The Impact of Climatology and Systematic Errors upon the Skill of DERF Forecasts

TIMOTHY W. BARKER AND JOHN D. HOREL

Department of Meteorology, University of Utah, Salt Lake City, Utah

28 February 1989 and 29 June 1989

ABSTRACT

The average skill of 61 thirty-day forecasts of 500 mb geopotential height from the Dynamic Extended Range Forecast (DERF) experiment are investigated. These forecasts were made using the National Meteorological Center's Medium Range Forecast model and starting from initial analyses during the winter of 1986/87. The impact upon extended range forecast skill of the removal or retention of systematic errors and low-frequency climate variability is studied. If the systematic error is removed a posteriori at each forecast lead time, then the skill of forecasts of time averages may be improved considerably. The magnitude of this improvement is difficult to quantify with forecasts from a single season. Nearly all of the skill at extended range in the DERF experiments arises from the successful forecast of low-frequency fluctuations in the large-scale circulation.

1. Introduction

Over the past few years, the National Meteorological Center (NMC) and other operational weather prediction centers have begun to investigate the feasibility of making extended range numerical predictions. During the winter of 1986/87, NMC produced a large number of experimental 30-day forecasts using the operational Medium Range Forecast (MRF) model. This Dynamical Extended Range Forecasting (DERF) experiment provides a large dataset for investigating the skill of such forecasts initialized with actual data and using a state-of-the-art numerical prediction model. Results from this experiment are summarized by Tracton et al. (1989). In addition, NMC has made this unique and extensive dataset available to the meteorological community for further analysis.

Numerous predictability studies have shown that it is unlikely that individual weather patterns can be predicted more than a few weeks in advance when considered over a large sample of forecasts for a large geographic domain. Many investigators have suggested, however, that postprocessing of extended range numerical forecasts may lead to useful forecast skill at extended range. We were motivated to perform this study on the basis of a preliminary attempt to identify the forecast skill of quasi-stationary regimes within the DERF experiment following the approach of Horel and Roads (1988). The MRF model showed considerable skill in forecasting quasi-stationary periods at lead times of 10–20 days. Further investigation determined that most of the apparent skill resulted from processing steps related to the elimination of systematic errors and the choice of the climatology and was not related to the occurrence of quasi-stationary periods.

We focus here upon the impact of two factors that affect extended range forecast skill: 1) the removal or retention of systematic errors, and 2) the removal or retention of low-frequency climate variability. Our aim is to clarify the magnitude of the differences in forecast skill that can be produced by varying the analysis technique with regard to systematic model errors and base climatologies. We will show that if the systematic error can be removed, then the forecast skill of time averages may be improved substantially. Further, we will demonstrate that nearly all of the skill at extended range in the DERF experiments arises from the successful forecast of low-frequency fluctuations.

2. Data processing

During the DERF experiment, a total of 108 forecasts were made once a day beginning at 0000 UTC 14 December 1986 and continuing until 0000 UTC 31 March 1987. Initialized analyses and forecasts at daily intervals out to 30 days are available at all standard levels for selected variables. We will focus upon the winter season and limit our analysis to the first 61 forecasts of Northern Hemispheric 500 mb geopotential height (north of 25°N), which are initialized between 14 December 1986 and 12 February 1987. Since our long-term winter climatology ends on 14 March, the last 30-day forecast that can be verified is initialized on 12 February. The height data in spherical harmonic form are truncated at rhomboidal 10 and then transformed to a 2.5° × 5.0° latitude/longitude grid for this

Corresponding author address: Dr. John D. Horel, Department of Meteorology, 819 Wm. C. Browning Building, University of Utah, Salt Lake City, UT 84112.

© 1989 American Meteorological Society

Unauthenticated | Downloaded 05/25/24 02:25 AM UTC

analysis. This truncation serves as a filter of small scale features in the height field which are not of explicit interest for extended range forecasts.

Pattern correlations between anomaly maps poleward of 25°N (hereafter referred to as anomaly correlations) are used to measure skill. The anomalies are produced by subtracting the climatological mean height at each grid point from the forecasts and analyses. In the following section, results will be presented in terms of ensemble averages of the anomaly correlations derived from all 61 forecasts. Since the anomaly correlation is restricted to be between -1 and 1, the distribution of correlations is skewed if, on average, the correlations are close to those limits. To correct the skewness of the distribution of correlations, the correlations are transformed using the Fisher-z transformation so that they resemble more closely a normally distributed population. For ease of interpretation, the ensemble averages of the transformed correlations are then transformed back to obtain the ensemble mean anomaly correlations.

