FORECAST SKILL AND PREDICTABILITY AT THE EXTENDED RANGE USING MONTE CARLO ENSEMBLE INTEGRATIONS FROM A GENERAL CIRCULATION MODEL

D. P. Baumhefner
National Center for Atmospheric Research*, Boulder, Colorado, USA 80307

1. INTRODUCTION

The possibility of deterministic numerical forecasts for periods longer than the typical synoptic time–scale was originally suggested as the major goal of the Global Atmospheric Research Program (Panel on International Meteorological Cooperation, 1966; U.S. Committee for the GARP, 1969). This possibility was also optimistically underscored by the leading scientists involved in the operational production of long range forecasts at that time (Namias, 1968). Reliable estimates of predictability error growth for daily time–scales (Lorenz, 1982; Baumhefner, 1985) which lead to potential limits of forecast skill on the order of 10 days seem to preclude any chance of this goal being realized. However, new estimates of error growth which utilize time averaging of simulated atmospheric flows extend the limit of daily skill into the monthly time range (Tribbia and Baumhefner, 1988). This potential skill has been confirmed in a series of landmark papers (Miyakoda *et al.*, 1983; Miyakoda *et al.*, 1986; Miyakoda *et al.*, 1987) which showed conclusive evidence of forecast skill at extended range for certain situations. Recently, various operational forecasting centers have tested the concept of numerical extended range forecasting with similar success. The European Centre for Medium Range Forecast (ECMWF) studies (Molteni *et al.*, 1987, Brankovic *et al.*, 1987) have concentrated on the influence of horizontal resolution while the British Meteorological Office research (Murphy, 1988) has emphasized the importance of ensembling forecasts. The National Meteorological Center experiments (Tracton and Kistler, 1988; Kistler *et al.*, 1988) illustrate the behavior of forecast skill from a very large sample of contiguous forecasts in time. A consensus of opinion that can be drawn from the previous experiments result in the following conclusions: 1) Accurate forecast skill at the extended range is episodic in nature, 2) Unlike daily forecasting, systematic model errors become a major component of the total error, and 3) Ensemble averaging of individual forecasts is beneficial for verification and may be useful in predicting skill in an *a priori* fashion.

---

* The National Center of Atmospheric Research is sponsored by the National Science Foundation

Based on these general conclusions, an alternate approach to the extended range prediction problem is suggested and tested. Since the systematic component of the forecast error at extended range is manifested by a drift of the model from the observed state to the model climatology, the use of a low resolution climate model (GCM) which is designed to minimize this drift is considered. If the climate model forecast skill is comparable to the higher resolution forecasts, then a relatively large ensemble of forecasts can be integrated from similar initial states for the same cost as one high resolution forecast. The primary benefits of such a large ensemble include a reasonable estimate of the average forecast and an accurate assessment of the spread of the forecasts.

A set of forecast experiments was designed to test this hypothesis. Forecast integrations to 30 days were made using the latest version of the National Center for Atmospheric Research (NCAR) General Circulation Model at T31 resolution for a 12 case sample of recent wintertime situations. An ensemble of forecasts was constructed for each initial condition by adding perturbations to the existing operational analysis. This procedure is referred to as a Monte Carlo forecasting scheme (Leith, 1974) in which a finite sample of equally likely initial states are determined for a given time. Therefore, the amplitude, spectral distribution and vertical structure of the perturbations were carefully chosen to simulate the current uncertainty present in current observation/analysis techniques. The ensemble size was arbitrarily selected to be 10, based on arguments presented by Leith (1974). Preliminary results indicate that this sample describes the spread of possible forecasts in a reasonable manner.

The actual forecast skill was evaluated for each member of each ensemble by comparing it to the observed atmospheric behavior. The spread of the forecast skill and the skill of the average of each ensemble was compared as well. Cases were selected from the verification statistics which typified accurate and inaccurate extended range forecasts, and examined synoptically. The potential limit of predictability of the twelve atmospheric states was estimated by calculating the evolving differences among the individual forecasts in each ensemble. Case to case variation of predictability as well as the average behavior was studied. Finally the relationship of the dispersion of forecasts within an ensemble to the actual skill of the ensemble average was investigated by correlating the average skill to the averaged predictability estimates, and the average skill and its spread to the spread of the forecasts. A summary of the experimental design, data, and the model used is found in Section 2. The forecast skill evaluation is found in Section 3 and the estimates of predictability error growth in Section 4. Sections 5 and 6 contain the discussion of predictability/skill relationships and some general conclusions.

2. EXPERIMENTAL DESIGN

The construction of the Monte Carlo sample of initial states is a crucial element in this experimental design. Following the guidelines of Leith (1974), each state in the sample should reflect a possible initial condition of the atmosphere at a given time. This restriction is quite different from a lagged average approach in which the perturbations include both model and initial data error and do not necessarily represent a possible initial state of the atmosphere. An attempt is made to simulate the current observational and analyses errors by imposing three general constraints, which were obtained from an analyses error diagnosis paper by Daley and Mayer (1986) and a Global Weather Experiment analysis comparison by Baumhefner (1985). The first constraint limits the amplitude of the perturbation in a global RMS sense to the currently estimated analysis errors for various atmospheric variables. The second constraint orders the spectral distribution of the perturbation such that only the smallest scales beyond wave 30 are totally in error (*i.e.*, the difference variance equals the total variance of that scale). The third constraint vertically filters the perturbation by projecting onto the normal modes of the model retaining only the first four components. This process builds in a vertical coherence to the perturbation.

The global analyses, in this case the National Meteorological Center (NMC) operational global optimal interpolation system, was used without modification as the first member of the ensemble. Nine perturbations were generated from the analyses using a geographically random distribution and imposing the previously discussed constraints. A perturbation was applied to all the independent variables of the model. Typical values of hemispheric Root Mean Square (RMS) difference among members of the ensemble are about 1.2° for the temperature fields, 1.5 m sec^{-1} for the velocity fields and .3 g kg^{-1} for the moisture fields. These values produce height differences at 500 mb of the order of 20 m RMS. Each vertical level in the model was perturbed with a different random distribution but with the previously mentioned amplitude and spectral limits. The resulting difference fields, when mapped geographically, resemble a forecast error field very early in an integration with many maxima near the scale of the Gaussian grid. The typical value for the maxima range from three to four times the RMS value for the specific field in question.

Twelve different wintertime initial states were chosen from the NMC global analysis archive in order to sample as many flow regimes as possible. The forecasts begin on the first of December and the first of January for the winters of 1981–82 through 1986–87. The forecasts were integrated using a long term climatological monthly average of the global sea surface temperature (SST) derived from the Comprehensive Ocean

Atmospheric Data Set (COADS), which was then interpolated daily during the forecast. A subset of forecasts have also been made using the observed monthly mean SST (as analyzed by Climate Analysis Center, NMC) and interpolated to daily values (Baumhefner *et al.*, 1988). This paper will discuss only the climatological SST results. The original and perturbed analyses were initialized with a non-linear normal mode scheme (Errico and Eaton, 1987). Only the first six vertical modes and frequencies shorter than 30 hours were initialized. The initialization produces only minor changes the perturbation structure, except for the geopotential height fields. The amplitudes of the height perturbations are reduced by a factor of two or three in some cases.

All forecasts were produced from a T31–12 layer version of the NCAR general circulation model (CCM1), documented by Williamson *et al.* (1987). The new model, when compared to the previous version, includes increased vertical resolution in the stratosphere, improvements in the radiation code, refinements in the cloud parameterization and numerics, and changes in the boundary layer formulation. A long-term, annual cycle control run with this model produced a much improved climate simulation compared to the previous R15 climate model (CCM0), especially in the Tropics and Southern Hemisphere (M. Blackmon, personal communication). The Northern Hemisphere wintertime simulation of the stationary waves has also improved, although at the expense of losing some variance in the lower frequency variability.

