TECHNICAL MEMORANDUM

119

A pilot study on the prediction of medium range forecast quality

S. Grønås

Research Department

December 1985

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

1. INTRODUCTION

The predictability of the Centre's forecasting system has increased steadily since operational forecasting was introduced in 1980. Potential limits for the predictability, as worked out by Lorenz (1982), indicate that substantial improvements are still possible and that the overall forecast quality will continue to improve.

The quality of the medium range forecasts varies from day to day. Sometimes there is skill in the forecast up to ten days, occasionally only the short range products are useful. This spread in the predictability might become smaller as the general forecast quality increases. However, the day to day variation in quality is partly caused by natural variations of the predictability of the atmosphere and is likely to remain a problem for all future medium range forecasting. Assessment of the reliability of the forecasts will be one of the most important parts of the Centre's work. A priori forecasting of the forecast quality will hopefully be introduced at the Centre as more understanding of the problem is gained.

The users of medium range products, the forecasters, have to use all relevant knowledge when assessing the reliability of today's forecast. The regular publication of verification results is of great value, as are the studies of the performance and systematic errors of the Centre's forecasting system, i.e. Arpe and Klinker (1986), Akyildiz (1985).

There have been some studies directly related to the reliability of forecasts both within the Centre and the national services. Many forecasters seem to relate quality to consistency between the most recent forecast and those from previous days. Grönaas (1982) investigated spells with high and low predictability and found relations between forecast quality and the flow regimes. Ernst Klinker (personal communication) found significant relation between medium range forecast quality in Europe and synoptic scale activity in the Pacific.

In recent years some research on weather regimes and their stability and treatment by models has been undertaken. Simple models (e.g. Charney and DeVore (1979)), have shown that there is a possibility of more than one equilibrium regime in the atmosphere, and observational studies, Sutera (1985), seem to support this idea. Eventually such research will be applied to the prediction of forecast quality.

In this study the error evolution and the quality of spectrally truncated height fields are investigated. Such fields, truncated at total wavenumber 10 by a triangular truncation, are found to be valuable in medium range forecasting, Persson (1984). We assume that if the forecast system fails to predict the larger scales contained in the T10 flow, then the detail in the forecast is also useless. On the other hand, if the larger scales are well predicted in the medium range, then the truncated forecast is useful for weather prediction (see Persson).

For the truncated fields forecast quality is studied in relation to the following three factors: consistency with the previous forecasts, the flow type and influence of the Pacific region. The study could be considered as a pilot project to gain some experience before starting more extensive investigations of the variation of forecast quality.

2. DATA BASES FOR PREDICTABILITY STUDIES

This study uses only 500 hPa spectrally truncated fields; the truncation is at total wavenumber 10. The spectral fields are transformed to a 5 x 5 degrees latitude/longitude regular grid. Verification statistics are computed for a European area which covers 35 to 75° N and 25° W to 40° E. Four winter months November, December, January, February 1984/85 are investigated along with the four summer months May, June, July, August 1985. The anomalies are computed in relation to monthly mean ECMWF analyses for the period 1979-1984.

For more extensive studies, analyses and forecasts from all major NWP centres are available through the WMO/CAS NWP Intercomparison project. Over five years, forecasts of surface pressure and 500 hPa height up to five days have been collected from most centres, and up to 8 days from ECMWF and Japan. Details about the WMO/CAS data set is found in WMO technical document WMO/TD No. 60 (1985).

Predictability studies should be encouraged. At the Centre a suitable subset of the archived data could easily be provided together with software for unpacking, handling spectral coefficients, verification routines, etc. Five years of data are now available.

3. ERROR EVOLUTION

To study the error evolution the 500 hPa T10 analyses and forecasts up to ten days are plotted together with the errors for the winter sample 1984/85 and the summer sample from 1985 for the northern hemisphere north of 25°N. The maps are plotted so that verifying analyses and lagged forecasts are vertically underneath each other. Also plotted are the anomaly correlation coefficients, as a quality measure for Europe, and a parameter which measures the consistency to the previous lagged forecasts. The consistency correlation is simply the correlation coefficient between the anomaly of the lagged forecast, for instance between anomalies from today's 4 day forecast and the 5 day forecast from the previous day. The maps are studied to provide some descriptive model for the error evolution. This seems to be a difficult task, and more work should be done in this area.

Some interesting case studies have been carried out which concern the evolution of errors into the medium range caused by errors in the analysis. Cats and Åkesson (1983) studied the divergence of medium range forecasts made by analyses separated by one day. They found that small differences in the analyses in a baroclinic region resulted in large downstream differences in the medium range forecast. A similar amplification of small pertubations in a complex, unstable flow was found by Hollingsworth et al. (1985). They found the development theory of Simmons and Hoskins (1979) very useful in documenting these effects, and suggested that this theory describes one of the essential mechanisms for the loss of forecast skill through unstable amplification of analysis errors.

In this study, only a few brief statements on the error evolution will be given. In the 500 hPa T10 forecast the first two or three days is the initial stage of the error development. During this period isolated quasi-circular error patterns with amplitudes of a few decametres develop. The location of these errors might be anywhere, but errors connected with the major jets are quite frequent. The scale of the errors is relatively small.

After a further two or three days the forecasts are characterised by the development of error trains from the isolated areas of errors. Some of the errors might be interpreted as a pertubation of a baroclinic flow. From these pertubations trains of errors, develop downstream and also upstream of the disturbance.

This behaviour is simpler to that found in the case studies mentioned above and the evolution seem to be in accordance with the downstream development theory of Simmons and Hoskins.

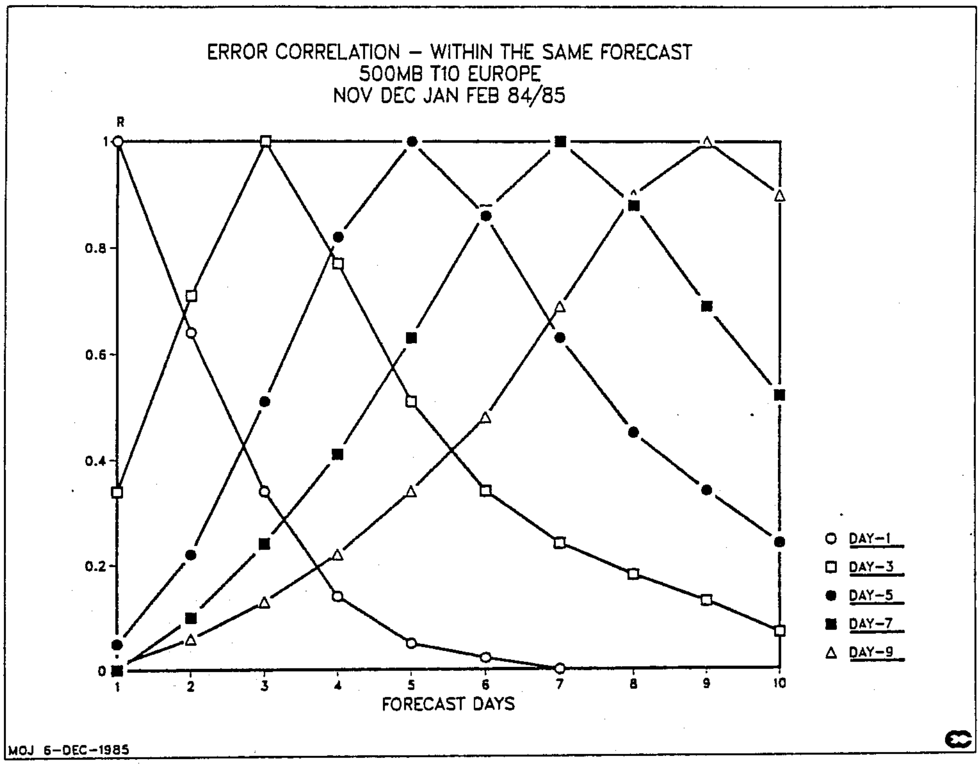
After 4-5 days such error trains - with amplitudes of more than 10 decametres - are found in nearly all forecasts. However, their direction and the apparent displacement is not necessarily along the 500 hPa flow, and the trains might be caused by errors other than those in the analysis.

Within the train development stage the scales of the errors seem to increase, and as error patterns become well developed, they will also be more stationary. The latest part of the forecast is often characterised by slow moving patterns, and the amplitudes seem to increase steadily during this stage.

