Frequency Dependence in Forecast Skill

H. M. VAN DEN DOOL AND SURANJANA SAHA

Cooperative Institute for Climate Studies, Department of Meteorology, University of Maryland, College Park, Maryland

(Manuscript received 31 March 1989, in final form 17 August 1989)

ABSTRACT

A method is proposed to calculate measures of forecast skill for high, medium and low temporal frequency variations in the atmosphere. This method is applied to a series of 128 consecutive 1 to 10-day forecasts produced at NMC with their operational global medium-range-forecast model during 1 May–5 September 1988. It is found that over this period, more than 50% of the variance in observed 500 mb height fields is found at periods of 18 days or longer. The intuitive notion that the predictability time of a phenomenon should be proportional to its lifetime is found to be qualitatively correct; i.e., the low frequencies are predicted (at a given skill level) over a longer time than high frequencies. However, the current prediction skill in low frequencies is far below its potential if one assumes that for any frequency the predictability time scale ought to be equal to the lifetime scale. In the high frequencies, however, the current prediction skill has already reached its potential; i.e., cyclones are being predicted over a time comparable to their lifetime; i.e. 3 to 4 days. We offer some speculations as to why the low frequency variations in the atmosphere are so poorly predicted by our current state-of-the-art models. The conclusions are tested, and found to hold up, on a more recent dataset covering 10 December 1988–16 April 1989.

1. Introduction

In an intuitive sense, the predictability of a specific atmospheric phenomenon has to be proportional to its own lifetime. Nobody tries to forecast the timing of tomorrow's cumulus clouds, or the passage of an extratropical cyclone two months ahead of time. Apparently we restrict prediction of certain phenomena to forecast lead times less than, or at most comparable to, the lifetime of these events. This strategy allows us, in practice, to first observe the early stages of a particular development before attempting to forecast the rest of its life cycle. The proportionality of lifetime and predictability time plays an important role in the theory of predictability of two-dimensional flows (Lorenz 1969; Lilly 1973).

If there is any truth in the above assumption one must expect the predictability of short-lived events to be short, while long-lived events should be predictable out to much larger lead times. Because there is considerable energy at low frequencies in 500 mb height variations (Blackmon et al. 1984) there is, in that fact alone, some basis for hope in long-range weather prediction (LRWP).

Every day operational centers such as the National Meteorological Center (NMC) in Washington, D.C., produce operational numerical weather predictions (NWP) of large-scale fields such as 500 mb height, out to 10 days ahead. The question is: How well are the high, medium and low frequencies, as observed in nature, forecast by these state-of-the-art models? Is prediction skill (a practical measure, subject to improvement) a function of lifetime in the way we envision predictability (a property of the fluid) to be a function of lifetime?

Surprisingly, to our knowledge, that question has never been answered directly. This can be explained primarily from the fact that it is not evident how one filters forecasts in time. Also, at one particular time one cannot determine the state of the atmosphere at a given temporal frequency. It is only after the fact that one can analyze the observations in terms of variations at certain frequencies and ask the question: How well were these variations predicted by a particular NWP model? Readily available standard verification scores include root-mean-square (rms) errors and anomaly correlations (AC) as a function of lead time. These and related scores are often broken down for several distinct spatial-wavenumber bands (Boer 1984; Savijarvi 1984; Dalcher and Kalnay 1987) and may, by relating space and time scales, indirectly answer our question. However, the relation between time and space scales is nontrivial (Blackmon 1976; Straus and Shukla 1981). Therefore we offer here a direct approach by applying a filter in the time domain before calculating various verification scores.

Some readers may believe that a forecast out to 10 days deals with time-scales less than 10 days. But that

Corresponding author address: Dr. H. M. van den Dool, Cooperative Institute for Climate Studies, Department of Meteorology, University of Maryland, College Park, MD 20742.

is not the case as seen from the following example. One can make a model for the annual cycle (and nothing else) and integrate that model 10 days into the future and ask the question: How well did we forecast this phenomenon with a 365 day time scale? Similarly we can ask: How well does NWP predict in a given low-frequency band 1, 2, . . . , 10 days ahead? Of course a direct Fourier transform of ten data points in time would erroneously fold the low-frequency variance into the resolved periods (i.e., 2–10 days).