The forecasts were integrated using the seasonal cycle version of the model and were evaluated directly without any post processing, adjustments, or removal of systematic components of error.

3. FORECAST SKILL

Each member of the twelve ensembles was evaluated for forecast accuracy using several measures of skill and for various time-means. In this paper only the 500 mb height anomaly correlation (AC) for 30 day time-means is presented. The AC is calculated by deriving a forecast anomaly from an eight year (1980–1987) climatology of the month being forecast and correlating that pattern with the observed anomaly for the same period. Figures 1–2 present a summary of skill for each ensemble and an average of the ensembles for two geographical areas.

Examination of the hemispheric scores reveals several interesting characteristics. The score of the average forecast (the sum of the individual members) from each ensemble is in all cases except DEC83 better than the mean value of the individual members. In over half the sample the increase in skill is equal to half the standard deviation of the ensemble skill range. This result clearly highlights one advantage of the ensemble approach to extended range forecasting. The case by case scatter of skill is quite large; in agreement with previously cited work. It is remarkable, however, that six cases exhibit correlations above .5 for the average forecast, with only two cases, DEC83 and DEC84, completely useless with AC values below .2. The rest of the cases are only of marginal use in a hemispheric sense; however, in some local areas, there is still some skill.

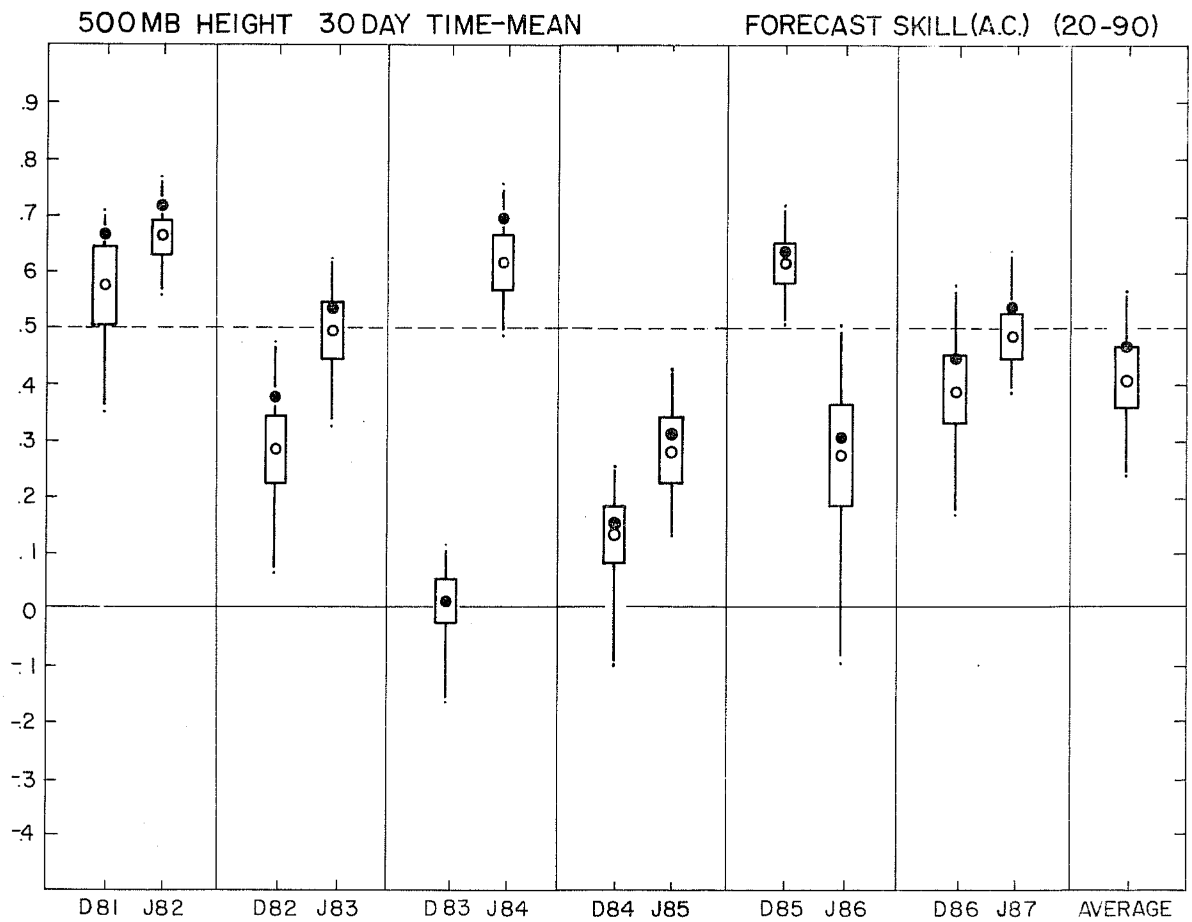

Fig. 1. Anomaly correlations of forecast skill 30 day time–mean 500 mb height field with observed state for latitude band 20–90N. Dates along abscissa indicate starting times of ensemble forecasts. D equals DEC 1 00Z, J equals JAN 1 00Z, 81 equals 1981, AVERAGE equals arithmetic mean of 12 ensembles. Open circle equals mean skill score of ensemble. Open bar equals standard deviation of scores within ensemble. Vertical solid line equals range of score within ensemble. Solid circle equals skill of average forecast of ensemble (sum of individual members).

The spread or range of skill within each ensemble is quite large, typically covering 20–30 correlation points. Fortunately, there is still considerable coherence between the skill of the average forecasts and that of the individual members. For example, none of the individual member forecasts have skill below .5 in JAN82, JAN84, or DEC85. Particularly large ranges of skill occur for the marginally useful cases of DEC82 and JAN86. The relatively large spread of skill from realistic analysis error estimates illustrated here should inject a note of caution when interpreting skill for a single or small sample of forecasts for a given date.

The average skill for the 12 case sample for the average forecast is a respectable .47. This value confirms the usefulness of the low resolution model in making extended range forecasts. For the sum of the 12 cases the average forecast is seven points better than the average skill of all the forecasts. The average standard deviation and range cover 11 and 33 points of skill respectively. These ranges indicate that, on the average