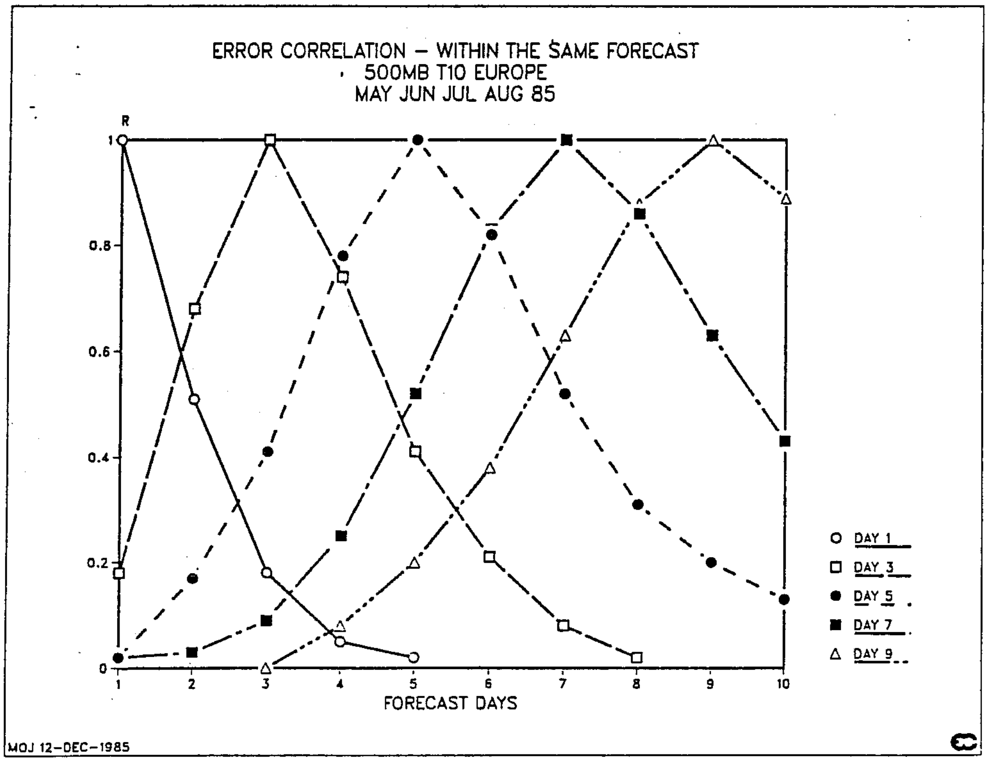
Within a western European area, negative error structures are often seen after 5-6 days. These errors seem to develop in different ways.

Error correlation between the forecasts, both within the forecast and with previous forecasts, are computed for Europe for winter and summer. Figs. 3.1a and 3.1b show mean correlations within the forecast for the two periods. The correlations seem to support the experience from the error maps described above. The errors at day 1 are hardly correlated with the errors further out in the forecast. This indicates that the errors are moving and developing. The correlation between the errors increase all the time and the correlation with the next forecast step is always higher than that to the previous step. The most rapid increase is found up to day 5, indicating that the errors are growing. After day 5 the scale of the errors seems to have increased and the patterns might be more stationary. As expected, all error correlations are higher in winter than in summer.

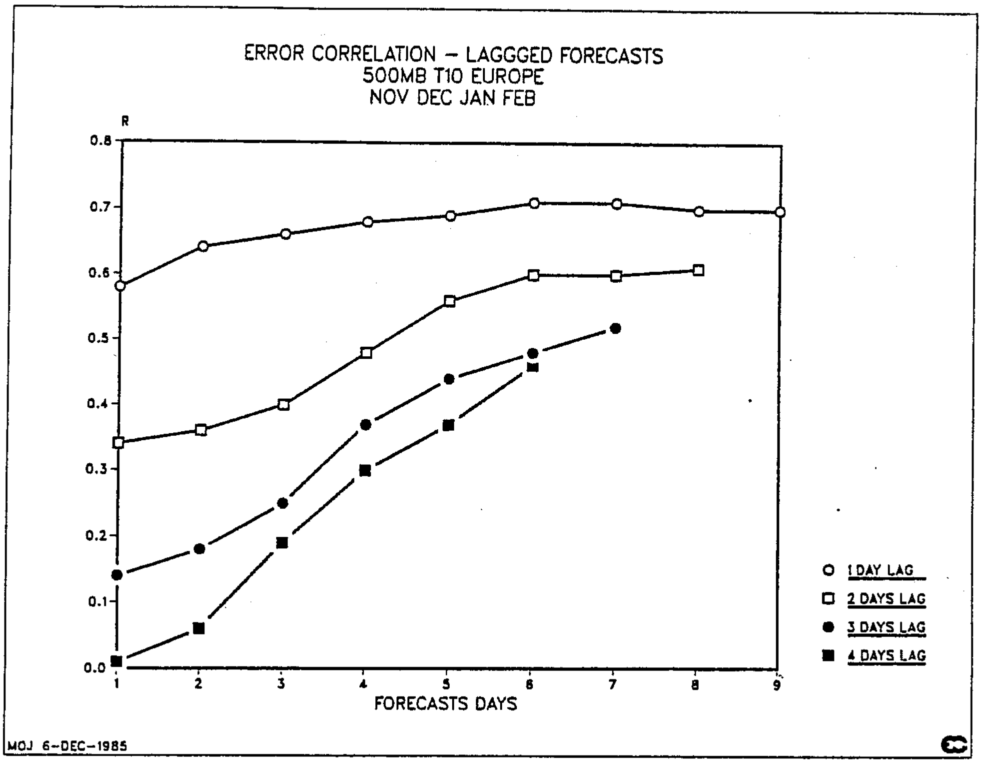
The errors of today's forecast are, as expected, correlated with the errors in the forecast from the previous days. Error correlations of lagged forecasts are shown for Europe in Figs. 3.2a and 3.2b for winter and summer. For a lag of one day the correlations are significant from day 1 onwards and increase in winter to 0.7 after day 4 and stay steady at this value thereafter. For larger lags, the error correlation is much lower, especially for the first days of the forecast.



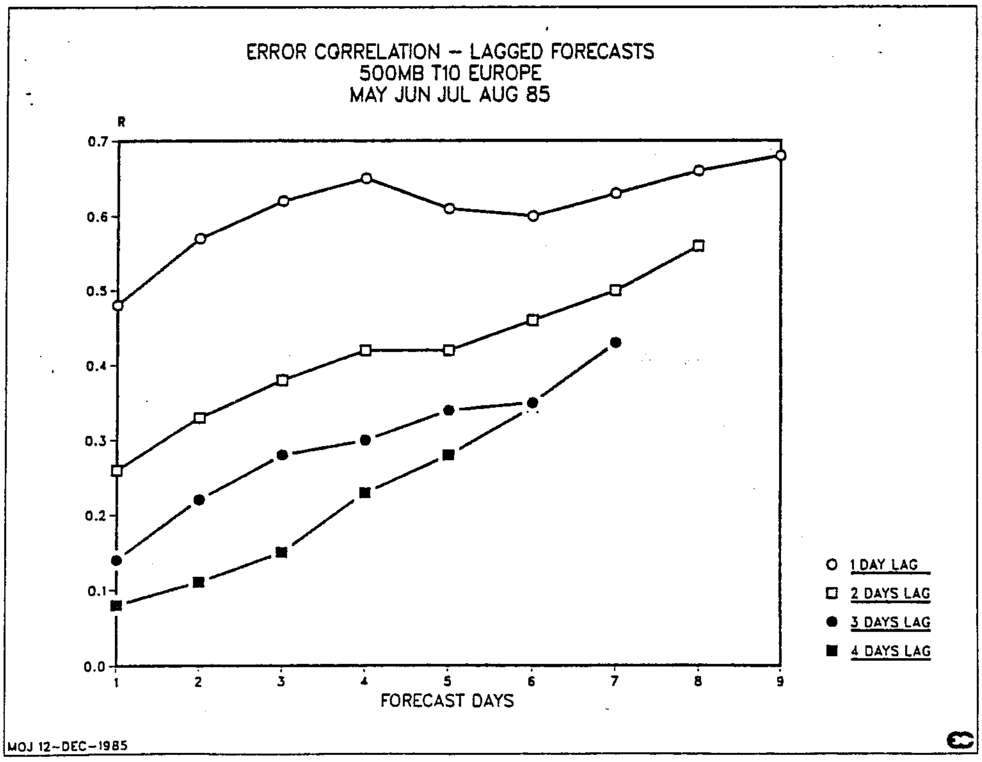
3.1a Correlation of forecast errors of Days 1, 3, 5, 7, 9 with the adjacent forecast days for the period November 1984 to February 1985, 500 hPa height fields truncated at wavenumber 10.