The impact upon forecast skill of including or excluding low-frequency variability is investigated by using two datasets of anomalies derived from two distinct climatologies. First, a daily climatology of 500 mb geopotential height is used that is based upon 17 years of NMC operational analyses from 1965 to 1982 (Horel 1985). Deviations from this climatology include low-frequency variations which reflect how analyses during the 1986/87 winter differ from other winters. We will refer to these anomalies as those which include low-frequency variability. Second, in a manner analogous to Roads (1989), a climatology is used that approximates the average height at each grid point during the winter of 1986/87 in a least-squares sense:

$$\Phi(x, t) = a(x) + b(x)t + c(x)t^2$$

where t is the day within the DERF experiment and the coefficients $a(x)$, $b(x)$, and $c(x)$ are calculated by a least-squares fit to the first 91 DERF analyses at each grid point x. Deviations from this climatology lack interannual and low-frequency intraseasonal variations as a result of the quadratic fit. We will refer to these anomalies as those which have had low-frequency variability removed. It should be remarked that this approach for determining the climatology is not feasible operationally since the analyses during the entire winter must be known in advance.

Figure 1 shows the hemispheric average (poleward of 25°N) of 500 mb geopotential height determined from these two climatologies as a function of time dur-

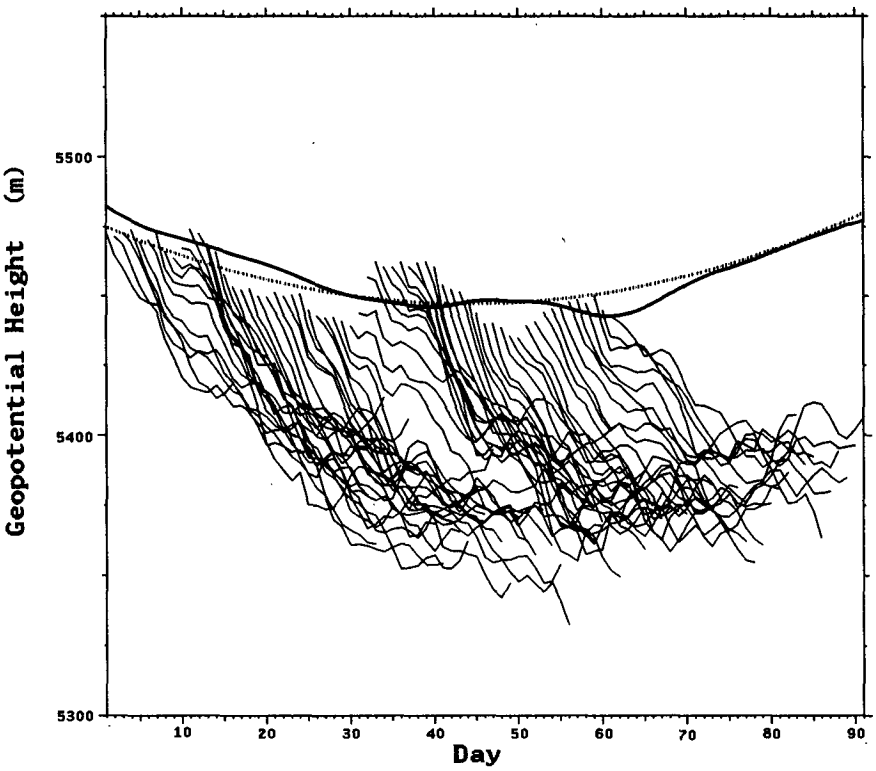

FIG. 1. The 500 mb height averaged over the domain poleward of 25°N during each of the first 61 thirty-day forecasts as a function of verifying day. The heavy solid line denotes the 17-year climatology while the heavy dashed line denotes the climatology derived from a least-squares fit to the first 91 DERF analyses.

Unauthenticated | Downloaded 05/25/24 02:25 AM UTC

ing the winter season. Overall, the differences between the hemispherically averaged heights determined from the two climatologies are nearly insignificant; however, in certain geographical areas and during parts of the winter season the differences between these two climatologies are quite large. For example, over the Gulf of Alaska during January, the average differences are larger than 150 m (not shown).

As noted by Barnett and Graham (1988), Tracton et al. (1989), and Roads (1989), the MRF model has substantial systematic (time average) errors during the DERF period which are inhomogeneous in space and time. The overwhelming feature of the model's bias, however, is to lower the hemispherically averaged 500 mb height, especially for forecasts at long lead times. In Fig. 1, the hemispherically averaged 500 mb height is shown for each forecast day of the 61 forecasts in our sample. During each of the 61 forecasts, the hemispherically averaged 500 mb height drops roughly 100 m in 30 days. The relatively consistent negative slope of the curves in Fig. 1 during the first 10 days of each model integration indicates that the magnitude of the hemispherically averaged bias is growing at nearly a constant rate. For longer forecasts, the changes in the model's "climate drift" from day to day appear more random.