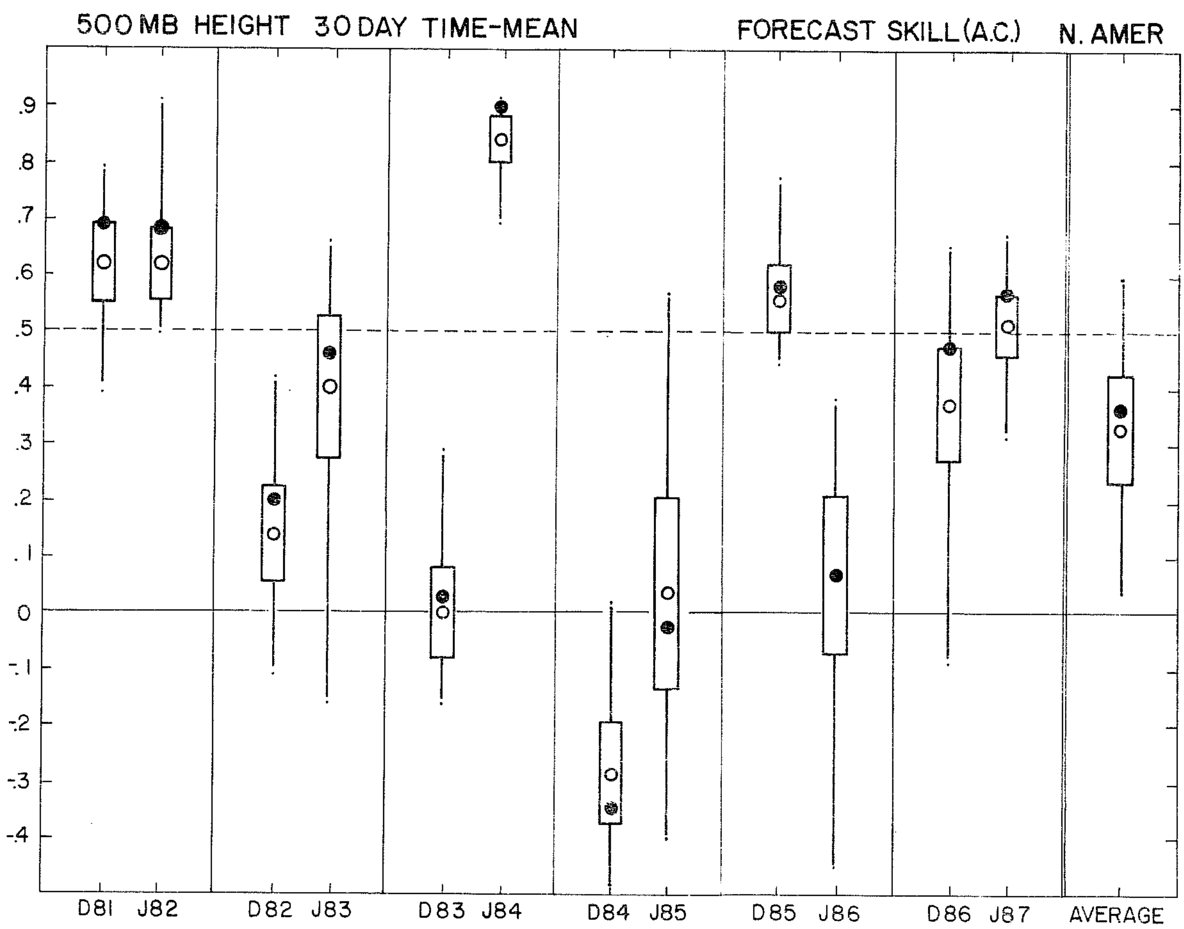

Fig. 2. Same as Fig. 1, except for limited region of 150–60W and 25–70N.

for a ten member ensemble using realistic analysis errors, one should expect several skillful forecasts and several mediocre ones.

When skill is evaluated over regional domains the size of North America, as in Figure 2, the results are by and large similar, with considerable correlation of skill between the two verification areas. Because the area averaging is less and typically only a single flow pattern is measured, the skill and spread of skill are much more variable. This characteristic produces spectacular successes (JAN84) and dismal failures (DEC84) yielding a mean skill that is slightly less accurate than the hemispheric value. The high variability of skill strongly points out the need to discriminate the skillful forecasts before they verify. Note that averaging individual members becomes much less effective when all the forecasts in the ensemble are already unskillful.

The sample of 12 cases allows for a few direct comparisons of skill with other models of higher resolution. The skill of four cases for DEC81, JAN82, DEC82, JAN83, can be compared with the T21–T42 results of Molteni *et al.* (1987). In all cases the NCAR T31 skill was equivalent to or better than the T42 skill. Surprisingly, the range of skill from the T31 forecast ensembles did not even overlap the relatively poor skill of the T21 results in three of the four cases. A comparison of a single case of JAN87 found in Tracton and Kistler (1988) also showed equivalent skill compared to their R40 forecast. Further direct comparisons of low resolution monthly forecasts with higher resolution experiments (Baumhefner, 1988) show superior skill from a R15 model in six of the eight cases used by Miyakoda (1986). Indirect evidence also suggests that low resolution climate models can be quite successful in extended range forecasting. The 108 day average skill of .39 for an R40 model (Tracton and Kistler, 1988) is equivalent to this 12 case sample of .40 for the mean skill.

It is important to validate the impressions given by the AC scores by examining the actual forecast and observed anomalies synoptically. The JAN82 and DEC83 cases were selected for this inspection because they represent typical examples of skillful and unskillful forecasts. Figure 3 shows the observed anomaly for JAN82 along with the average forecast from the ten member ensemble and the best and the worst individual members. The observed pattern features a strong negative center over North America downstream from a broad positive anomaly covering the entire Pacific. The Atlantic–European area exhibits a train of anomalies with positive values over Western Europe. The average forecast accurately predicts the Pacific/North American couplet. The European sector patterns are less accurate; however, all the major features are still identifiable. The average forecast does not suffer from smoothing normally associated with averaging; instead the forecast amplitude of the anomalies are too strong. The AC score of .71 indicates the excellent correlation of phase in this forecast. The same basic features are present in both the best and the worst forecast, indicating not much change of phase; however, there are considerable changes in amplitude. This characteristic of the forecasts results in a rather tight range of AC from .76 to .55. Note the worst forecast is particularly poor over Europe where negative values are found in place of