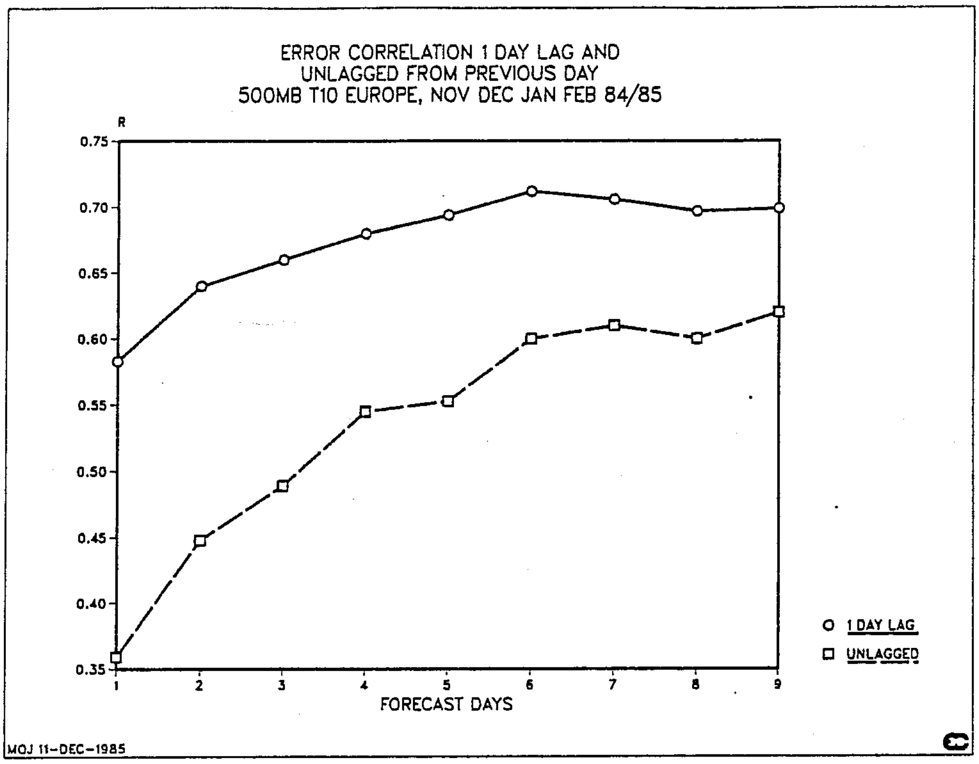
3.1b As 3.1a, but for period May to August 1985.



3.2a Correlation of forecast errors between neighbouring forecast runs lagged by 1, 2, 3 or 4 days for the period November 1984 to February 1985, 500 hPa height fields truncated at wavenumber 10.



3.2b As 3.2a, but for period May to August 1985.



3.3 As 3.2a, comparing the correlation between the height errors of two subsequent forecast runs with and without taking a one day lag into account.

In Fig. 3.3 the error correlation for a lag of one day is shown together with the unlagged correlation, which is the correlation between today's errors and the errors at the same forecast step the previous day. The lagged error correlation is significantly higher than the unlagged one is throughout the forecast. This indicates that the errors are highly dependent on the flow.

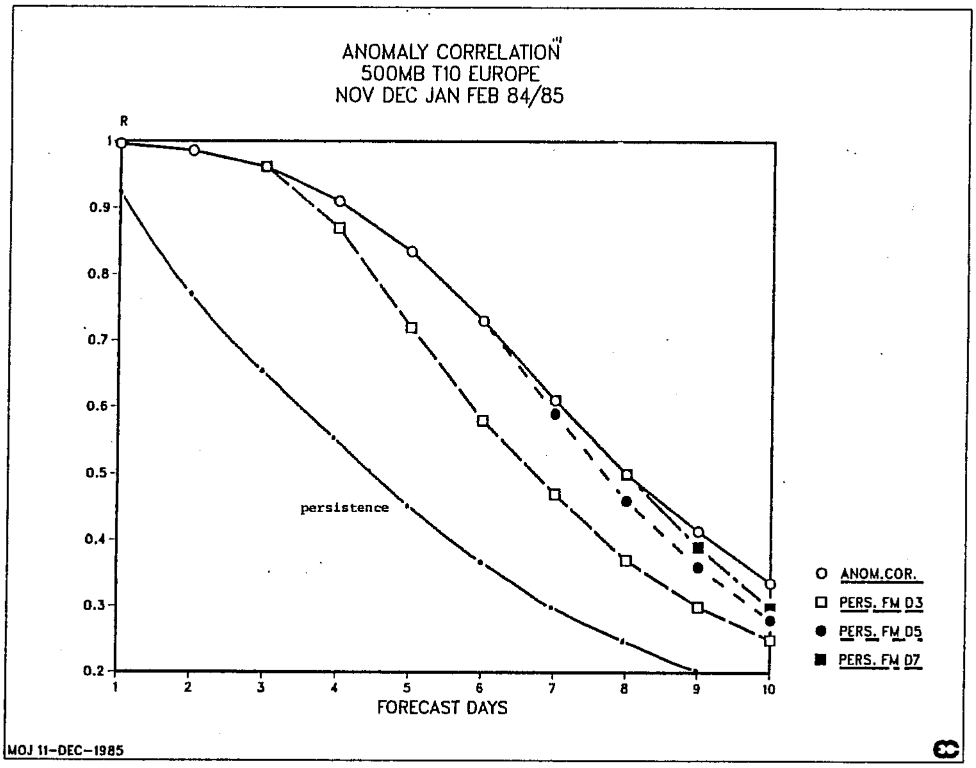
4. FORECAST QUALITY

Anomaly correlation coefficients of the model forecast for the European area are shown for winter 1984/85 in Fig. 4.1. The forecast quality stays high throughout the first 4-5 days of the forecast and drops off rapidly between day 5 and day 7. From day 8 to day 10 the decay in the anomaly correlation is again somewhat smaller. This is in accordance with other verification results.

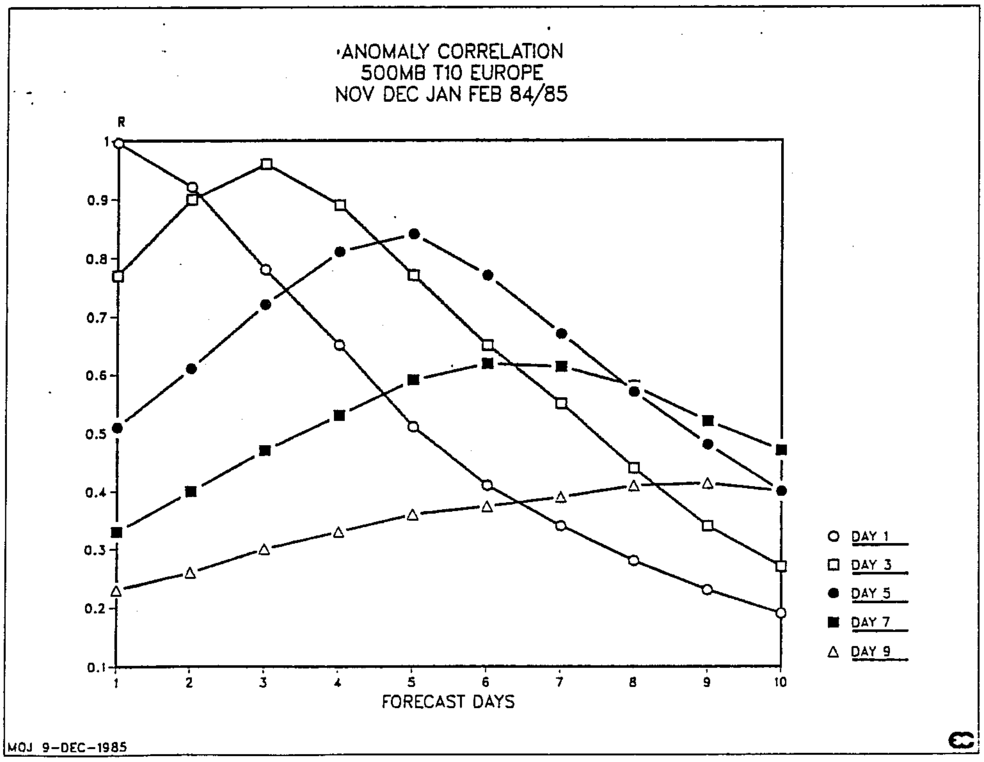
Numerical products are usually considered to be useful as long as the anomaly correlation is above 0.6. When this measure is used, the predictability of the forecasts is more than 7 days in winter. When persistence is used to predict three days forward, the score will also be approximately 0.6, that is about the same as the score after 7 days with the model forecasts.

Fig. 4.1 also shows the persistence scores using the day 3, day 5 and day 7 forecasts for the rest of the forecast period; this will be referred to as the day n persistence. For the winter sample the day 5 and day 7 persistence scores are as high as those for the forecast one day forward. However, further on in the forecast the quality of the model prediction is higher than persistence. This means that the model adds some information to the forecast even at the latest days of the integrations.