We investigate the impact of systematic error by performing our analysis on datasets in which estimates of the systematic error have been removed or retained. In a manner analogous to Miyakoda et al. (1986), Roads (1989), Tracton et al. (1989), and Chen (1989), the systematic error is defined here as the average difference over all 108 cases between the forecast and verifying height at each grid point and each forecast lead time. For the DERF experiment, this systematic error is equivalent to the difference between the model's forecast of the seasonally averaged height anomalies and the observed seasonally averaged height anomalies. The systematic error of the 30 day forecasts is shown in Fig. 2c. The largest systematic errors are found at high latitudes over the Bering Sea and over Iceland with smaller errors over Canada.

The systematic error defined in this manner is only an estimate of the "true" systematic error, which could be determined if extended-range forecasts were made for many years with the same numerical model. When forecasts are available from one season only, it is difficult to separate the true systematic error from the interannual variability of the forecasts and analyses. The observed interannual variations determined from the 500 mb height analyses during days 31–91 of the DERF experiment are shown in Fig. 2a. Below normal heights are evident over the central Pacific Ocean and above normal heights are found over Canada. The interannual variations determined from 30 day forecasts of 500 mb height made from initial conditions on days 1–61 (i.e., verifying on days 31–91) are shown in Fig. 2b. Although there is a clear tendency for the seasonally averaged anomalies in Fig. 2b to be negative, it is impossible to determine what part of this anomaly field is due to the model's ability to forecast the observed interannual variation in Fig. 2a versus the model's tendency to lower the heights. For example, the small positive height anomalies in Fig. 2b over Canada may arise from a combination of negative systematic errors of the model being compensated by a successful forecast of the positive height anomalies observed in that region during this season.

By making assumptions about the model's ability to reproduce the observed anomaly field for this winter, we can estimate the impact of removing the true systematic error. The most optimistic assumption is that the model is able to reproduce exactly the seasonal anomaly field in Fig. 2a for all forecast lead times. This assumption is implicit in the definition of systematic error used in the studies of Roads (1989), Tracton et al. (1989), and Chen (1989). This optimistic assumption is likely to be valid for forecasts of short duration, but it is unlikely that extended range forecasts are capable of reproducing the interannual variations exactly. Hence, as a more pessimistic alternative, we assume that the model's forecast of the amplitude of the seasonal anomaly field in Fig. 2a decreases linearly as a function of forecast duration such that the amplitude reaches zero at 30 days. Although both assumptions are quite arbitrary and simplistic, it is clear by comparing Figs. 2a and 2c that the systematic error is not dominated by the observed interannual variations. Thus, our assumptions regarding the model's forecast of interannual variations are of secondary importance. Further, when both interannual and low-frequency intraseasonal variations are removed from the data by using the least-squares climatology, these different assumptions have little impact on our measure of model skill.

To summarize, we perform our analysis for the four combinations of including or excluding the low-frequency signal and the systematic error. The cases which compare anomalies relative to the 17 year climatology, and thus include interannual and low-frequency intraseasonal variations, are labeled in subsequent figures with an I. For the cases which compare anomalies relative to the least-squares climatology, and, hence, exclude low-frequency variations, the I is omitted from the label. Likewise, cases which include the systematic error are labeled with an S and those cases which have had the systematic error removed omit the S label. In addition, forecasts from the MRF model are labeled with an M and compared to persistence forecasts, labeled with a P. Tracton et al. (1989) perform many of their calculations with a long-term climatology and without removing any systematic errors. Thus, their analysis essentially corresponds to our analysis labeled MIS, i.e., model forecasts which include the low-frequency signal and systematic errors. The MIS case can be considered to be the reference case to which a variety