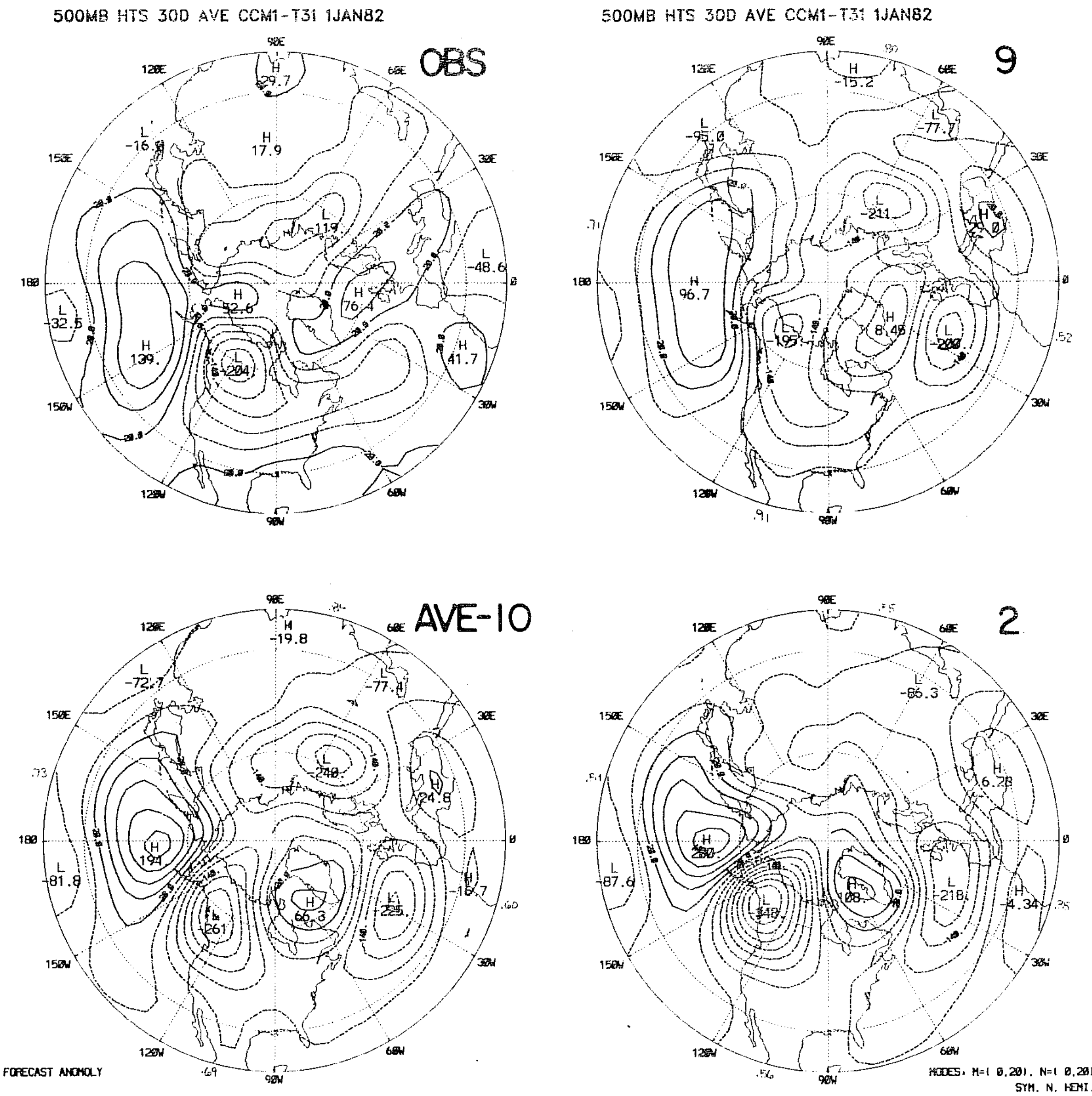

Fig. 3. Observed and forecast anomalies from 1980–1987 climatology of 30 day time-mean 500 mb height field for JAN 1982 over Northern Hemisphere. Upper left equals observed anomaly, lower left equals ten-member average forecast anomaly, upper right equals best forecast anomaly of ensemble (case 9), lower right equals worst forecast anomaly of ensemble (case 2), contours interval equals 40 m. First contour equal to ± 20 m.

positive values. Clearly, from a synoptic point of view, this 30 day integration contains a great deal of useful information.

The forecasts for DEC83 illustrate an entirely different story. The observed patterns are quite different from JAN82 with a strong ridge over Alaska and negative values over the Pacific. The Atlantic pattern is also reversed in sign; however, the negative anomalies in North America and Eastern Europe are similar to JAN82. The average ensemble forecast has little correspondence to the observed anomalies with the

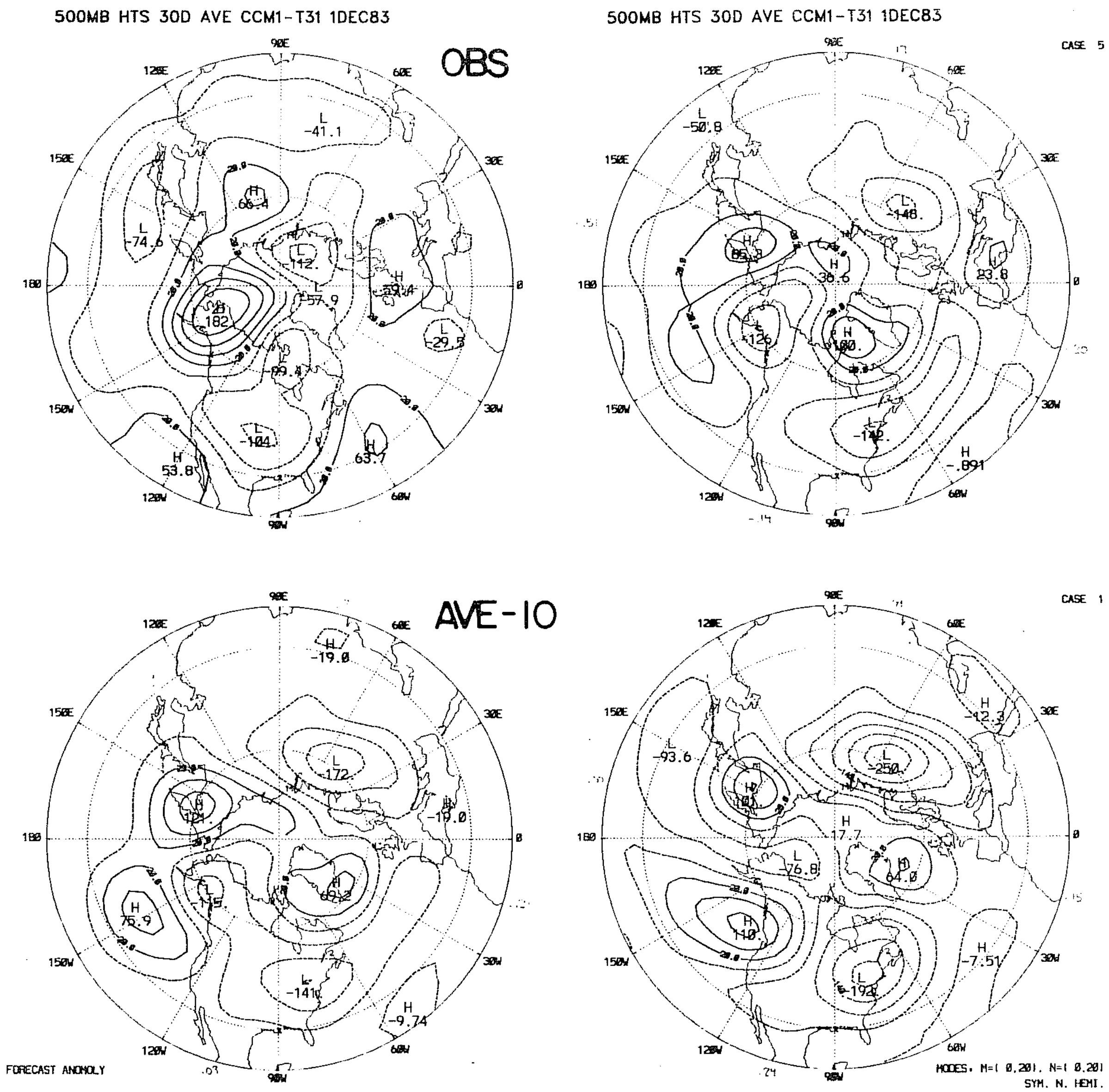

Fig. 4. Same as Fig. 3 except for DEC 1983 with worst forecast anomalies (case 5) in upper right and best forecast anomalies (case 1) in lower right.

exception of eastern North America and Eastern Europe. In fact, most features are out of phase by more than 90° which confirms the AC calculation of .01. The large blocking anomaly over Alaska is out of phase by nearly 180°. The best forecast member in this case, with only a meager AC score of .11, shows significant differences over the Pacific compared with the worst member. However, in general the range of patterns in the ensemble do not depart radically from the average forecast. As indicated by the skill score there is virtually no useful information contained in this forecast.

4. PREDICTABILITY ESTIMATES

A highly beneficial by-product of the Monte Carlo approach of forecast ensembling is the implicit estimate of predictability error growth. Using the classical method of twin forecasts (Panel on International Cooperation, 1966; Baumhefner, 1985; Tribbia and Baumhefner, 1988), which assumes that the model can realistically replace the atmosphere, an estimate of predictability is derived from the ten-member forecast ensembles. Forty-five pairs of forecasts were differenced for each ensemble and the average difference, its standard deviation, and range were calculated in a manner analogous to evaluating forecast skill. These scores then represent estimates of the best possible forecasts that can be achieved (*i.e.*, the upper limit of skill) because they assume that the "model" that replaced the observed atmosphere has no error. Therefore, by definition, the only errors present in this predictability estimate calculation are those due to initial state deficiencies.

The predictability results for 30 day time-means at 500 mb are presented in a similar fashion as the forecast skill for easy comparison in Figures 5-6. Concentrating first on the average behavior of the 12 ensembles, the average difference AC score for the predictability estimate is .75, which is 30 percentage points better than the measured forecast skill. It is noted here that this value of predictability for 30 day time-means is more optimistic than that found in Tribbia and Baumhefner (1988). The reasons for this discrepancy are linked to differences in initial data sets, formulation of models, and sampling strategy. The differences between these predictability estimates underscore the relative uncertainty in these calculations and point to the need for additional research on the subject. The average range of predictability is also quite large and comparable to the forecast skill range. This implies a strong sensitivity of the initial state to the type and size of the perturbation used in the Monte Carlo ensemble. Therefore the sample size must be large enough to detect this range of uncertainty otherwise the predictability (spread) estimates will be severely biased. Clearly, if these predictability estimates are correct, then considerable skill can be gained at extended range by reducing model error.