Verification of the first two years of ECMWF operational forecasts by Grönaas (1982) showed pronounced phase errors in the forecast after a few days. Fig. 4.2 shows forecast quality, in terms of anomaly correlations, displayed in a special way for days 1, 3, 5, 7 and 9. For each forecast day all the ten forecast steps are offered as predictors. The curves show that the forecast from the correct day gives the best prediction, and that this is the case even at day 9, although the differences to day 7 and 8 are very small. This means that our verification results now show no sign of model phase errors. Fig. 4.2 also indicates that for all steps in the forecast, there is a better correlation to



4.1 Anomaly correlation of 500 hPa height forecasts for Europe truncated at wavenumber 10 during November 1984 to February 1985. See text for further explanation.



4.2 Anomaly correlation of 500 hPa height forecasts for Europe truncated at wavenumber 10 during November 1984 to February 1985. For days 1, 3, 5, 7 and 9 of the forecast the correlation with each of the ten verifying days of the forecast range is shown.

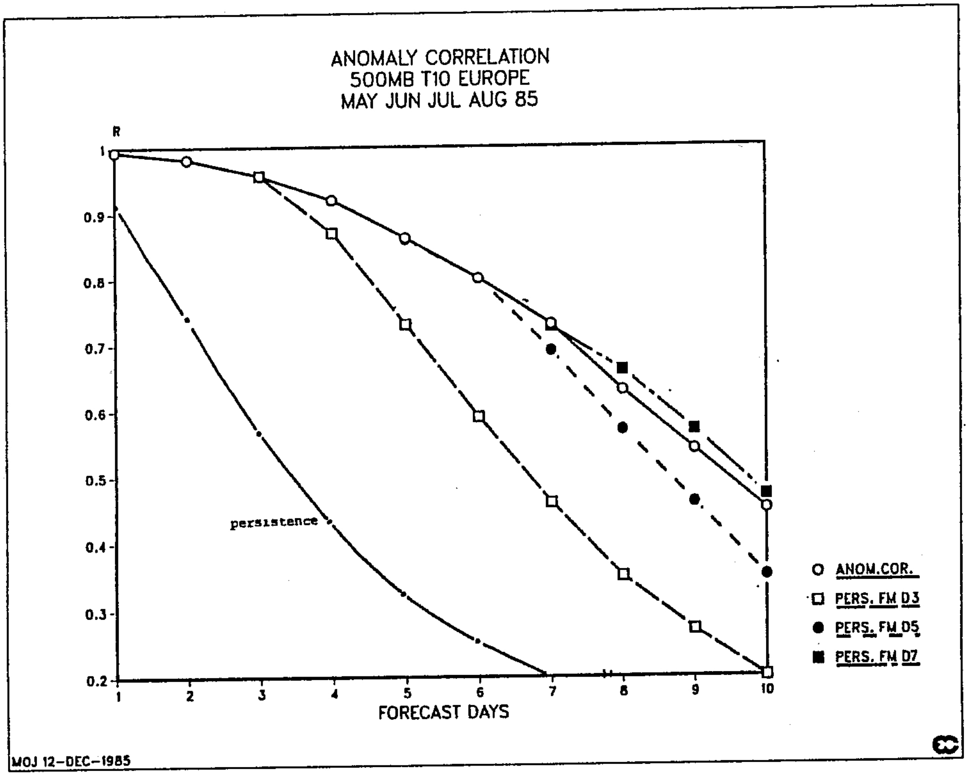
the previous forecast step than to the following one. This is as expected since the the forecast quality decreases throughout the forecast. Arpe and Klinker (1986) also confirm the improvement in the model phase speed which they link to the introduction of the spectral model.

The verification results from the summer sample are found in Figs. 4.3 and 4.4. Surprisingly the drop in forecast quality after day 4 is less than in the winter sample and the predictability is more than 8 days. Most forecast verification show better results in winter compared to summer and the high values are difficult to explain. Although the anomaly correlation is high, the model does not seem to add information to the forecast after day 7, since the 7 day persistence gives better scores than the model. This is also indicated in Fig. 4.4, which shows that the day 6 forecast gives better score for day 7 than the day 7 forecast itself; further, the day 7 forecast is the best predictor for the day 9 forecast. The high absolute score is probably of modest value for practical forecasting, since the T10 forecast has much less energy in summer and is less useful for forecasting purposes.

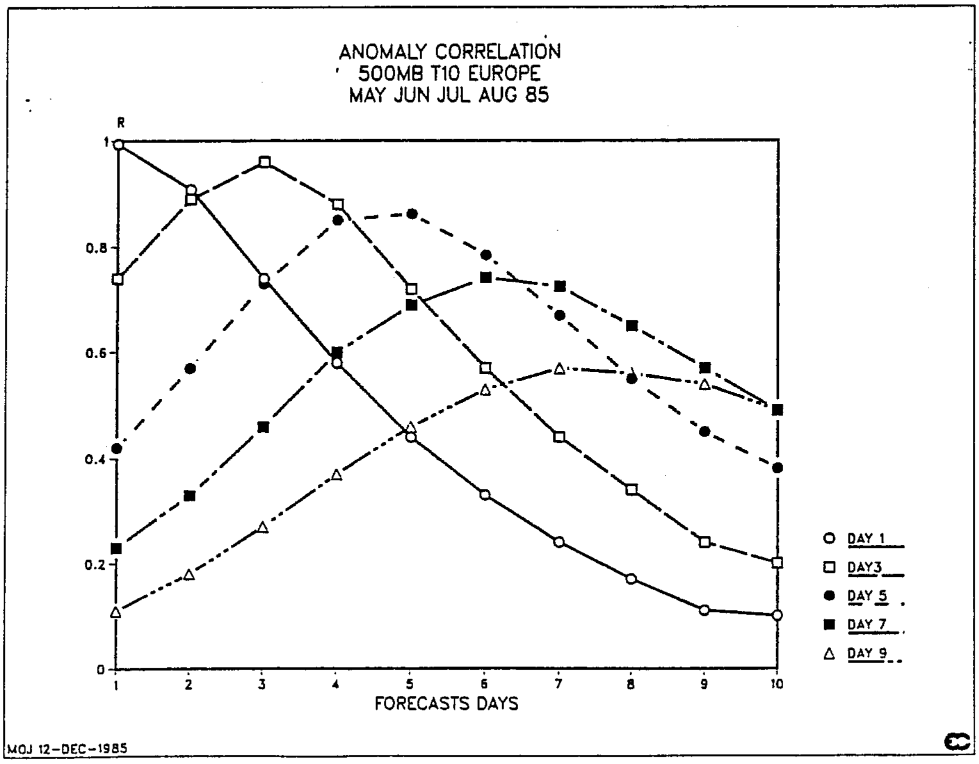
Neither of the summer verification results show any sign of phase errors.

The anomaly correlation for our sample varies considerably from day to day. As for the general verification results, there are spells with high and low predictability. In Fig. 4.5 is plotted the correlation coefficient between today's anomaly correlation and the anomaly correlation for lagged forecasts from the previous days. For one day lag the correlation is above 0.7 for the first three days into the forecast and around 0.6 later on. Similar results, but with lower values, are found for a two days lag. For a greater lag than this, all correlations are small, but especially in the first days of the forecast. The result shows that the intercorrelation of forecast quality between consecutive days is too weak to use persistence of forecast quality as a predictor of the quality for medium range forecasts.