Unauthenticated | Downloaded 05/25/24 02:25 AM UTC

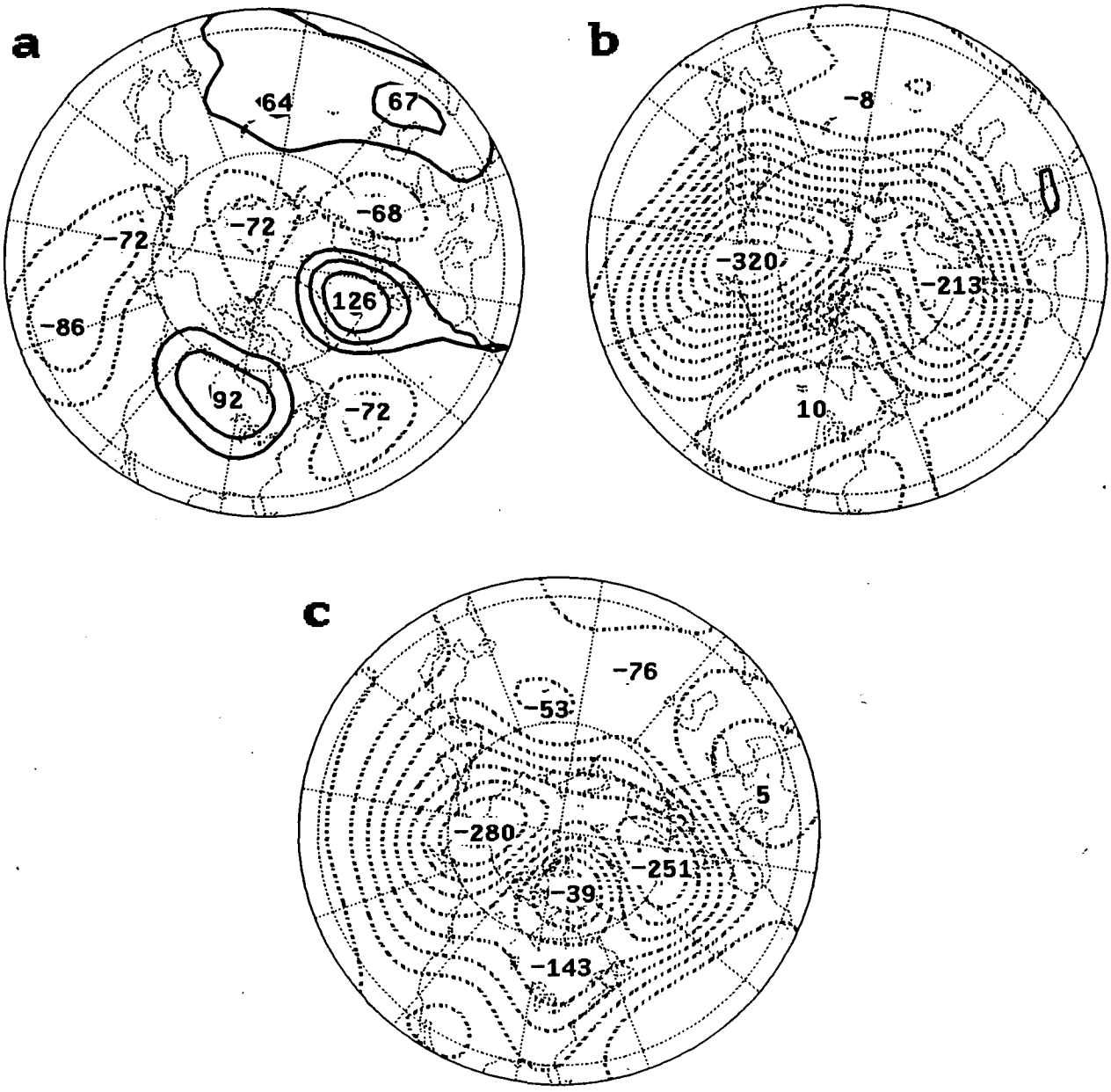

FIG. 2. (a) Sixty-one-day average of observed 500 mb height anomalies from days 31–91 of the DERF experiment. (b) Average of the 500 mb height anomalies from 30-day forecasts made from initial conditions on days 1–61 and verifying on days 31–91 of the DERF experiment. (c) Systematic error (forecast minus analysis) of 30-day forecasts based upon the entire 108-day forecast sample. The contour interval is 30 m with the positive (negative) contours solid (dashed) and the zero contour omitted.

of other postprocessing procedures have been applied. Roads (1989) performs his calculations with a least-squares climatology and with the systematic error removed, corresponding to our analysis labeled M. As described above, the ambiguity inherent in the removal of systematic error for the MI case requires special treatment. We assume an upper bound for forecast skill to be our case where the observed interannual signal is exactly duplicated by the model at all lead times. We assume a lower bound to be our case where the amplitude of the observed interannual variations diminishes linearly as a function of forecast length.

3. Results

First, the effects of the presence or absence of low-frequency variability upon forecast skill are investigated as a function of forecast length. Figure 3 shows the anomaly correlations of model and persistence forecasts with the corresponding verifying analysis when aver-