There is a class of literature in which the prediction skill (Miyakoda et al. 1986) or predictability (Shukla 1981; Tribbia and Baumhefner 1988) of monthly means is discussed. This monthly mean is taken over daily forecasts (day 1–day 30) generated by NWP in an attempt to isolate the "low frequency variations." However, a monthly mean is a crude filter, not only for its spectral properties but even more because it averages very good (day 1 . . .) with very poor (. . . day 30) forecasts, thus yielding a product that, skillwise, has a very short effective lead time. [Some of the advantages and disadvantages of averaging daily forecasts into a time-mean forecast have been listed in Van den Dool (1985).] We will describe a time filtering technique that allows us to study skill of forecasts of low (as well as medium or high) frequencies as a function of forecast lead time (day by day).

In section 2 we discuss data and analysis—the time filter in particular. Section 3 features the results while some possible implications and speculations are offered in section 4.

2. Data, analysis and time filtering

We study a set of 128 contiguous 10 day forecasts made at NMC once daily at 00Z, by the operational medium range forecast (MRF) model. The period covers 1 May 1988 to 5 September 1988. The fields used in this study are 1000, 500 and 250 mb height; the spatial coverage is global. The MRF model is a global spectral model truncated triangularly at zonal wavenumber 80 (T80). In our analysis we use only T20 data as it contains the bulk of the height variance.

Before discussing the problem of time filtering, it is necessary to first define the anomaly correlation (AC) as

$$\mathrm{AC} = \frac{\sum_t \sum_{n,m} F'A'}{\{\sum_t \sum_{n,m} F'^2 \sum_t \sum_{n,m} A'^2\}^{1/2}} \qquad (1)$$

where we sum over time t, and over two-dimensional (n) and zonal wavenumber (m). Here F' is the forecast anomaly: forecast (F) − climatology (C) and A' is the observed anomaly: initialized analysis (A) − C. The monthly climatology (based on 1978–85) was interpolated from the nearest two calendar months to the date (t) in question. Note that in (1) we sum in spectral space while more often the AC is obtained by summing in gridpoint space with appropriate weighting. Abbreviating covariance and standard deviation to "cov" and "sd," respectively, we can write symbolically:

$$\mathrm{AC} = \frac{\mathrm{cov}_{FA}}{\mathrm{sd}_F \cdot \mathrm{sd}_A}. \qquad (1a)$$

There are probably two basic options for time-filtering forecasts. The first is to take each 1 to M day forecast separately and apply a time filter to the M elements. [Averaging over a month falls in this category (Miyakoda et al. 1986).] Often M (10 in our study) is too small for any meaningful time filtering. Fourier analysis aliases/folds the high/low frequencies into the five resolved frequencies, and a numerical filter cannot be applied on just ten elements. An additional problem is the inhomogeneity of the time series due to climate drift.

The second option, followed here and detailed below, is to place 128 N-day forecasts ($N = 0, 1, \cdots, 10$) in time sequence. These 128 N-day forecasts, verifying on successive days, constitute, in the second option, a time series to which the time filtering can be applied (see details below). It is the length of the time series (128 in our case) that allows us to study high, low and medium frequencies at a given lead time N. This yields a detailed answer to the question about forecast skill in a given frequency band as a function of lead time. Even for $N = 1$, forecasts out to only 1-day, it is possible to calculate the forecast skill in the low frequencies. Contrast this to option 1 where the only way to address low frequencies is to increase the lead time.

One can argue whether a set of consecutive forecasts is a time series. Some ambiguity could exist about the notion frequency, as seen from the following extreme example. Suppose a model predicts, every day, pure persistence (i.e., frequency zero in a given model run). A Fourier analysis of such forecasts, in the manner proposed here, will yield a full spectrum nevertheless. Such ambiguities probably do not exist in forecasts made by state-of-the art models. Another difficulty to consider is serial decorrelation. For large N, forecasts decorrelate not only from the observations but also from each other; i.e., a time series of 8 day forecasts is more random in time than the verifying observations. This leads to spurious transfer of variance towards the high frequencies, and hence sd_F at a given frequency may no longer be trusted. However, because sd_F, for each frequency band, should ideally be equal to sd_A we replaced (1a) by a pseudoanomaly correlation AC* given by

$$\mathrm{AC}^* = \frac{\mathrm{cov}_{FA}}{\mathrm{sd}_A^2}. \qquad (1b)$$

and hence the verification depends only on the covariance.