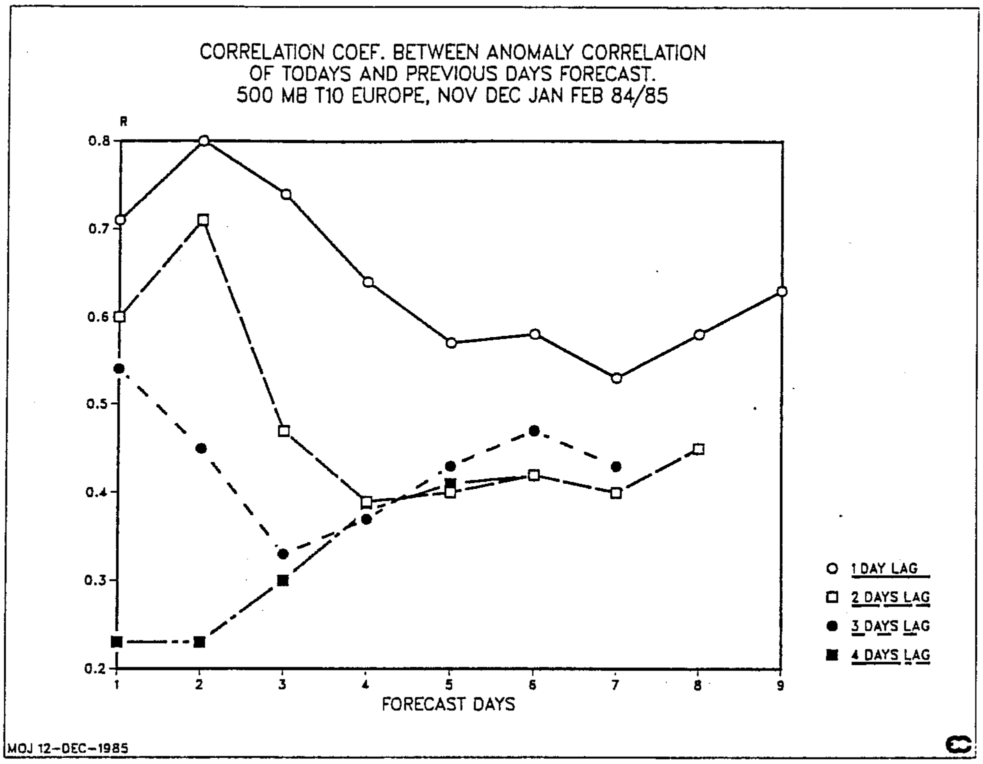
For both winter and summer samples the anomaly consistency correlation to the previous forecasts is computed. Figs. 4.6 and 4.7 show how the consistency decreases throughout the forecast. If we let a correlation of 0.7 be a lower limit for consistency, we find consistency up to 7-8 days to the previous day



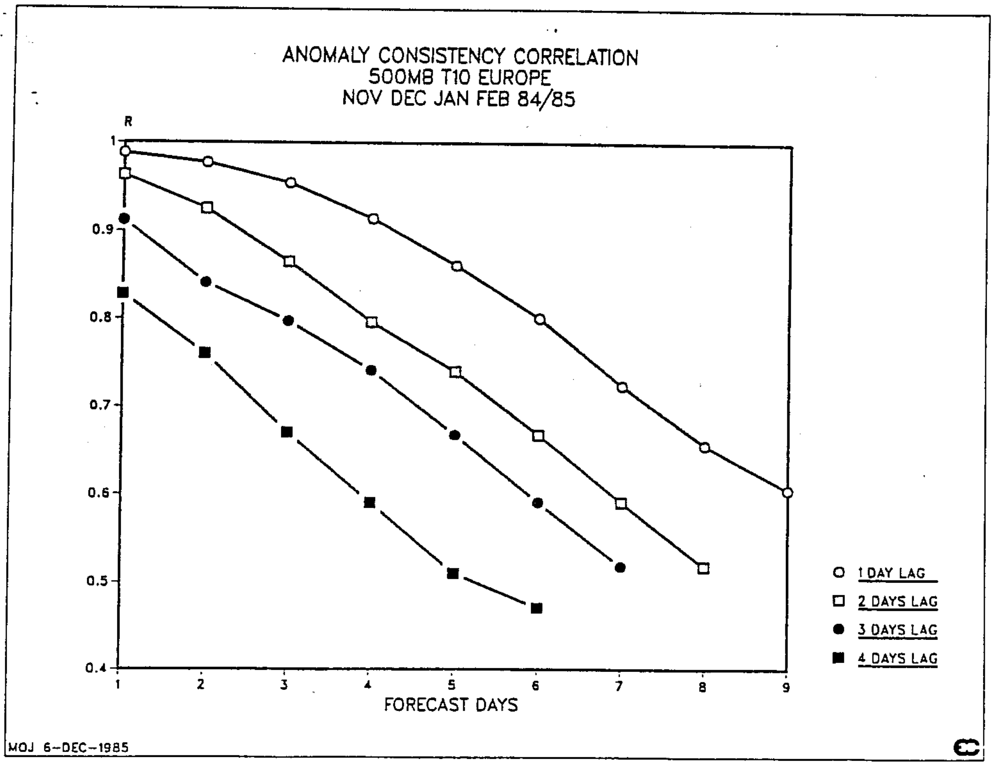
4.3 As Fig. 4.1, for the period May to August 1985.



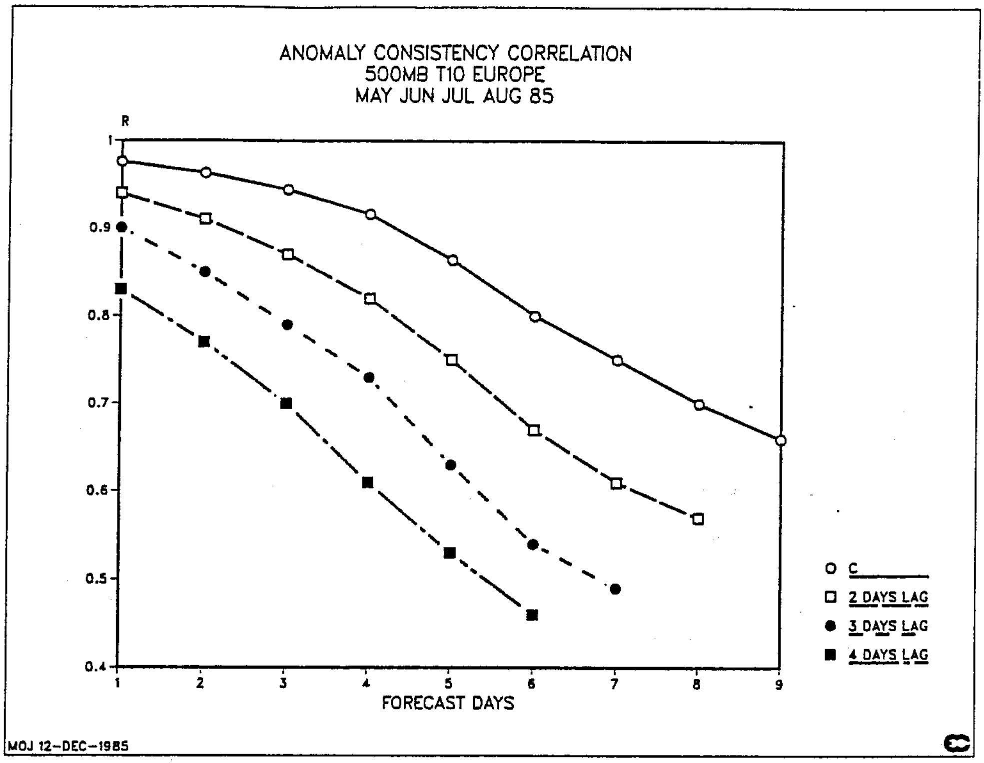
4.4 As Fig. 4.2, for the period May to August 1985.



4.5 Correlation coefficient between the anomaly correlation of neighbouring forecasts lagged by 1, 2, 3 and 4 days 500 hPa height fields truncated at wavenumber 10 over Europe for November 1984 to February 1985.



4.6 Anomaly consistency correlation between neighbouring forecasts lagged by 1, 2, 3 and 4 days, 500 hPa height fields truncated at wavenumber 10 over Europe November 1984 to February 1985.



4.7 As Fig. 4.6, for period May to August 1985.

in the winter sample, and 5-6 days to the forecast two days earlier. For forecast runs which are four days apart, consistency is lost after three days into the forecast. For our sample there are rather small differences in consistency between winter and summer.

Consistency statistics could be measured on a monthly basis along with other verification results. In this way we could know how the consistency varies with forecast quality.