There is a wide variation in average predictability and its scatter depending on the initial state. For example, DEC82 shows the lowest average AC and also the largest spread about the mean. Whereas DEC85 predictability is very high with virtually no spread at all. The causes for this case dependent variability are unclear; however, its presence may signal possible relationships with forecast skill.

The regional behavior of the predictability estimates (Figure 6) is similar to the previously analyzed forecast skill with larger uncertainty and lower average values of AC. The characteristics of the individual cases of DEC82 and DEC85 are similar to the hemispheric estimate and three other cases also show the large scatter of DEC82. The rapid dispersion among forecasts for regional areas is somewhat discouraging and

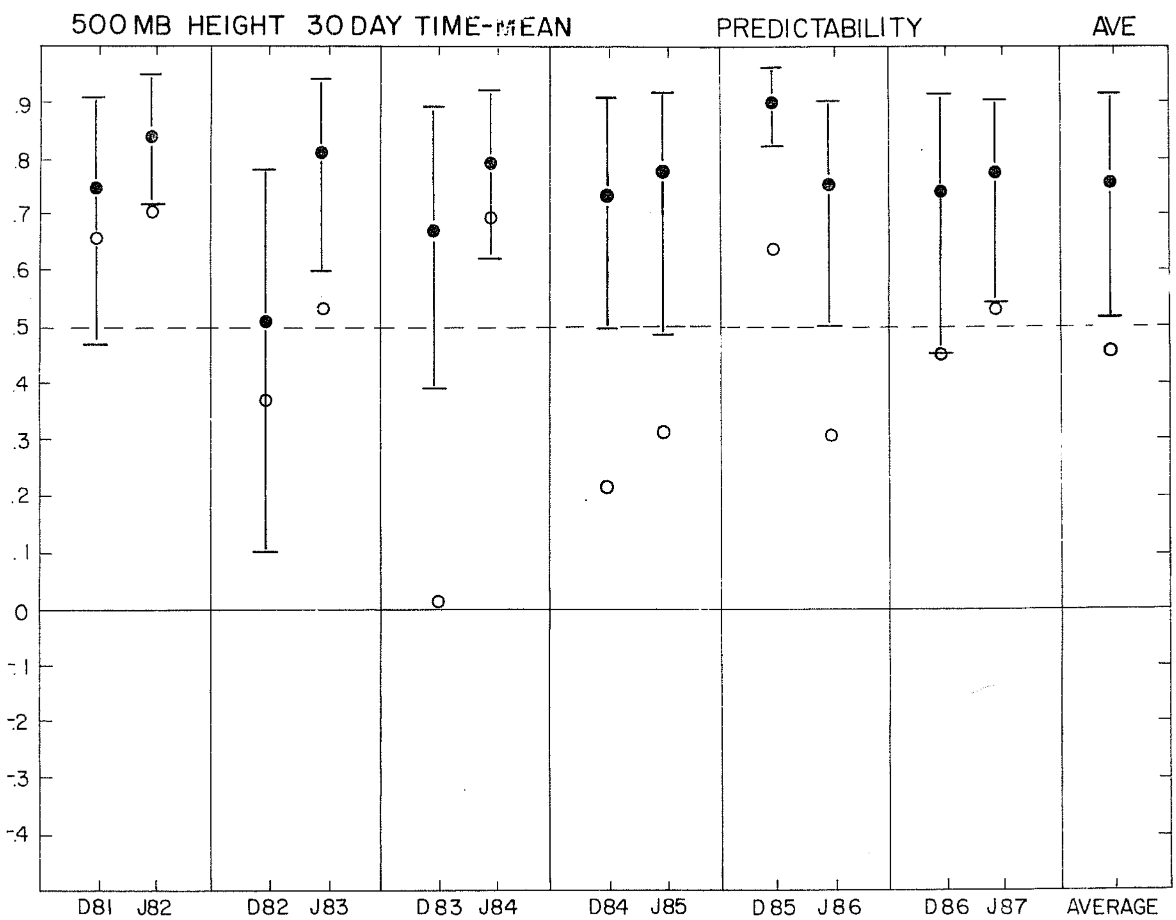

Fig. 5. Anomaly correlations of forecast predictability for 30 day time-mean 500 mb height fields for latitude band 20–90N. Solid dot equals average of 45 pairs of forecasts for each ensemble. Vertical bar equals range of scores. Open circle equals Score of average forecast from Fig. 1. Abscissa values same as Fig. 1.

probably plays a large part in the wide range of forecast skill as well. This uncertainty factor or *de facto* signal to noise ratio will play an important role in the accuracy of *a priori* estimates of forecast skill.

5. PREDICTABILITY-FORECAST SKILL RELATIONSHIPS

The primary function of a Monte Carlo ensemble of extended range forecasts would be to link the characteristics of the predictability estimates previously described to the actual forecast skill in some fashion. This kind of calculation was the original goal of stochastic-dynamic prediction first outlined by Epstein (1969). In this experimental design, however, the uncertainty relationships are not directly calculated, but must be derived. A detailed examination of Figures 1–2 and 5–6 together suggest several possible relationships.