We will follow the second option. By taking advantage of some properties (detailed below) of the anomaly correlation it turns out that time filtering the observations only is enough for our goals. Hence we decided to not time filter the forecasts at all. The observations [i.e., the anomaly time series $A'(t)$] are time filtered as follows. For each (n, m) within the T20 truncation, a time series is constructed such that:

$$A'(t) = \mathrm{Re}(A'_r(t)) + \mathrm{Im}(A'_i(t))$$

where (t = 1,128) and real and imaginary refers to the spatial spherical harmonics decomposition. These series were then subjected to a discrete Fourier transform where

$$A'_r(t) = \sum_{k=0}^{64} (a_k \cos(kt) + b_k \sin(kt)) \qquad (2)$$

and similarly (and independently) for $A'_i(t)$. Filtered time series were then obtained by transforming back after zeroing out certain (a_k, b_k) as desired. We constructed three broad frequency bands named high, medium and low, i.e., high: periods of 2–6 days (k = 22–64), medium: 6–16 days (k = 9–21) and low: 18 to 128 days (k = 0–8). A fourth band named "all" contains all frequencies, i.e., the unfiltered data. The high-frequency band coincides very nearly with the frequency range referred to as bandpass filter (Blackmon 1976; Wallace et al. 1988) used to study transient eddies and to define storm tracks. The low-frequency band is obviously of prime interest for long-range weather prediction (LRWP).

Time filtering the observations will be enough for some of the vertification statistics. For example, the term in (1a) that describes the decrease of AC with forecast lead time is cov_{FA}, while sd_F and sd_A are (or ought to be) nearly independent of forecast lead time. Because of temporal orthogonality [using (2)] and summing over all time [as in (1)], the product $F'A'$ is nonzero only for those frequencies k that have nonzero amplitude (a_k, $b_k \neq 0$ in (2)) in both the $A'(t)$ and $F'(t)$ time series. Hence filtering F', is, when studying cov_{FA}, done implicitly by filtering A' alone. Note that this method does not work for rms errors because explicitly filtered $F'(t)$ would need to be available for that purpose, and therefore we will not discuss rms verification statistics in this paper. Finally, we point out that even though the power spectrum of the F' time series is distorted (as a result of which sd_F cannot reliably calculated) the estimate of cov_{FA} is not biased. Energy in the F' series that has been transferred to the wrong (= higher) frequencies is not expected to covary with the observations at these frequencies.

Of course Eq. (2) is not a perfect filter. The time series is not periodic, for example. A digital filter [example: Wallace et al. 1988, Table 1) has better characteristics in this regard but would yield filtered data only in the interior portion of the record of 128. Also our (or any other) definition of high, medium and low is somewhat arbitrary. But with only broad band resolution sought, Eq. (2) will be a good enough spectral transform to distinguish prediction skill in low, medium and high frequencies.

3. Results

a. *Unfiltered statistics*

The AC as a function of lead time for 500 mb heights during the 128 day period chosen is shown in Fig. 1. On the global domain AC decreases to 0.6 in 4.2 days. At day 5 the AC is clearly positive but from Fig. 1 alone we cannot tell whether we predict (with partial success) the second generation cyclones or the early stages of low frequency events.

It is worthwhile to discuss, on the basis of Fig. 1, how we define the time over which we can (currently) predict the atmosphere. At lead times between 1 and 10 days the forecasts are neither perfect (AC = 1) nor worthless (AC = 0). We therefore propose to use an integral time scale. In the spirit of Leith (1973) who sought to calculate the time between effectively independent data points from the autocorrelation we chose here the e-folding time (AC = 0.37) as the integral time over which we can predict the atmosphere. This is appropriate as long as AC decreases exponentially (actually the decay is very slightly steeper than exponential). For a red noise process the integral time scale is uniquely determined by the autocorrelation at one lag only. Here we replace the autocorrelation by the AC, which is close to an autocorrelation function fol-

