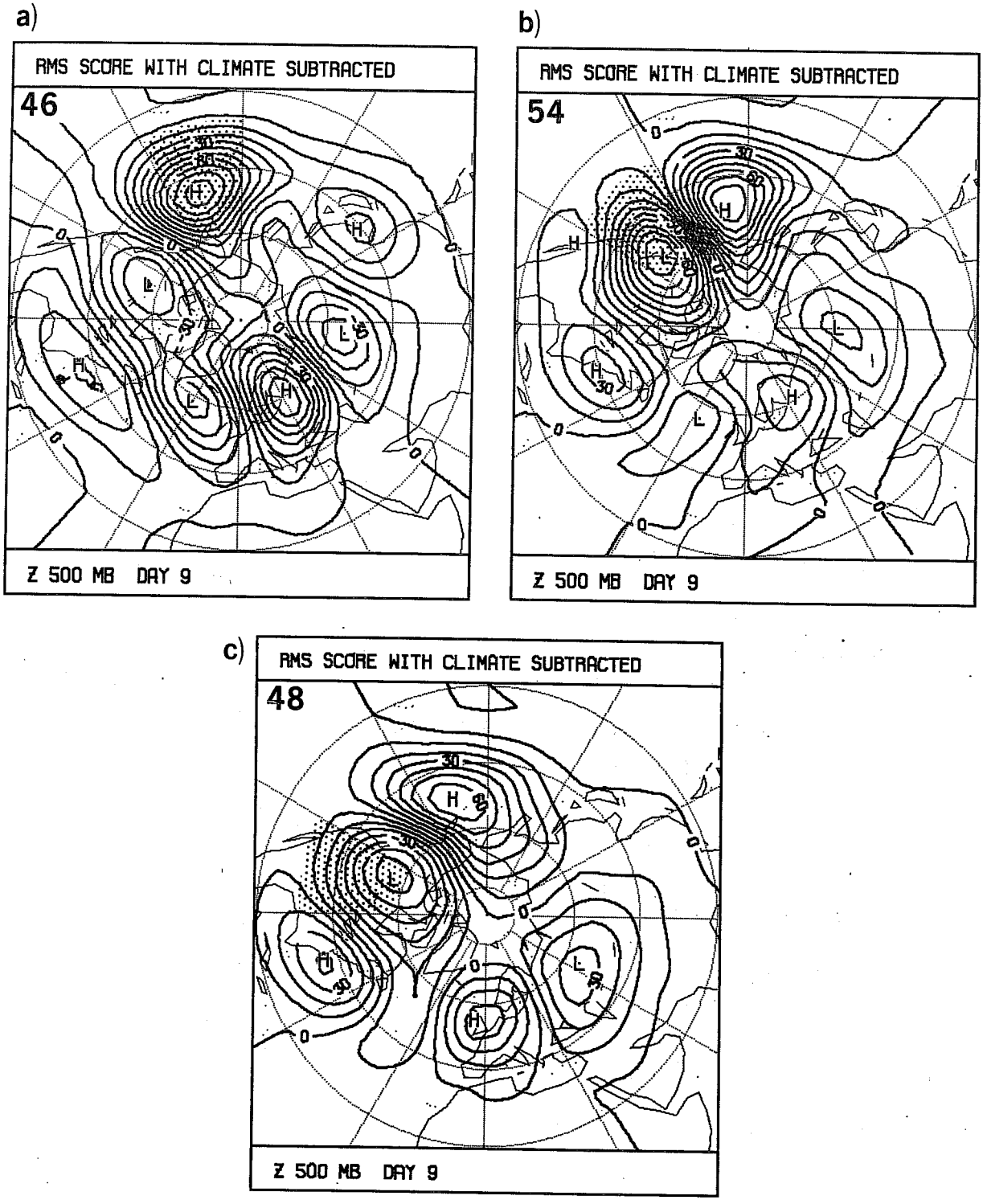


Fig.18 As Fig.17 but for day 9 forecasts.