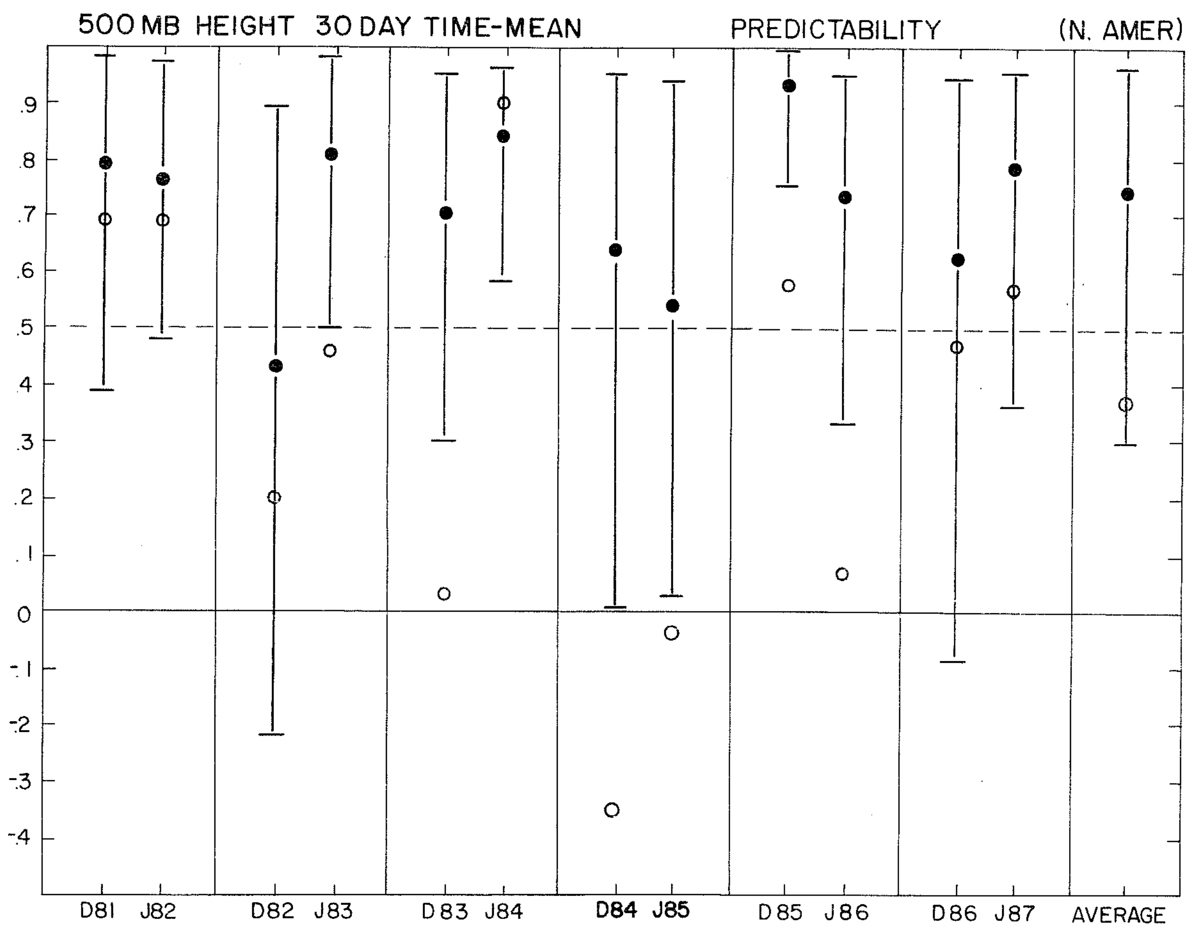

Fig. 6. Same as Fig. 5 but for limited region of 150–60W and 25–70N.

First, the average value of forecast skill appears related to the average value of predictability. A correlation of these values would indicate a relation between slower predictability error growth and accurate forecast skill. Secondly, the average forecast skill may be correlated to the spread of the predictability estimates. This relationship would imply that highly dispersive situations are more difficult to forecast. And finally, the spread of forecast skill looks similar to the spread of predictability which may

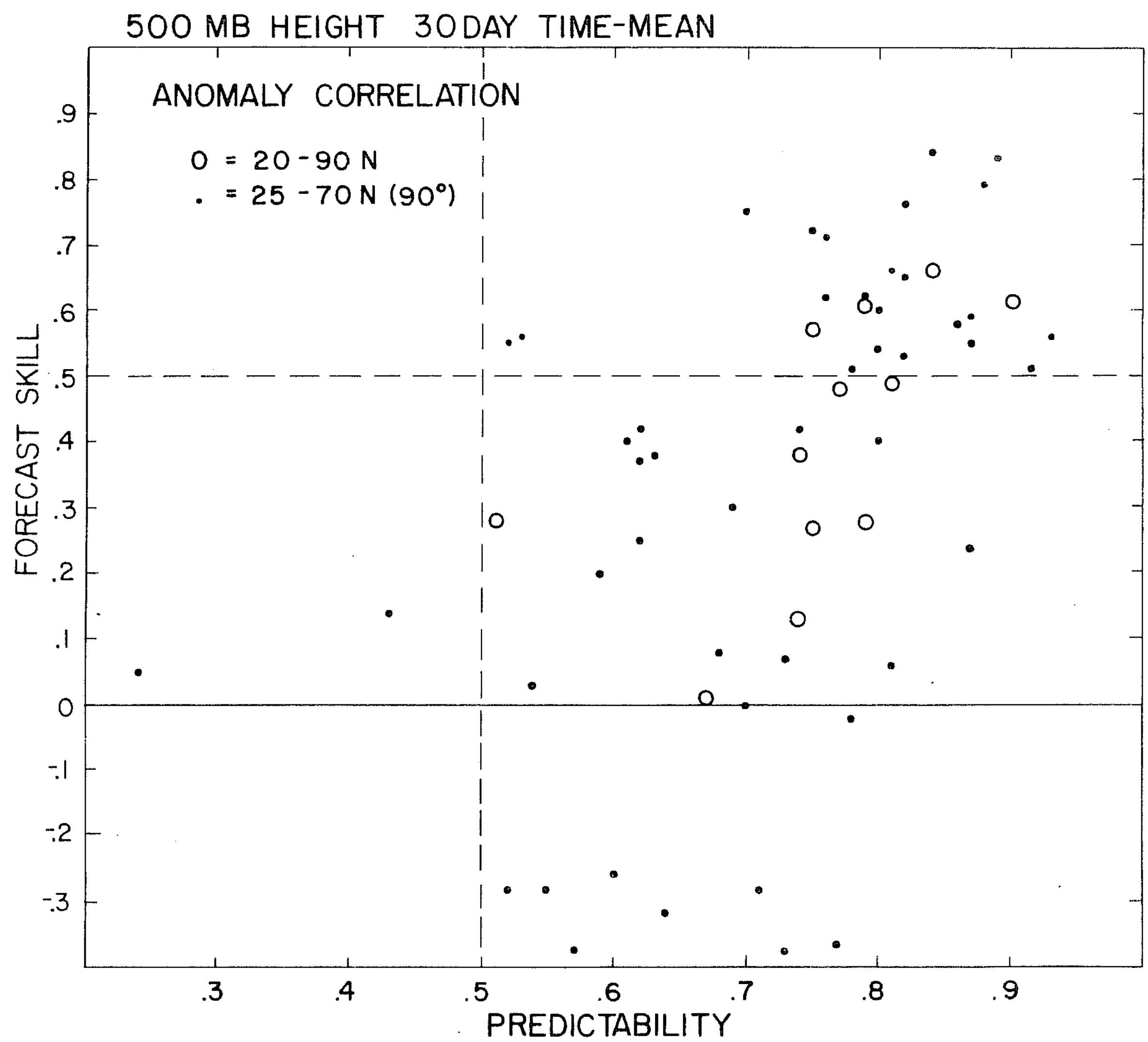

Fig. 7. Scatter Plot of the average anomaly correlation of 500 mb height 30 day time–means for each 10–member ensemble. Abscissa equals average of the differences between forecasts (predictability value). Ordinate equals average forecast skill. Open circle equals scores calculated over hemisphere from 20–90N. Dot equals regional scores from 90° windows from 25–70N (first window starts at 150W).

in turn be related to average skill. For this relationship to hold, the errors caused by initial condition uncertainty would have to be of the same order as the modeling error.

The first relationship is tested by constructing a scatter-plot of the average values of forecast skill and predictability for all 12 cases and both the hemispheric and regional domains (Figure 7). Four windows, 90° of longitude wide, were used to calculate the regional scores. The hemispheric domain scores show the greatest correlation with