Unauthenticated | Downloaded 05/25/24 02:25 AM UTC

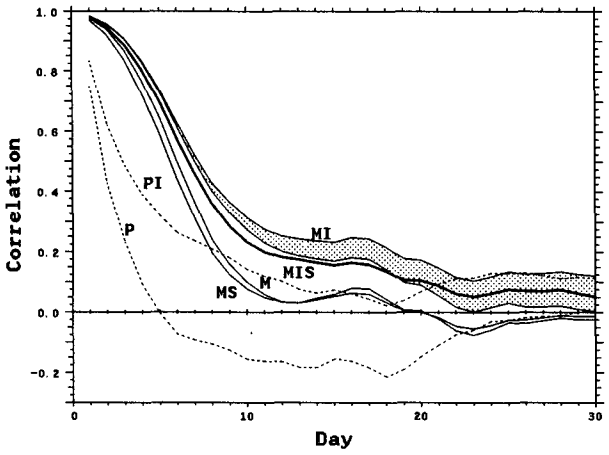

FIG. 3. Ensemble mean over 61 cases of the pattern correlations between forecasts of 500 mb height anomalies and the verifying height anomalies as a function of forecast lead time. Curves are labeled according to the following convention: M denotes model forecasts, P denotes persistence forecasts, I denotes forecasts which include the low-frequency signal, and S denotes forecasts which include the systematic error. The MIS curve is denoted by a heavy line for reference. The shaded band denotes a range of possible values for the MI curve. See text for further details.

aged over the 61 forecasts. The time at which the anomaly correlation drops below 0.6 is often used as an indication of the limit of useful skill when considering synoptic scale features over an entire hemisphere (for example, see Hollingsworth et al. 1980). For persistence forecasts, useful skill is evident for only 36 hours if the low-frequency variability is removed (curve P) and for 2 days if the low-frequency variability is retained (curve PI). For the model forecasts, if the low-frequency variability is absent, then the anomaly correlation averaged over this sample of 61 forecasts drops below 0.6 at about 5 days (curves M and MS). If the low-frequency variability is retained (curves MI and MIS) then the ensemble mean correlation remains above 0.6 for slightly longer than 6 days. Thus, including the low-frequency signal increases the mean time for useful skill by about one day during this particular winter.

After approximately 12–15 days, all of the curves in Fig. 3 appear to reach asymptotic limits. Persistence forecasts in which the low-frequency variability is removed have negative skill until roughly 25 days, after which the anomaly correlation remains near zero. Persistence forecasts in which the low-frequency variability is retained have small positive (0.1–0.2) correlations out to 30 days. For the MRF forecasts with the low-frequency variability removed (curves M and MS), the anomaly correlations asymptote at extended range to zero. If the low-frequency variability is retained (curve MIS), then the skill beyond lead times of 20 days is comparable to that of persistence forecasts. Thus, the MRF model has skill (albeit small) at making extended forecasts of weather patterns which arises from the presence of low-frequency climate variability and which is of the same magnitude as persistence forecasts.

The impact of systematic errors upon forecast skill of instantaneous forecasts is small (compare curves M with MS and shaded area associated with MI to the curve of MIS in Fig. 3). If the low-frequency variability is removed (curves M and MS), then there is virtually no difference between the ensemble mean anomaly correlations with (curve MS) or without (curve M) systematic error. The difference between forecasts with and without the low-frequency signal when the systematic error is retained (curves MIS and MS) suggests that the model does forecast some low-frequency variability at extended range, although the amount of skill in these forecasts is small.

If the low-frequency variability is retained, the effect of removing systematic error depends upon the assumptions made in defining the systematic error. Our pessimistic assumption is that the model forecasts no interannual variations by 30 days. Thus, the skill of the forecasts made with this assumption is near zero for extended range forecasts (lower bound of MI curve in Fig. 3). Likewise, our optimistic assumption is that the model perfectly reproduces the observed interannual variations, which increases the skill of extended range forecasts to that of persistence forecasts (upper bound of MI curve).

Although the skill of forecasts at long lead times is small, there is considerably more skill in making forecasts of time averages. In Fig. 4 we show the forecast and analysis maps for a particular 30 day average which was well forecasted by the model. Figures 4a and 4b show the average analysis for the 30-day period beginning 0000 UTC 23 December 1986 with the low-frequency signal removed and retained respectively. As should be expected, the 30-day average of the height anomalies has relatively small amplitude relative to the least-squares fit climatology (Fig. 4a). Relative to the 17-year climatology, the observed 30-day average of the height anomalies is similar to the 61-day average shown in Fig. 2a and is characterized by below normal heights centered over the Aleutians and Greenland with above normal heights over Canada and over the eastern Atlantic (Fig. 4b). Figures 4c and 4d show the 1–30 day forecasts made by the MRF model verifying during this period with the systematic error retained in both cases. In both figures, the 500 mb height anomaly field is dominated by below normal heights around the hemisphere as a result of the climate drift of the model. With the low-frequency signal removed (Fig. 4c), the forecast bears little resemblance to the verifying 30-day mean and the anomaly correlation between these two maps is only 0.01. With the low-frequency signal retained (Fig. 4d), there is considerable agreement in the locations of the relative minima and maxima, e.g., negative (positive) height anomalies near the Aleutians (western United States); however, the model's systematic error has weakened substantially the positive height