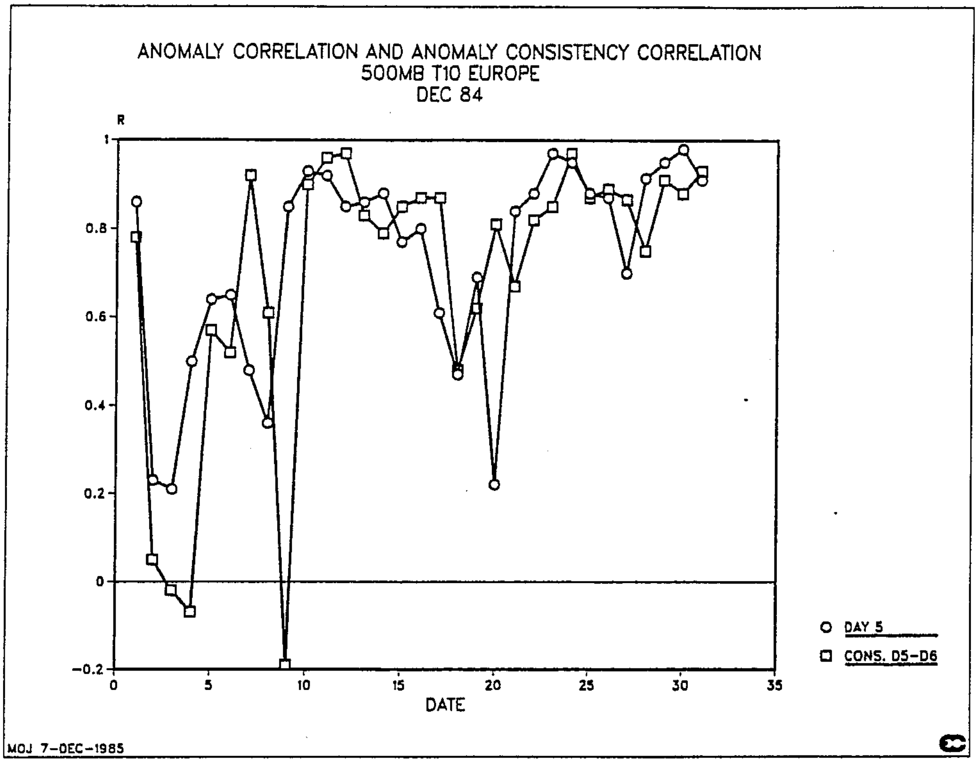
5. FORECAST QUALITY RELATED TO CONSISTENCY AND FLOW PATTERN

Forecasters will often inspect the consistency of a forecast by comparing it with those from previous days and equate consistency with forecast quality: if the consistency is high, then the forecast quality is supposed to be high as well. However, consistency might very well just indicate that the errors of the forecasts are intercorrelated.

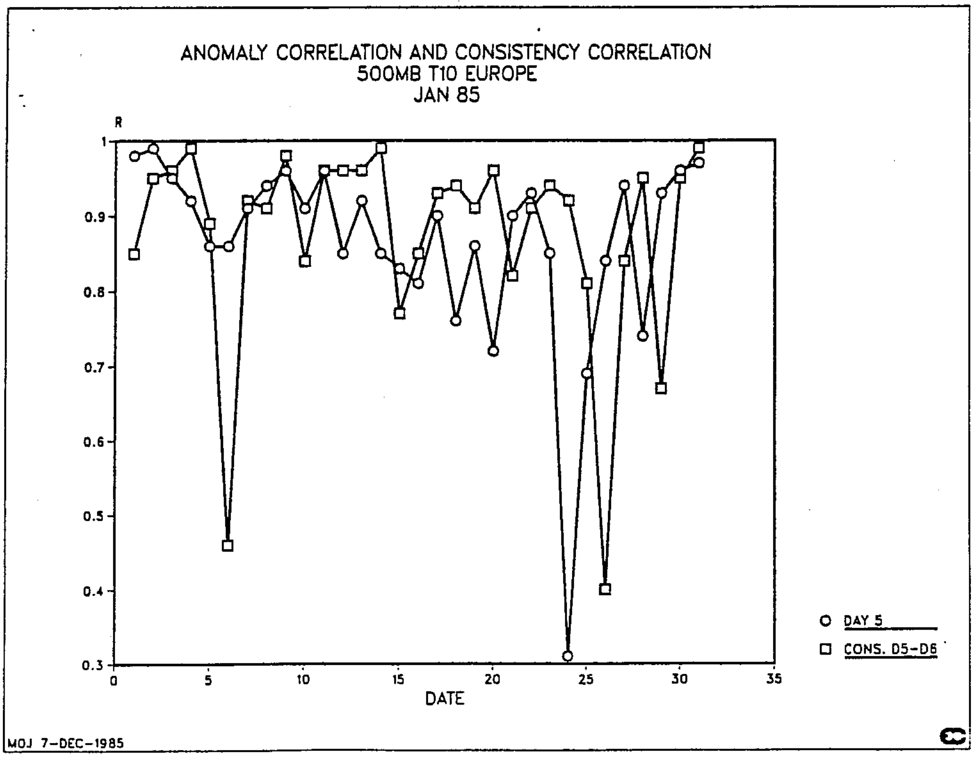
Figs. 5.1a and 5.1b show the day 5 anomaly correlation for Europe together with the one day consistency correlation for each day in December 1984 and January 1985. The correlation coefficient between the two series is 0.49 for the winter sample. The curves show that for some periods there is a relationship between quality and consistency. Sometimes the consistency drops radically from one day to the other, and in such cases the consistency does not tell us much about the forecast quality. Occasionally the new forecast will be less accurate than the previous forecasts, however, the latest forecast will probably be the best.

The correlation coefficient between anomaly correlation and consistency is plotted for winter 1984/85 in Fig. 5.2 for 1, 2 and 3 days of lagged forecasts. For a one day lag, the day for which the best relation is found is indicated.

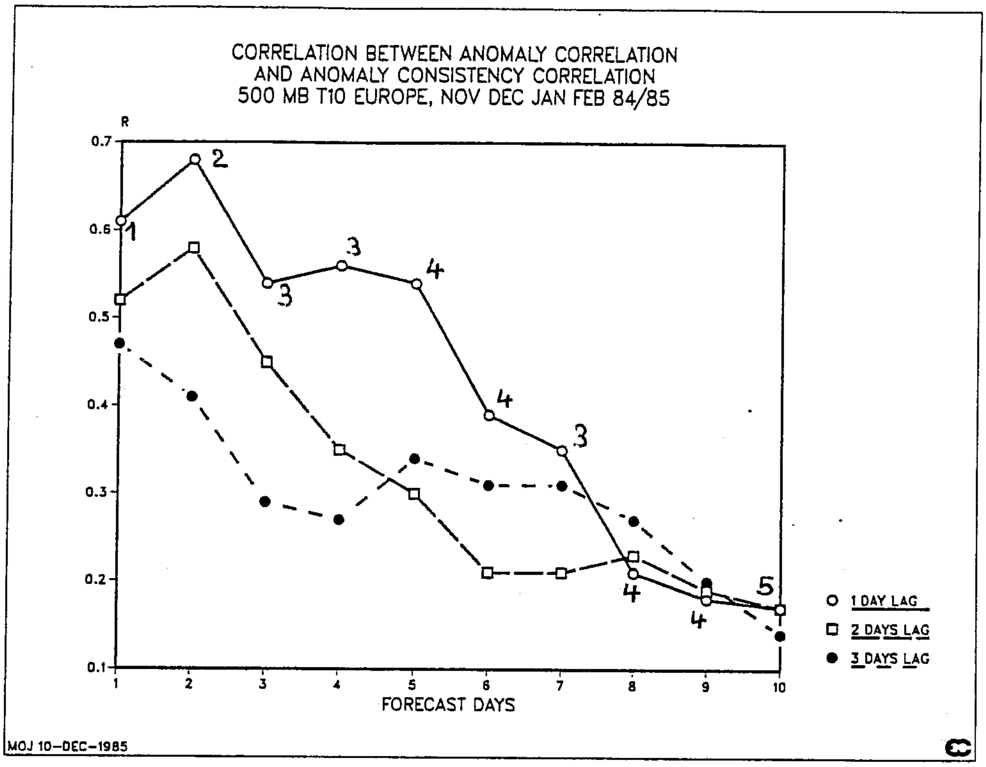
The relation between quality and consistency is highest during the first days of the forecast. After day 5 it drops off to small and probably insignificant values. Consistency at day 4 is the best predictor for quality at day 5 and also for the small correlations after day 5 the consistency at day 4, day 5 is the best predictor. After day 5 of the forecast, consistency seems to indicate consistency in the errors.



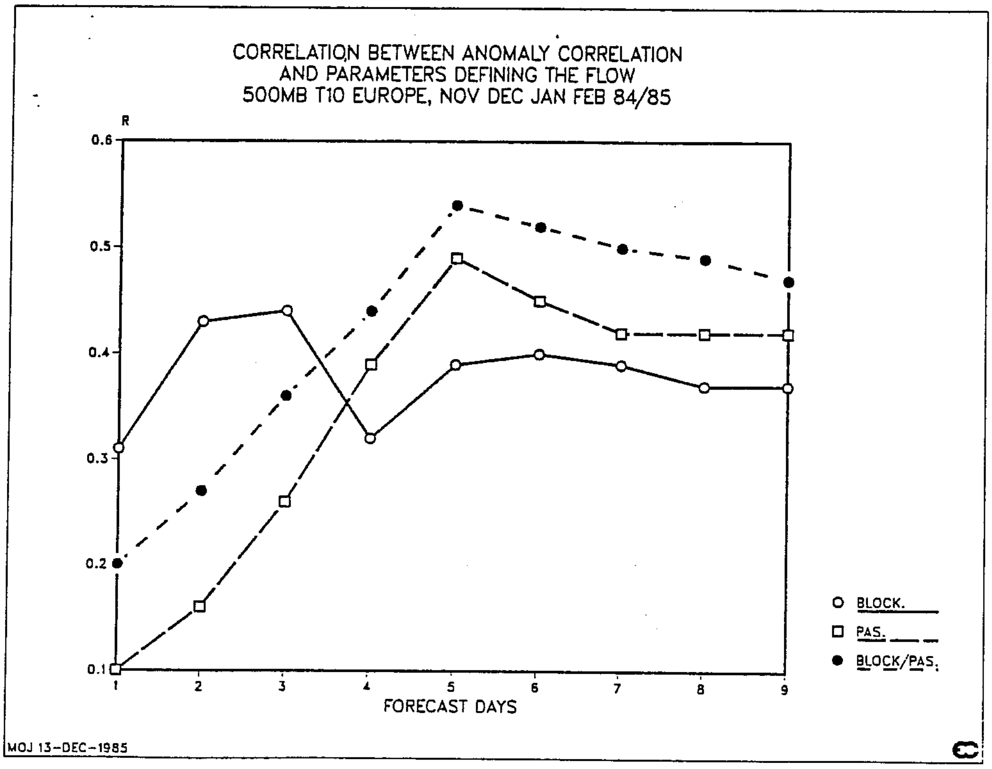
5.1a Daily anomaly correlation and anomaly consistency correlation with previous forecast for day 5, 500 hPa height fields truncated at wavenumber 10 over Europe in December 1984.