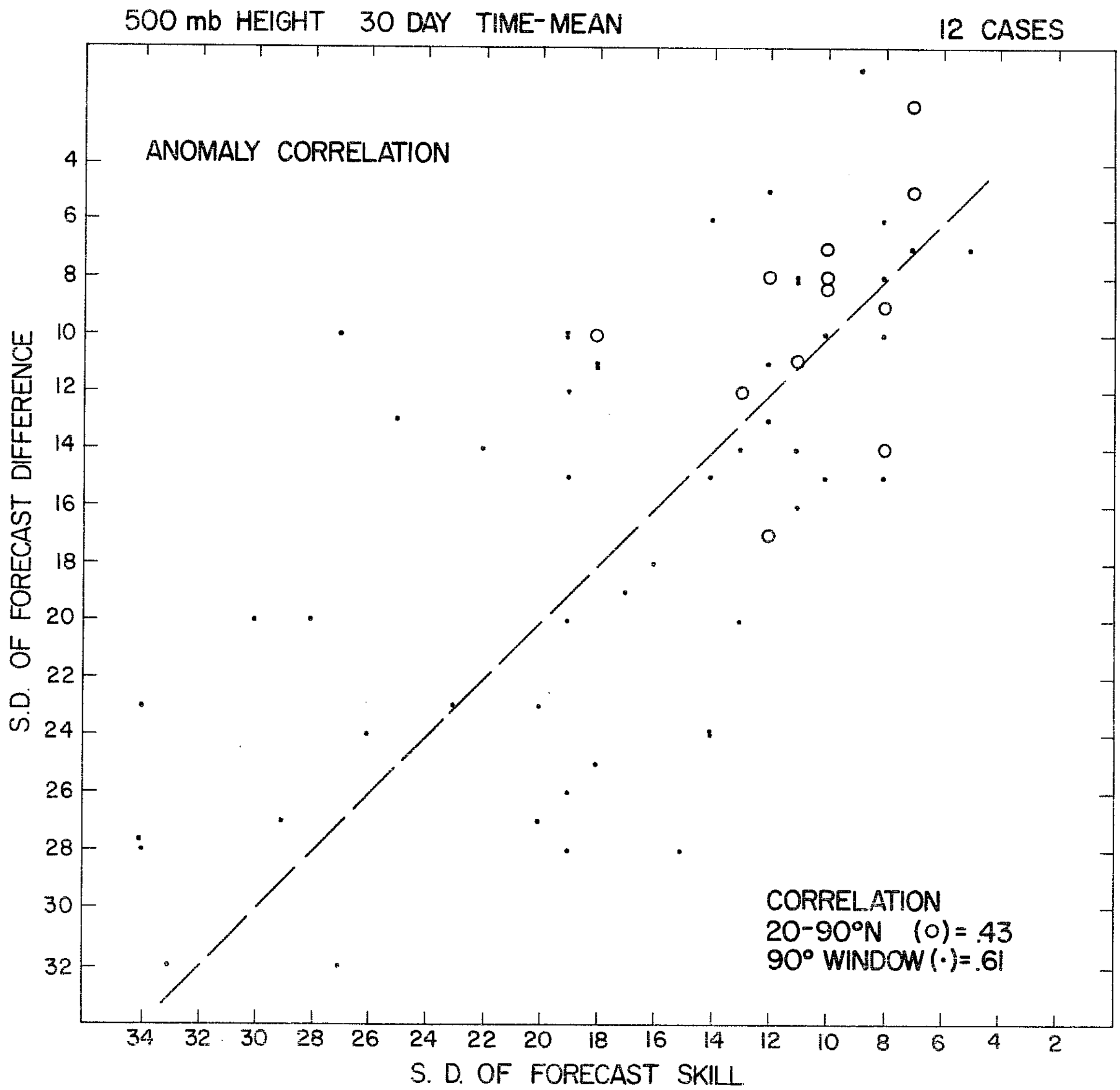

Fig. 8. Scatter plot of the dispersion of forecast skill compared to dispersion of predictability estimates for each ensemble of 12 cases. Dispersion measured as the standard deviation of each sample of the anomaly correlation of 30 day time-means of 500 mb height. Abscissa equals spread of forecast skill. Ordinate equals spread of predictability. Open circle and dot same as Fig. 7.

a value of .59, but the sample size of 12 is quite small. The regional score sample of 48 shows less correlation (.49) and larger scatter. Overall, there appears to be some relationship between skill and predictability in an average sense. Further subdivision and/or expansion of the sample size may be necessary to sharpen this relationship.

The second possible connection between forecast skill and predictability relates the dispersion of forecasts to skill. This relationship is referred to as forecast agreement in Kistler *et al.* (1988). In this case the average standard deviation (forecast skill) of each ensemble is correlated against the spread of the ensemble as measured by the standard

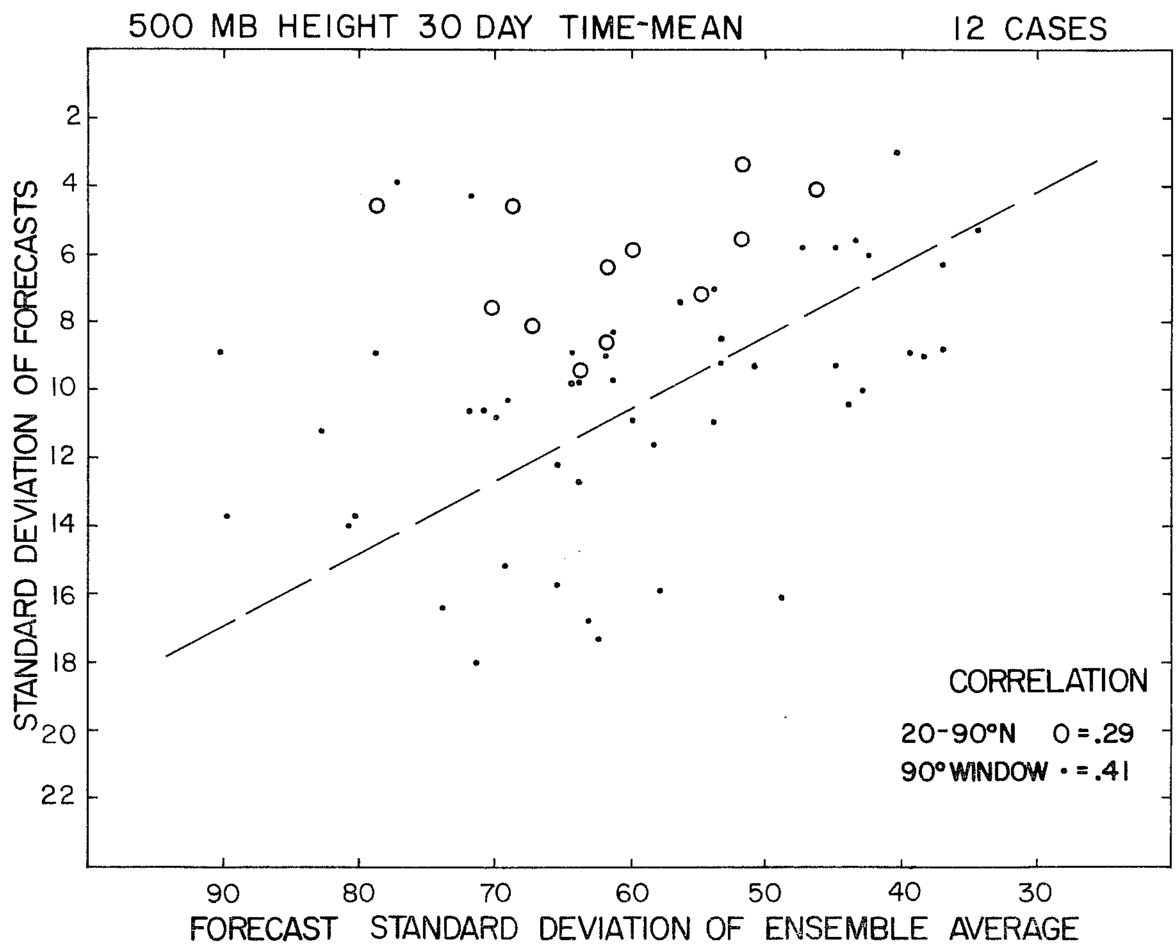

Fig. 9. Scatter Plot of average forecast skill of 30 day time-mean 500 mb heights compared to spread of the skill for 12 cases. Abscissa equals average forecast skill of ensemble measured by standard deviation of difference between forecast and observed condition. Ordinate equals spread of forecast skill within ensemble as measured by standard deviation. Open circle and dot (same as Fig. 7).

deviation of the range. The scatter plot looks very similar to Figure 9 and is not shown. The hemispheric values exhibit very little correlation whereas the regional values show only a slightly stronger relationship. The dispersion of forecasts is greater in the regional domains which may account for the better correlation. The correlation with forecast agreement found by Kistler *et al.* (1988) was much better than these results indicate. Again there appears to be a connection between these variables, but it is very weak and needs further study.