Unauthenticated | Downloaded 05/25/24 02:25 AM UTC

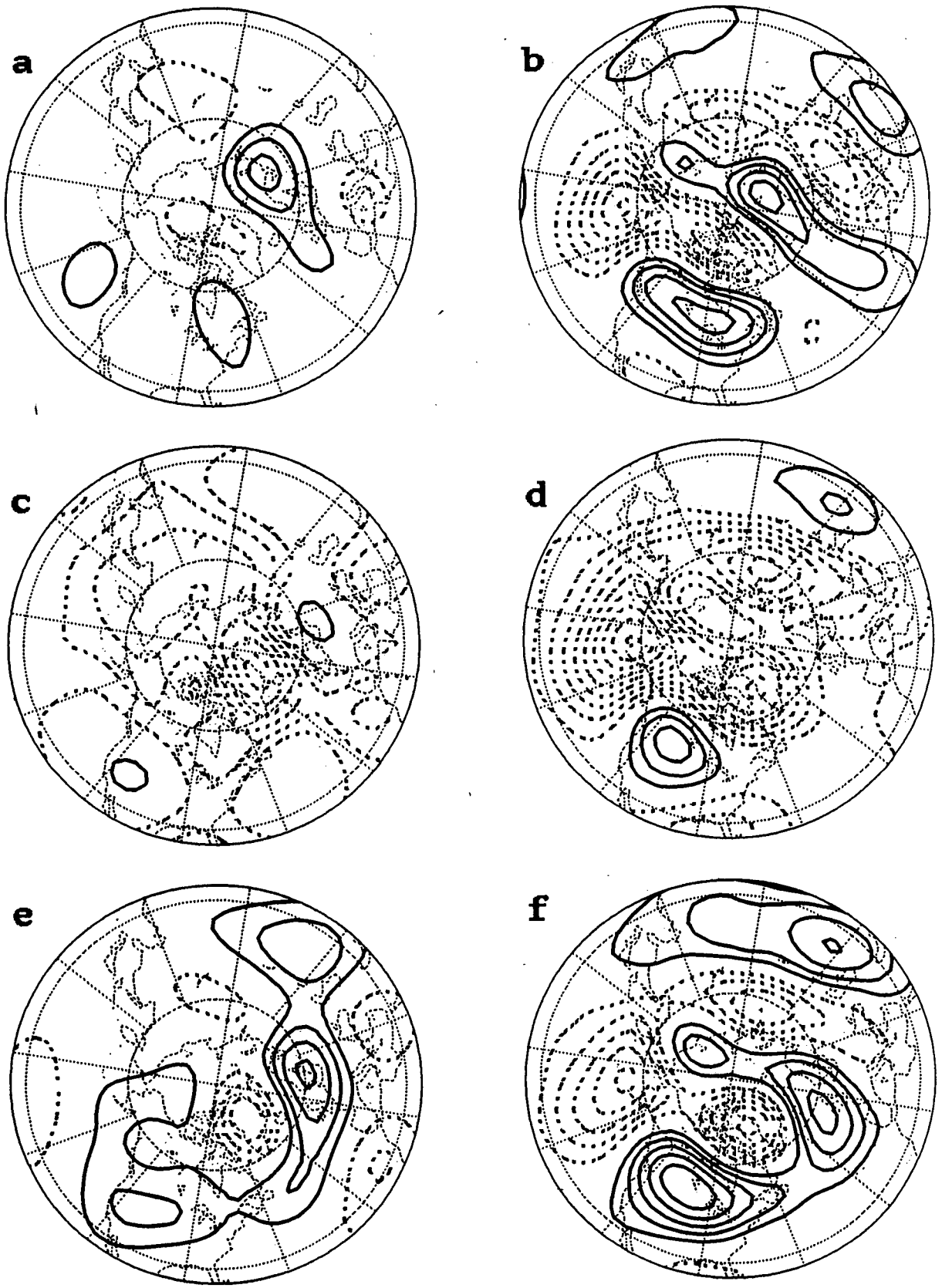

FIG. 4. Spatial maps of 500 mb height anomalies averaged over the 30-day period beginning 0000 UTC 23 December 1986 (day 9 of the DERF experiment). The contour interval is 30 m with positive (negative) anomalies denoted by solid (dashed) contours; the zero line is omitted. (a) Verifying 1–30 day mean with the low-frequency signal removed. (b) Verifying 1–30 day mean with the low-frequency signal included. (c) 1–30 day forecast with the low-frequency signal removed and systematic errors included. (d) 1–30 day forecast with the low-frequency signal retained and systematic errors included. (e) 1–30 day forecast with the low-frequency signal and systematic error removed. (f) 1–30 day forecast with only the systematic error removed.