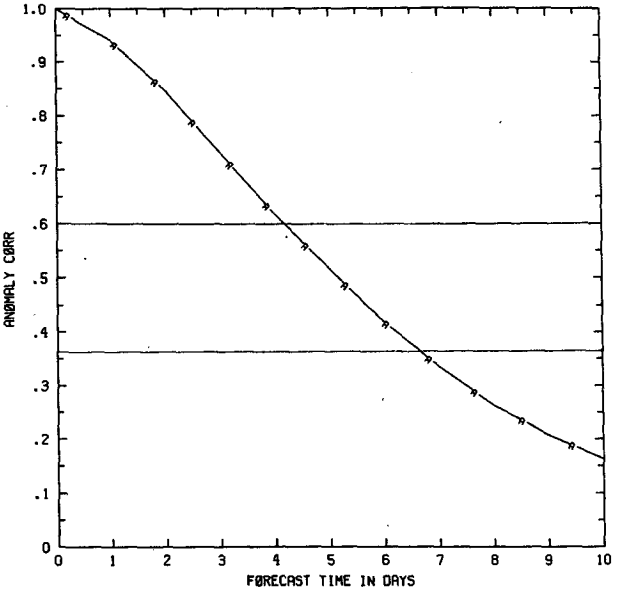

FIG. 1. The anomaly correlation [Eq. (1) in text] for global 500 mb height fields as a function of forecast lead time. The period is 1 May 1988 through 5 September 1988. Horizontal lines are drawn at anomaly correlation scores of 0.6 and 0.37.

lowing the fluid's motion. (We do not propose the *e*-folding as a practical limit of predictability.) Using these notions the prediction time is 6.7 days for all frequencies together.

Another statistic, which is less often published, is the AC as a function of spatial scale, shown here in Figs. 2a and 2b. In order to construct Fig. 2a we first summed the three terms in Eq. (1) over all n for each m before applying the equation to yield AC as a function of m, AC(m). Similarly Fig. 2b was obtained by summing over all m for each n. Shown are AC(m) and AC(n) for lead times of 1, 3, 5, 7 and 10 days. There is obviously a loss of forecast skill at all scales but this loss is clearly dependent on spatial scale. Very soon into the forecast the spatially short waves lose prediction skill, and by day 5 only m = 1 to 4 and n = 1 to 8 surpass the 0.6 AC point. The results for AC(n) agree broadly speaking with those by Boer (1984) in his Fig. 3. We also note the bias-type error in n = 0 (global mean mode; Saha and Alpert 1988) and m = 0 (the zonal mean modes; White 1988).

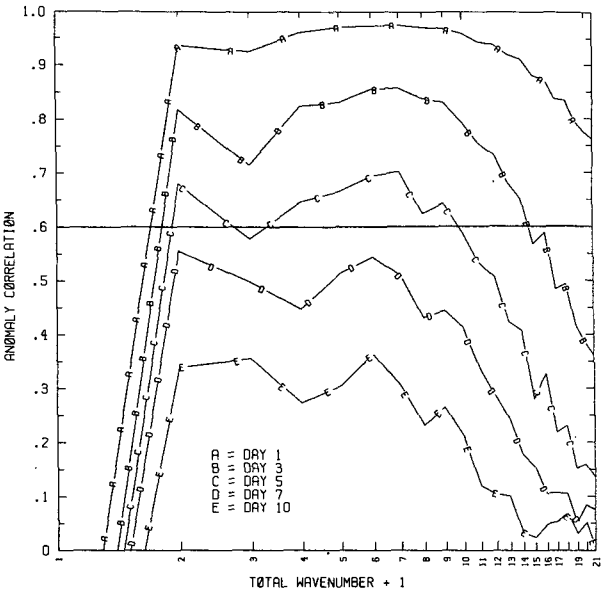

FIG. 2b. As in Fig. 2a but for two-dimensional wavenumber.

b. *Time-filtered statistics*

In Table 1 we tabulate some definitions and quantities pertaining to the time filtering. The overall observed temporal variance in 500 mb height, integrated over all spatial scales, is seen to reside for 15.2%, 26.6% and 58.1% in the high, medium and low frequencies, respectively. Again (compare Blackmon et al. 1984) we are impressed by the large amount of variance in the long time scales. (Of course the 15% height variance in the high frequencies may translate into a lot more variance in traditional weather elements.)