5.1b As Fig. 5.1a, for January 1985.



5.2 Correlation between anomaly correlation and anomaly consistency correlation with previous forecasts lagged by 1, 2 and 3 days; 500 hPa height fields truncated at wavenumber 10 over Europe in November 1984 to February 1985. The numbers along the graph for the 1 day lag indicate the forecast day for which the best consistency correlation was found.



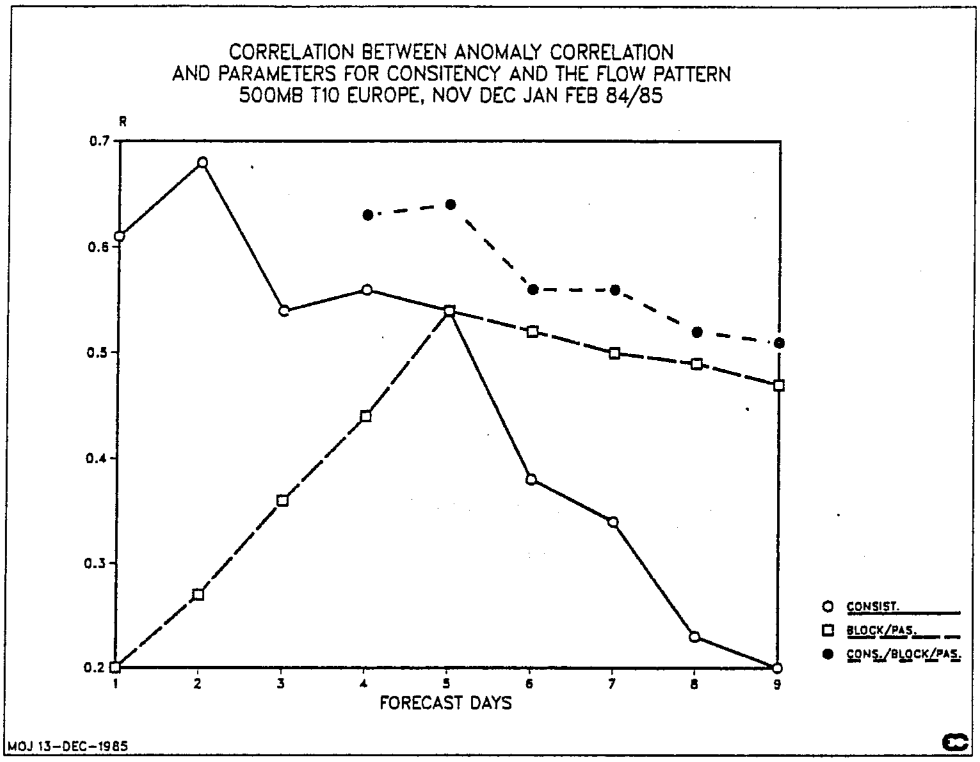
5.3 Correlation between the anomaly correlation for the 500 hPa height forecasts (T10) out to day 9 and three parameters defining the flow pattern, a block over Europe/Atlantic, a blocked flow over the eastern Pacific/western North America and thirdly the coexistence of both events. The anomaly correlation is computed over Europe during November 1984 to February 1985.

Consistency with forecasts from more than one day ago adds little information to the quality of the forecast. Averaging of the predictors in time, both within the forecast and to previous lagged forecasts, does not seem to have any significant effect.

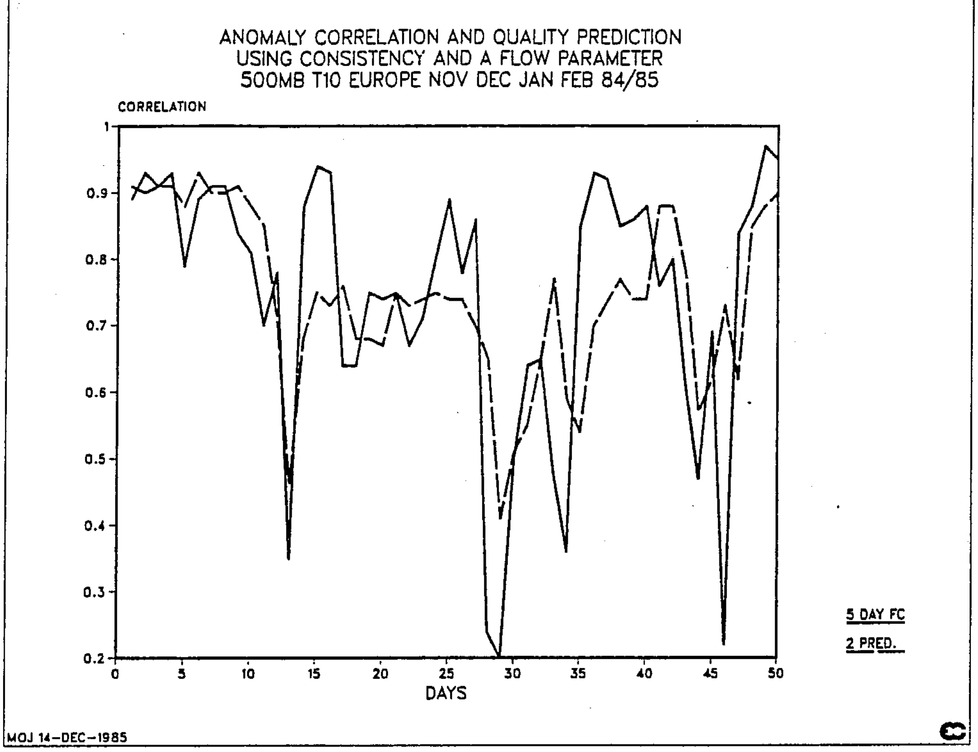
The verification results indicate that the quality of the forecast is related to the flow regimes. Grönaas (1982) found that when persistent quasi-stationary large-scale synoptic features like blocking and cut-off lows are present in the Atlantic/European area in the analysis, or predicted within the first days of the forecast, the medium range forecasts have above average scores for Europe. Here we have used this result and formed a binary predictor (called the blocking parameter) for our winter sample: this predictor is true if the criteria above is fulfilled and false if such large-scale features are absent. The decision is made subjectively.

The other idea followed here is the relation between Pacific cyclonic activity and forecast quality in Europe suggested by Ernst Klinker. For our sample the possibility of the flow preventing error propagation from the Pacific to the Atlantic/ European area is considered. When the analysis and first two days of the forecast indicate persistent ridges or a weak flow at the western coast of America and in the Pacific just west of America, then we say that this condition is present. All winter days are inspected subjectively and a binary predictor is (the Pacific parameter) formed. As a third predictor a new binary predictor is defined from the first two. When both conditions - large scale Atlantic/ European features and small Pacific influence - are present, the combined predictor is true, otherwise false. It should be mentioned that these binary predictors are very persistent and they can only explain slow variations in the forecast quality.