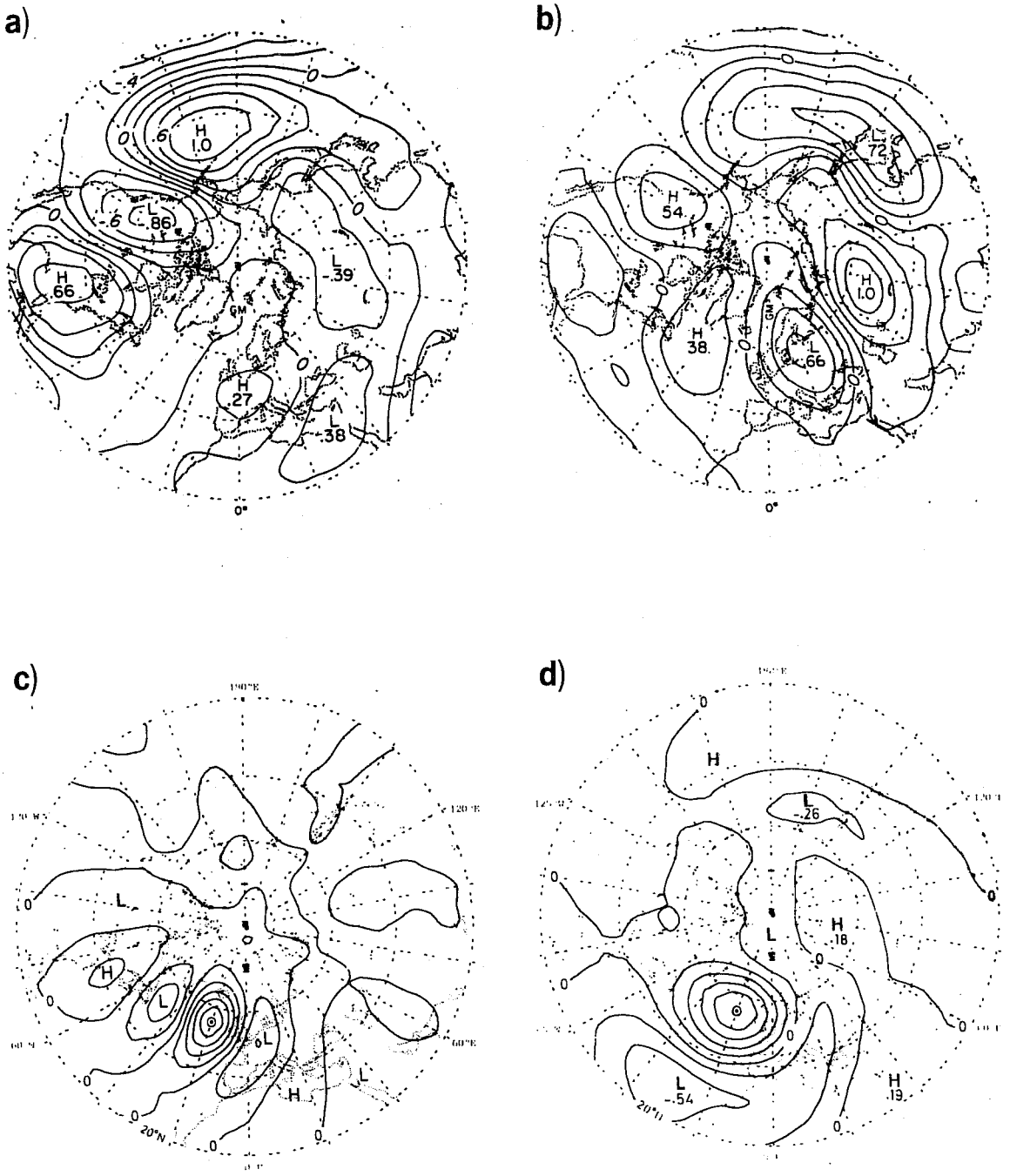


Fig.19 One point correlation maps showing the correlation coefficient between 500mb height at base points a) 45N,165W, and 55N,75E, and 500mb height at other gridpoints. From Wallace and Gutzler (1981), using monthly mean data. c) Base point at 55N,20W and data band-pass filtered. d) Base point at 55N,20W and data low-pass filtered. From Wallace and Blackmon (1983).