Unauthenticated | Downloaded 05/25/24 02:25 AM UTC

anomalies observed over Iceland. The anomaly correlation between the maps in Figs. 4b and 4d is 0.57, which indicates that the forecast exhibits considerable skill. However, it should be remarked that much of the skill of this particular forecast arises from the successful prediction of the first 10 days of the forecast period.

Also shown in Fig. 4 are the 30-day forecasts with the systematic error removed (using the optimistic assumption described in the previous section) and the low-frequency variability excluded (Fig. 4e) and included (Fig. 4f). With the low-frequency signal removed (Fig. 4e), the forecast is not successful and the correlation between the anomaly maps in Figs. 4a and 4c is 0.38. With the low-frequency signal retained (Fig. 4f), there is considerable agreement in amplitude and phase between the verifying analysis and forecast. The correlation between the anomaly maps in Figs. 4b and 4d is 0.81, which further indicates the close similarity between these two fields.

In order to summarize the skill of the other 60 forecasts of 1–30 day averages and the skill of forecasts of time averages of other durations, Fig. 5 shows the ensemble mean correlation between forecasts from the MRF model of time averages lasting from $1 - n$ days and their verifying time averaged analyses, where n varies from 1 to 30. For comparison, the anomaly correlation between the initial analysis and the subsequent time average is used as a measure of the skill of a persistence forecast. Persistence forecasts of time averages of anomalies without low-frequency variability (curve P in Fig. 5) have useful skill for averages of height anomalies out to only 2 days. If the low-frequency variability is retained (curve PI), then persistence forecasts of subsequent time averages remain skillful for 6 days, and then asymptote for longer time averages to a pattern correlation of 0.37. Thus, during this winter, persistence forecasts based solely upon the initial analysis demonstrated considerable skill at predicting subsequent time averages for long averaging periods.

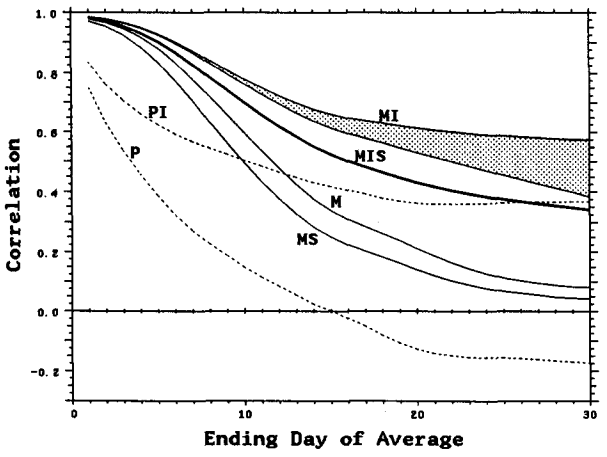

FIG. 5. Ensemble mean over 61 cases of the pattern correlations between forecasts of time averages and the verifying time averages. The ordinate denotes the length of the averaging period, i.e., 30 days denotes an average over the period 1–30 days. Persistence forecasts are made using the initial analysis only. Labeling as in Fig. 3.

The MRF model with both systematic errors and low-frequency variability removed reaches the 0.60 criterion for useful skill for time averages at roughly 10 days (curve M in Fig. 5). Including the systematic errors reduces the useful skill of time averages by 1½ days (compare curves M and MS). If both the systematic and low-frequency variability are included, then forecasts of time averages show skill out to 12–13 days (curve MIS). The skill of forecasts of longer time averages approaches that of persistence. For comparison, our anomaly correlation of 0.35 for 1–30 day averages is nearly identical to that of 0.39 reported by Tracton et al. (1989). The large differences in forecast skill between the MS and MIS curves indicate that the model does have skill in predicting time averages as a result of low-frequency variations in the circulation.

If the systematic error is removed while low-frequency variability is retained (shaded area in Fig. 5), then the ensemble mean anomaly correlations can be as high as 0.58 for time averages as long as 1–30 days if the forecasts of interannual variations are assumed to be perfect. More pessimistic estimates of the model's skill at predicting interannual variability lead to forecast skill of time averages comparable to that of persistence. Thus, the MRF model may exhibit considerable skill at making extended range predictions during this particular winter, if the true systematic errors were known and removed. Of course, it should be remembered that much of the skill in predicting long time averages arises from the successful prediction of the first few days of those periods.