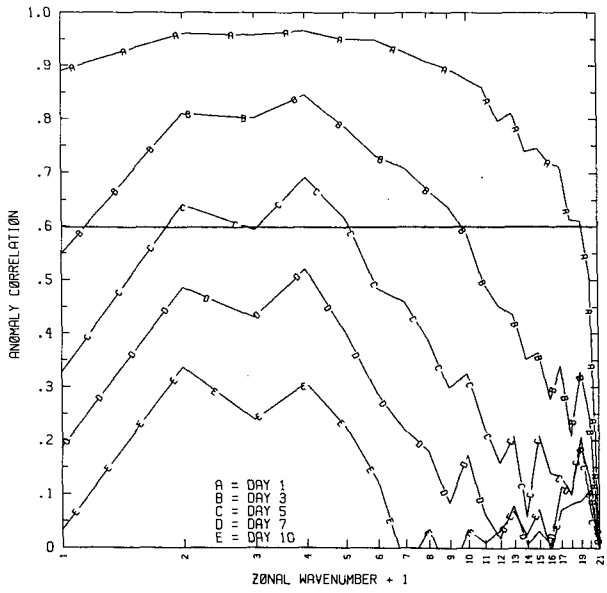

FIG. 2a. The anomaly correlation [Eq. (1) of text] for global 500 mb heights as a function of zonal wavenumber for forecast lead times day 1, 3, 5, 7 and 10.

As outlined in section 2 we can decompose cov_{FA} into the contributions from the three frequency bands. This is done in Fig. 3. Plotted are the covariances for the low, medium and high frequency bands as well as their sum labeled "all"—the covariance being additive. This figure reflects mostly the amount of variance present in each band, see Table 1, especially at short lead times. Nevertheless it is obvious that at large lead times (10 days) the low frequencies dominate in the total covariance. But some sort of normalization, as done in (1), would be desirable. We do not have time-filtered forecasts from which sd_F can be calculated for reasons explained in section 2, but given that sd_F, for each frequency band, should ideally be equal to sd_A we feel confident using the pseudoanomaly correlation AC* defined by (1b). Both quantities in Eq. (1b) can be calculated without filtering the forecasts and hence their ratio is known for each frequency band. The resulting pseudo AC* is shown in Fig. 4. The curve for "all frequencies" in Fig. 4 is nearly identical to that in

TABLE 1. Definition of frequency bands, the value of T_c, and the amount of height variance in each band.

	High	Medium	Low
Frequency (cycles/128 days)	22–64	9–21	0–8
Period (days)	2–6	6–16	18–128
Central period, T_c (days)	3.7	9.5	39.5
Observed variance (gpm^2)	729	1269	2773
Fraction of observed total variance (%)	15.2	26.6	58.1

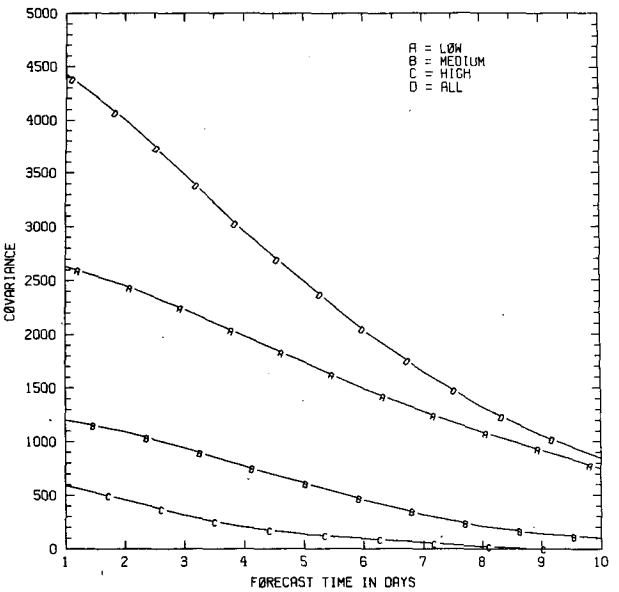

FIG. 3. The decomposition of cov_{FA} [numerator of Eq. (1b) in text] labeled "all," into the additive contributions from "low," "medium," and "high" frequencies, as a function of forecast lead time for global 500 mb height fields. Unit is (geopotential meters)^2.