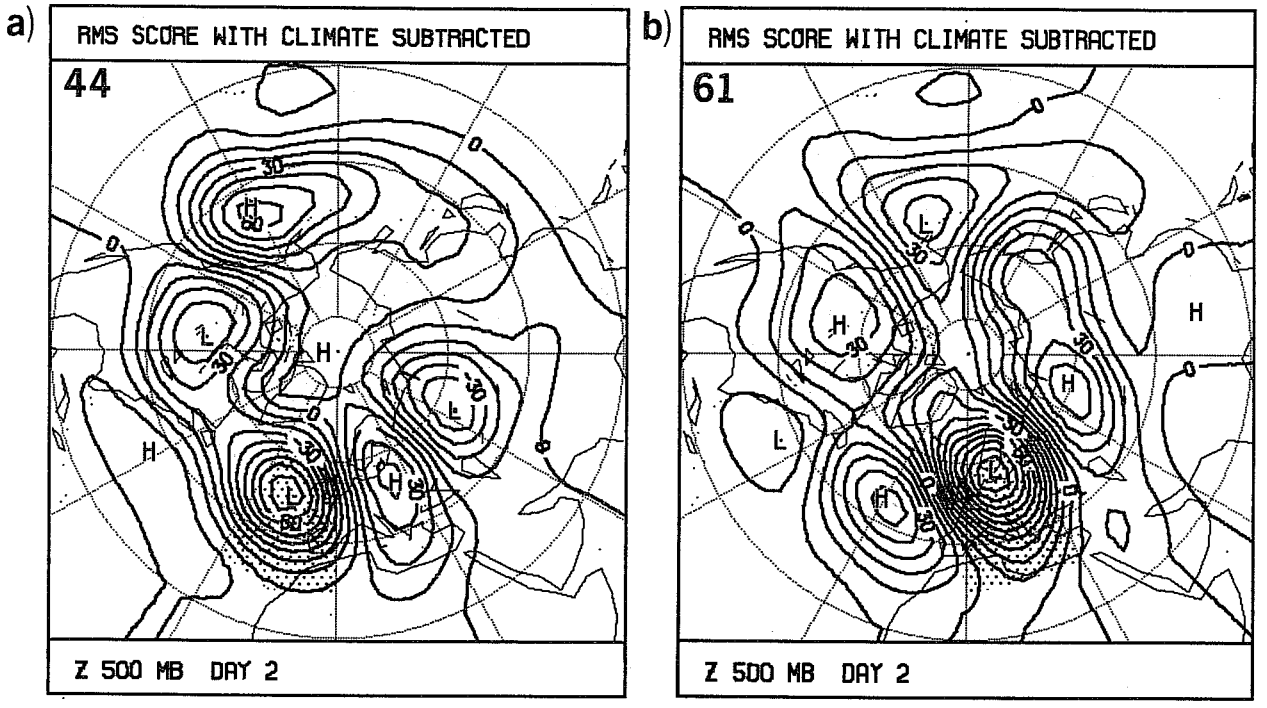


Fig.20 As Fig.17 but for regional area 30-60N and a)30W-0 , b) 0-30E.

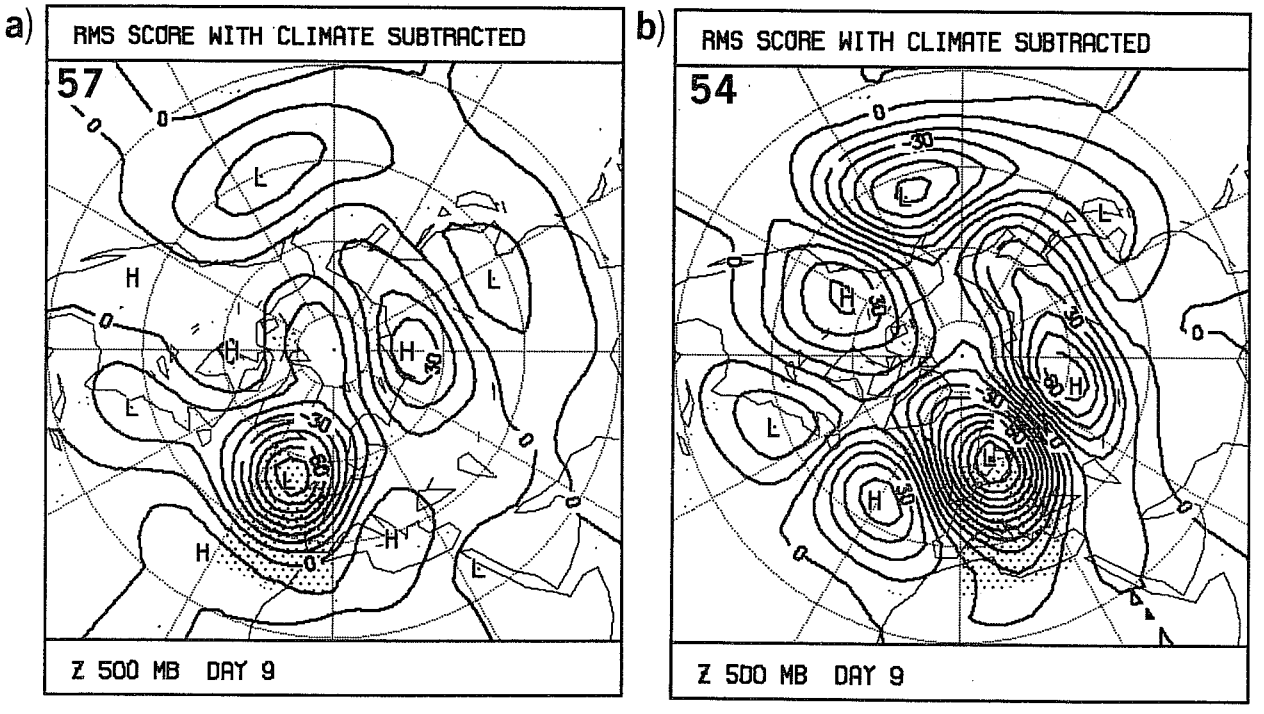


Fig.21 As Fig.18, but for regional area 30N-60N and a)30W-0 , b) 0-30E.