4. Summary and discussion

We have examined the average skill of a large number of 30-day forecasts of 500 mb geopotential height from the DERF experiment. We have restricted our analysis to be a determination of the impact of the removal or retention of low-frequency variability and systematic model errors. Removing the low-frequency signal from the forecasts and verifying analyses removes nearly all of the skill of instantaneous forecasts at long lead times. Removing the low-frequency signal has a similar impact upon the skill of forecasts of long time averages. The skill of instantaneous forecasts is not affected greatly by the removal or retention of systematic errors. The forecast skill of time averages that include low-frequency variability may be enhanced if the systematic errors are removed, but the magnitude of this improvement is difficult to determine from this dataset. Even with the uncertainty inherent in our study, the range of forecast skill for 1–30 day time averages with the low-frequency signal included and systematic errors removed may be high enough to be considered useful

Unauthenticated | Downloaded 05/25/24 02:25 AM UTC

in a practical sense. Thus, the model appears to have some useful skill at predicting deviations from the long-term climatology (at least in terms of whether the next month will have 500 mb heights higher or lower than normal).

The large differences in the ensemble mean correlations that are obtained in our study by simply varying climatology and treatment of systematic errors should promote some caution when evaluating other more elaborate postprocessing techniques. Comparing the results from studies which use different postprocessing approaches is often difficult. Our study helps to explain some of the differences evident between the analyses of Tracton et al. (1989) and Roads (1989). The skill of extended range forecasts, according to Roads, tended to be lower than that reported by Tracton et al. Much of this discrepancy appears to be due to Roads's use of a least-squares climatology that removes low-frequency variations. In addition, the results of our analysis indicate the possibility of more skill than that reported by Tracton et al. (1989) for the case when systematic errors are removed. The discrepancies between our analysis and that of Tracton et al. arise apparently as a result of differences in analysis approaches, including differences in the climatologies used and forecast sample size.

To further highlight that measures of forecast skill are sensitive to the analysis approach, we also performed all of the calculations using standardized anomalies, i.e., the anomalies were divided by the daily standard deviation of the 500 mb height field at each grid point. The use of standardized anomalies for calculating pattern correlations gives less weight to the storm track regions and more weight to the subtropics. The ensemble mean pattern correlations based upon standardized anomalies are roughly 0.05–0.10 higher compared to those shown in Fig. 5.

The possible large increase in skill of extended range forecasts produced by removing the systematic error is encouraging. Since changes to operational forecast models such as the MRF are currently made every few months, it is unlikely that the true systematic error of such a model can be determined; however, the goal of these model improvements is, in part, to remove systematic errors from the model forecasts. Thus, further improvements to the MRF model should allow the skillful signals, which are already present in the model, to become more evident without elaborate postprocessing.

Acknowledgments. We would like to thank J. Roads of the Scripps Institution of Oceanography for constructive comments on an earlier version of this manuscript and for providing the DERF data in a convenient format. We would also like to thank S. Tracton, Climate Analysis Center, and the reviewers for their useful comments. This research has been supported by NSF Grant ATM87-15360 and a NSF Graduate Fellowship awarded to T. W. Barker.

REFERENCES

Barnett, T. P., and N. E. Graham, 1988: Analysis of the dynamical extended range forecast experiment: A progress report. *Proc. of the Twelfth Annual Climate Diagnostics Workshop,* 401–411. [Available from the National Technical Information Service, U.S. Dept. of Commerce, Sills Bldg, 5285 Port Royal Rd., Springfield, VA 22161.]

Chen, W. Y., 1989: Estimate of dynamical predictability from NMC DERF experiments. *Mon. Wea. Rev.,* **117,** 1227–1236.

Hollingsworth, A., K. Arpe, M. Tiedtke, M. Capaldo and H. Savijarvi, 1980: The performance of a medium-range forecast model in winter—impact of physical parameterizations. *Mon. Wea. Rev.,* **108,** 1736–1773.

Horel, J. D., 1985: The persistence of the 500 mb height filed furing Northern Hemisphere winter. *Mon. Wea. Rev.,* **113,** 2030–2042.

——, and J. O. Roads, 1988: Sensitivity of regional predictability to flow characteristics. *J. Geophys. Res.,* **93,** 11 005–11 014.

Miyakoda, K., J. Sirutis and J. Ploshay, 1986: One-month forecast experiments—without anomaly boundary forcings. *Mon. Wea. Rev.,* **114,** 2363–2401.

Roads, J. O., and T. P. Barnett, 1984: Forecasts of the 500 mb height using a dynamically oriented statistical model. *Mon. Wea. Rev.,* **112,** 1354–1369.

——, 1989: Dynamical extended range forecasts of the lower tropospheric thickness. *Mon. Wea. Rev.,* **117,** 3–28.

Tracton, M. S., 1989: Application of dynamic extended range forecasting (DERF) to the monthly forecast problem. *Mon. Wea. Rev.,* **117,** 1606–1637.

Unauthenticated | Downloaded 05/25/24 02:25 AM UTC