Fig. 1 [except for assuming $sd_F = sd_A$; in reality the first is smaller than the second, a well-known model deficiency (Arpe and Klinker 1986; White 1988)].

From Fig. 4, a main result of this paper, it can be seen that AC* decreases to 0.6 after about 2, 4 and 5.5

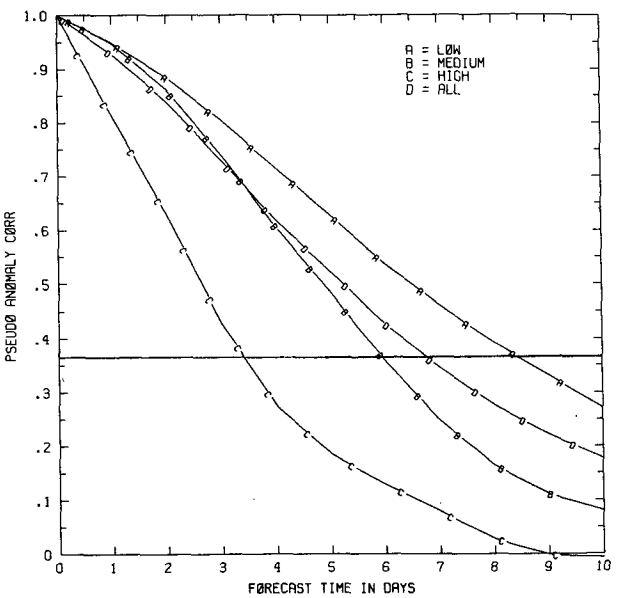

FIG. 4. The pseudoanomaly correlation [Eq. (1b) of text] for "high," "medium," and "low" frequencies as a function of forecast lead time for global 500 mb height fields. The curve labeled "all" refers to unfiltered data.

days for the high, medium and low frequency bands, respectively. In the same order the *e*-folding time scales are 3.4, 5.9 and 8.4 days. This ordering is consistent, at least qualitatively, with a priori expectations stated in the Introduction. At day 5, it turns out, there is virtually no skill (certainly no useful skill) in the forecasts of second or third generation cyclones. There is however, at first sight, considerable skill (AC* greater than 0.6) in the forecast of low frequency events 5 days into the future.

As previously noted, in Fig. 1 the AC* *e*-folding times are appropriate measures of the prediction time scale because AC* decreases very much like an exponential function (slightly steeper in fact). In Fig. 4 this is true for each of the frequency bands as well.

The lifetime of atmospheric events in a frequency band ($k1$, $k2$) is determined by the central period T_c which is calculated from

$$128/T_c = k_c = \sum_{k2}^{k1} k\phi_{zz}(k) / \sum_{k2}^{k1} \phi_{zz}(k)$$

where $\phi_{zz}(k)$ is the variance spectrum of the observed height field.

In Table 2 we compare the AC* *e*-folding times for the three frequency bands to the central period (T_c) of each band. The ratio of these two time scales is given in line 4. The ratio is about 1 for high frequencies, meaning that the integral prediction time scale is roughly the same as the mean lifetime of those events. In contrast, the ratio is only 0.21 for low frequencies. Apparently we have forecast skill only during the early one-fifth of the low frequency events. (Note that we have not removed the systematic forecast error, and there is no doubt that the low frequency AC* suffers more from the systematic error than the high and medium bands.)

A similar assessment follows from comparing *e*-folding time scales in Table 2, line 2, with the lead time at which the AC for a persistence forecast has decreased to $1/e$. Persisting the initial state yields a common control for skill. We did not actually persist the initial states for each band but calculated a proxy AC by

TABLE 2. For each frequency band: time at which AC* has dropped to 0.6 (1) and $1/e$ (2) respectively; central time T_c (3), persistence *e*-folding (5), and ratios of these times (4) and (6).