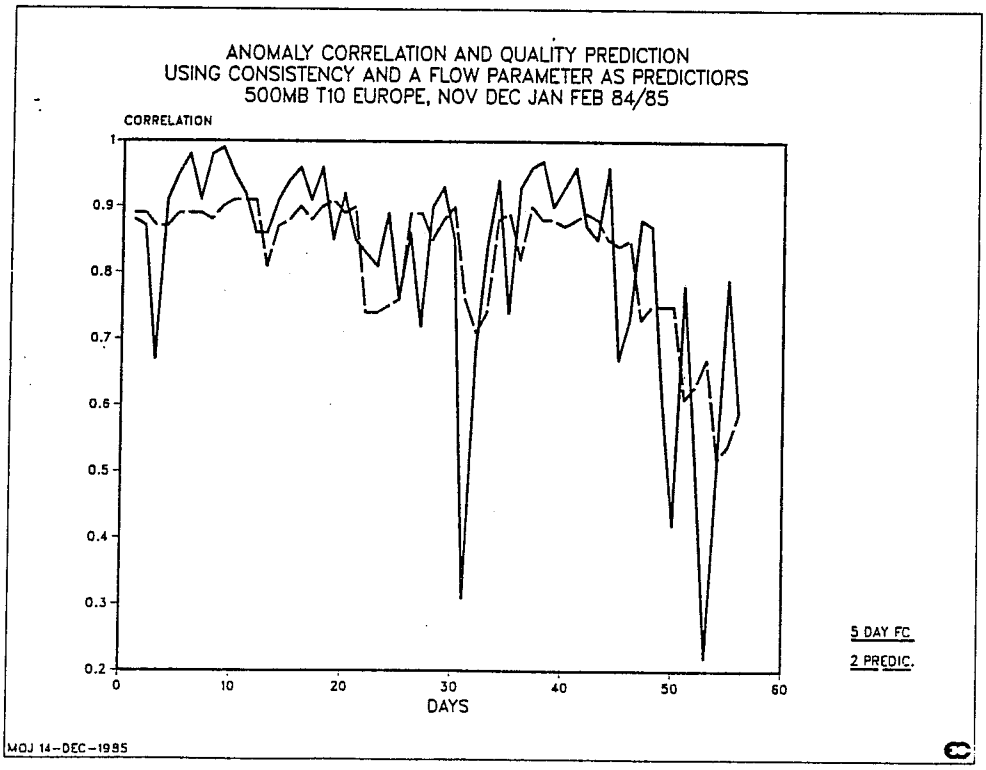
The correlation coefficient between the three predictors and the anomaly correlation for Europe are shown in Fig. 5.3. The first predictor has some correlation with the flow at all forecast steps, whilst the second predictor has, as expected, no influence at the beginning of the forecast. The highest correlation for the Pacific parameter is found at day 5 and thereafter there is only a small drop in the correlation. The combined predictor gives the best result for the medium range forecast, but the correlations are never higher than 0.5. However, this simple predictor does seem to divide the sample into forecasts which are above and below average. At day five in the forecast, the



5.4 Average correlation between the anomaly correlation for the 500 hPa height forecasts (T10) out to day 9 and the consistency with the previous forecast, the combined parameter defining the flow over the eastern Pacific and Atlantic, and the combination of the two. The area is Europe, November 1984 to February 1985.



5.5a Daily anomaly correlation and quality prediction based on consistency and a parameter defining the flow over the eastern Pacific and the eastern Atlantic, 500 hPa height forecast for day 5 over Europe truncated wavenumber 10 for the first 50 days of the period November 1984 to February 1985.



5.5b As Fig. 5.5a, continued for the second half of the period November 1984 to February 1985.

mean quality for days with combined effects of blocking/Pacific influence is 0.88 and only 0.68 when these effects are absent. For day 7 in the forecast the corresponding figures are 0.71 and 0.43. The differences represent roughly two days of predictability.

Finally, the consistency correlation and the flow parameters are combined. In Fig. 5.4 the two of the curves are taken from previous figures and show the correlation between the anomaly correlation and consistency (see Fig. 5.2), and to anomaly correlation and the combined flow parameter (see Fig. 5.3). The third curve shows the correlation when both these parameters are used as predictors. At days 4-6 in the forecast some improvements seem to be gained. The daily curves of anomaly correlation and prediction using these two predictors are shown in Fig. 5.5; the standard deviation of the residual is 0.14 and correlation 0.64.

These results indicate that there are only weak relations between forecast quality and predictors. For the first three days of the forecast, consistency alone seems to be a usable predictor. However, from days 4 to 6 the combined effect of consistency and the flow parameters seems to give the best information. Further on into the forecast, quality seems to be related to the flow alone.

6. SUMMARY AND RECOMMENDATIONS

Wide ranges of skill are found in the Centre's medium range forecasts. This aspect is of great importance to the forecasters, the users of the Centre's products, and à priori prediction of the forecast quality is desired. It is likely that the assessment of the reliability of the forecast will become one of the most important parts of the Centre's work.

Predicting the forecast quality is in some ways equivalent to forecasting the error growth in the model. In this study the synoptic evolution of the errors in the 500 hPa height forecast is investigated. Truncated T10 fields at the northern hemisphere north of 25° N is studied day by day for a summer and a winter sample. The evolution of the errors could be divided into three stages: an initial stage within the first two days of the forecast when quasi-circular patterns are developing. A following growing stage up to day 5-6 when error trains with alternative negative and positive structures are developing. Further on in the forecast, the amplitudes seem to grow more steadily, the scales are larger and the displacements for the structures are smaller.

Error correlation matrices for the samples for a European area is presented. The error correlations within the forecast seem to support the descriptions above. There exists a significant correlation between errors in lagged forecasts, especially in the last part of the forecast.

The error evolution is dependent on the flow and further synoptic research on this is highly recommended. To gain more experience about the growth of errors, error maps should be displayed on a daily basis. Since error trains are often seen in the baroclinic parts of the flow, error fields could be displayed together with thickness fields (i.e. height difference between 300 and 1000 hPa). Synoptic studies of the error growth in the early part of the forecast are also recommended.

It is important to find statistical methods to study error growth related to the flow. Canonical correlations of the EOFs have already been suggested and these should be tried.

For the winter sample the quality of forecasts for Europe is investigated in relation to the consistency of the forecast to previous forecasts, and to the flow represented by a few binary predictors. Quality is measured by daily values of the anomaly correlation for Europe. Consistency between the lagged forecast is measured by the anomaly correlation between them. Two binary flow parameters are subjectively derived. The first is true if persistent flow features exist in the beginning of the forecast in the Atlantic/European area (see Grönaas, 1982). The other parameter is true, when error propagation from the Pacific region is likely to be reduced (Klinker, personal communication). Both these predictors are highly persistent and could only explain the slow variations in forecast quality. A third binary predictor is made from the first two: when both blocking and small Pacific influence is present, this predictor is true.

No strong relationships between the quality and the predictors are found. In the first days of the forecast, consistency seems to be a useful predictor of forecast skill. However, after day 5 in the forecast, when the consistency becomes lower, there are just small correlations to this parameter.

There are some relations between the quality and the flow parameters. The combined binary predictor is a useful predictor from days 4-5 in the medium range forecast. The predictability in our sample is approximately two days longer for cases which contain blocking or little influence from the Pacific compared to cases without such features.

The results suggest that the relation between forecast skill and the flow should indeed be further investigated. Again canonical correlation on the EOFs seems to be a suitable statistical method.

So far, only modest information about the prediction of forecast quality is present. However, to gain experience, the Centre could start experimental quality prediction. On a daily basis subjective statements about the quality in terms of an average, above average and below average score, could be given.

References

Akyildiz, V., 1985, Systematic errors in the behaviour of cyclones in the ECMWF operational models, TELLUS, 37A, 297-308.

Arpe, K., and Klinker, E., 1986, Systematic errors of the ECMWF operational forecasting model in mid-latitudes, Quart. J. R. Soc., 112.

Cats, G., and Åkesson, O., 1983, An investigation into a marked difference between two successive forecasts of September 1982, Contrib. Atmos. Phys., 56, 440-451.

Charney, J.G., and DeVore, J.G., 1979, Multiple flow equilibria in the atmosphere and blocking, J. Atmos. Sci., 36, 1205-1216.

Grönaas, S., 1982, Systematic errors and forecast quality of ECMWF forecasts in different large-scale flow patterns. Interpretation of numerical weather prediction products, ECMWF seminar/workshop 1982.

Hollingsworth, A., et al, 1985, The response of numerical weather prediction systems to FGGE level II B data. Part I: Analyses Quat. J. Met. Soc., Vol. 111, Jan. 1985, No. 467.

Lorenz, E., 1982, Atmospheric predictability experiments with a large numerical model, TELLUS, Vol. 34, No. 6, 505-513.

Persson, A., 1984, The application of filtered forecast fields to synoptic weather prediction - Presentation of the products and recommendations of their use, ECMWF Operations Department Technical Memorandum No. 95.

Simmons, A.J., and Hoskins, B.J., 1979, The downstream and upstream development of unstable baroclinic waves, J. Atmos. Sci., 36, 1239-1254.

Sutera, A., 1985, Probability density distribution of large scale atmospheric flow, to be published at ECMWF.

WMO PSMP, Report Series No. 16, 1985, Results on the WMO/CAS NWP data study and intercomparison project for forecasts for the northern hemisphere in 1984, Technical Document WMO/TD No. 60.