The final relationship that was investigated compared the spread of skill to the spread of the predictability estimates within each ensemble. Figure 8 displays the correlation of these two parameters in terms of the standard deviation of each respective range of values. The results show the best relationship between variables that has been tested with the regional scores having a correlation of .61. Unfortunately this connection between the ranges of the ensemble skill and predictability is not very useful as a direct prediction of skill but only how dispersive the forecast sample will be. The relation of average skill to the range of skill is shown in Figure 9. The correlation is weak at best and therefore puts the good relationship between spreads at a disadvantage in predicting forecast skill.

6. CONCLUSIONS

In general the use of a low resolution general circulation model coupled with Monte Carlo Ensembles proved to be highly successful for extended range forecasting. The following conclusions briefly summarize specific aspects of the experiment.

Evaluation of forecast skill, in an average sense, showed that the low resolution results were as skillful as higher resolution forecasts at the 30 day time range. On a case by case basis, there was considerable variation in skill as seen in earlier studies. The spread of skill within each ensemble caused by initial condition uncertainties was uncomfortably large. If scores were evaluated regionally, the average skill was similar to the hemispheric values but the case by case variation and ensemble spread of skill became even larger. The forecast skill, as measured by the verification scores, agreed with synoptic impressions of the differences between forecast and observed anomalies.

The average predictability estimate, obtained from the forecast dispersion within the ensembles, was much higher than the current levels of forecast skill. As in the case of forecast skill there was large sensitivity of these estimates to different perturbations in each ensemble, different initial conditions, and geographical location of verification. Direct relationships between case forecast skill and case forecast dispersion were somewhat weak but on the whole still positive. The strongest correlation compared the dispersion of skill and the dispersion among the forecasts themselves.

Several comments regarding these conclusions are in order. On the optimistic side the large difference between current forecast skill and the upper limit of skill estimated by the predictability calculations imply that considerable improvement in skill can be achieved by reducing model error. On the pessimistic side the case to case variability of skill exhibited in this sample requires that a method must be developed to forecast forecast skill. The first cut at deriving such a method was not very encouraging. The large spreads within the ensembles represent a two–edged sword. They hopefully can be used in a more intelligent manner to determine an *a priori* forecast skill measure. However, they also imply a large uncertainty exists in the final estimate of skill which cannot be removed. The ultimate reality of numerical extended range forecasts will most likely balance between these two facts.

Future research on extended range prediction should concentrate on three areas; expansion of the sample size of initial states, determination of regime structure and their relationship to skill, and examination of the initial perturbation methodology and the accuracy of the resulting dispersion distribution. The current sample sizes are not nearly large enough to statistically relate the previously discussed parameters or to determine regime dependencies on forecast skill. Many of the future methods of forecasting the skill of extended range integrations will undoubtedly depend on the accuracy of the dispersion estimates, therefore extensive testing of Monte Carlo techniques are required.

7. ACKNOWLEDGEMENTS

Many thanks to R. Errico, who played a major part in the development of the perturbation scheme. T. Mayer and B. Eaton designed and debugged the verification code which proved to be quite helpful in analyzing the results. Many of the scientific ideas presented here were derived from numerous discussions with J. Tribbia, G. Branstator and M. Blackmon. Finally, I would like to thank the ECMWF for the opportunity to present these results to the international community.

8. REFERENCES

Baumhefner, D.P., 1984: Relationship between present large–scale forecast skill and new estimates of predictability error growth. *Workshop on Prediction of fluid motions.* American Institute of Physics Conference Proceedings #106, 169–180.

Baumhefner, D.P., 1985: Analysis and forecast intercomparisons using the FGGE SOP I data base. *Proceedings of the First National Workshop on the Global Weather Experiment, Vol II, Part 1.* National Academy press, Washington, D.C., **II**, Part I, 228–246.

Baumhefner, D.P., 1988: Experimental 30-day forecasts from several spectral GCM's. *Proceedings of the Twelfth Climate Diagnostics Workshop*, CAC, NMC, NOAA, 5200 Auth Road, WWB, Room 606, Washington, D.C. 20233.

Baumhefner, D.P., J.J. Tribbia and M.L. Blackmon, 1988: The influence of specified sea surface temperature and initial condition uncertainty on Monte Carlo extended range forecast ensembles. *Proceedings of the WMO/WCRPP Workshop on intercomparison of results from atmospheric and ocean GCM's.* ECMWF, May 1988, WMO technical document.

Brankovic, C., S. Tibaldi, U. Cubasch, 1987: Deterministic monthly forecasts at ECMWF. *Research Activities in Atmospheric and Oceanic Modelling* Report #10 WMO/TD–200, 6.25–6.26.

Daley, R. and T. Mayer, 1986: Estimates of global analysis error from the global weather experiment observational network. *Mon. Wea. Rev.*, **114**, 1642–1653.

Epstein, E.S., 1969: Stochastic–dynamic Prediction, *Tellus*, **21**, 739–759.

Errico, R.M. and B.E. Eaton, 1987: Non–linear normal mode initialization of the NCAR CCM. NCAR Technical note (NCAR/TN–303+IA), Climate and Global Dynamics Division, Boulder, CO 80303.

Kistler, R.E., E. Kalnay, M.S. Tracton, 1988: Forecast agreement, persistence, and forecast skill. *AMS Proceedings of the Eighth Conference on Numerical Weather Prediction*, Baltimore MD, 641–646.

Leith, C.E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102 #6**, 409–418.

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

Miyakoda, K., T. Gordon, R. Caverly, W. Stern, J. Sirutis, and W. Bourke, 1983: Simulation of a blocking event in January 77. *Mon. Wea. Rev.*, **111**, 846–849.

Miyakoda, K., J. Sirutis, J. Ploshay, 1986: One Month forecasting experiments without anomaly boundary forcings. *Mon. Wea. Rev.*, **114**, 2363–2401.

Miyakoda, K., J. Sirutis, and T. Knutsen, 1987: Experimental 30–day forecasting at GFDL. *Proceedings of Workshop on predictability in the medium and extended range.* ECMWF, Shinfield park, Reading, Berkshire, RG2 9AX United Kingdom, 17–50.

Molteni, F., U. Cubasch, and S. Tibaldi, 1987: 30 and 60 day forecast experiments with the ECMWF spectral models. *Proceedings of Workshop on predictability in the medium and extended range.* ECMWF, Shinfield park, Reading, Berkshire, RG2 9AX United Kingdom, 51–107.

Murphy, J.M., 1988: The impact of ensemble forecasts on predictability. *Quart. J. Roy. Meteor. Soc.*, **14**, 463–493.

Namias, J., 1968: Long-range weather forecasting — history, current status and outlook. *Bull. Amer. Meteor. Soc.*, **49 #5**, 438–470.

Panel on International Cooperation, committee on Atmospheric Science, NAS, 1966: The feasibility of a global observation and analysis experiment. *Bull. Amer. Meteor. Soc.*, **47 #3**, 200–220.

Tracton, M.S., R.E. Kistler, 1988: Activities in dynamic extended range forecasting at the National Meteorological Center, Washington, D.C. *AMS Proceedings of the Eighth Conference on Numerical Weather Prediction*, Baltimore MD, 715–718.

Tribbia, J.J. and D.P. Baumhefner, 1988: Estimates of the Predictability of low-frequency variability with a spectral general circulation model. *J. Atmos. Sci.*, **45**. (Accepted for October 1988).

U. S. Committee for GARP, Division of Physical Sciences, National Research Council, 1969: Plan for U.S. participation in the Global Atmospheric Research Program. National Academy of Sciences, Washington, D.C., 79 pp.

Williamson, D.L., J.T. Kiehl, V. Ramanathan, R.E. Dickinson, J.J. Hack, 1987: *Description of the NCAR Community Climate Model (CCM1).* NCAR Technical Note, NCAR/TN-285+STR, National Center for Atmospheric Research, Boulder, CO, 112 pp.