	High	Medium	Low
(1) AC* = 0.6 (days)	2.1	4.1	5.3
(2) AC* *e*-folding (days)	3.4	5.9	8.4
(3) T_c (days)	3.7	9.5	39.5
(4) (2)/(3)	0.92	0.62	0.21
(5) Persistence *e*-folding (days)	0.7	1.8	7.5
(6) (2)/(5)	4.9	3.3	1.1

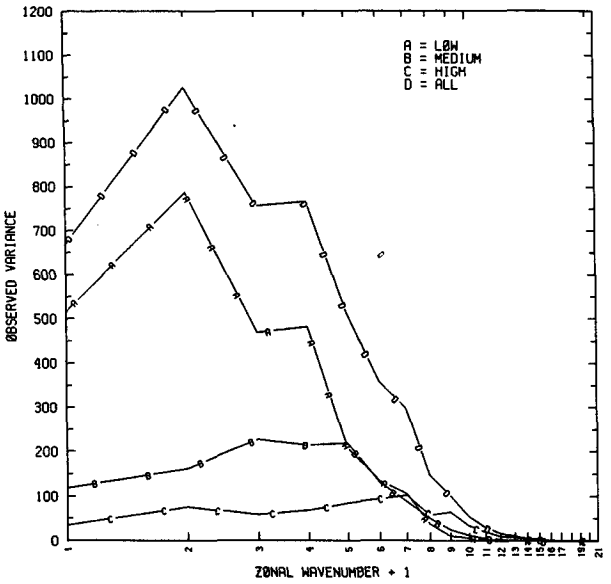

FIG. 5a. Time–space spectral decomposition of observed variance in global 500 mb height fields. As a function of zonal wavenumber, we have plotted the total variance (all), and the variance contributions from the low, medium and high frequencies (additive). Units are (geopotential meters)^2.

$$\mathrm{AC}(\tau) = \frac{\int_0^{T_c} \sin\omega(t-\tau)\,\sin\omega t dt}{\int_0^{T_c} \sin^2\omega t dt}$$

where $\omega = 2\pi/T_c$, and determined the lead time (τ) at which the proxy AC has fallen to 0.37 (roughly $0.19T_c$). Of course the persistence integral time-scale increases with decreasing frequency. Table 2 shows that the MRF model has skill five times longer than persistence in the high frequencies while barely outperforming persistence in the low frequencies. Therefore, the real gain of NWP over easily obtainable controls (such as persistence) is primarily in the high frequencies, a theme also underlying the results of Saha and Van den Dool (1988).

Next we discuss how well time and space scales are related. Figures 5a and 5b show the distribution of the observed variance for the three frequency bands and their total as a function of zonal (Fig. 5a) and total (Fig. 5b) wavenumber. In zonal wavenumber space, the maximum in variance is at $m = 1$ (generally from $m = 1$ to 4) for all, low and medium frequency bands. At high frequencies the distribution is surprisingly uniform with only a weak maximum at $m = 6$. Plotted against total wavenumber (Fig. 5b) the observed variance peaks very strongly around $n = 6$–8 (not unlike Savijarvi 1984, his Fig. 5 curves labeled transient) with a shift towards $n = 9$ to 10 for higher frequencies. Although there is some tendency for high wavenumbers to go along with higher frequencies, the separation is not very satisfactory, especially for zonal decomposition. Note for instance that the commonly applied separation into $m = 1$ to 3 (planetary), $m = 4$ to 9 and $m = 10$–20 (synoptic) (Bengtsson 1985, his Fig. 1a) has a lot of high and low frequency variance deposited into the "wrong" bin. In fact, there is almost no variance at all in the $m = 10$–20 band. Our results are similar to Blackmon's (1976) in this respect.

c. Geographical dependence

So far we have discussed results for the globe as a whole. For smaller domains AC* can be calculated as well, except that we have to sum in physical space with area weighting, rather than summing in spectral space as in (1). Results will be presented for the following domains: 30°–60°N (NH), 30°–60°S (SH), 30°S–30°N (TR) and (again) the whole globe (G). In Fig. 6 we have four panels, one for each frequency band. In each panel the decrease of AC* with forecast lead is given for each of the four domains. The description and conclusions above seem to apply more or less everywhere. The only noteworthy exception is the extremely low skill in the tropics (TR) in the low frequencies. (This is rather disturbing since tropical–midlatitude interaction is often quoted as a source for enhanced predictability in the higher latitudes.) We have to remember, though, that the height variance is very low in the tropics in all frequency bands. Except in this small section, all results pertain to the globe as a whole.

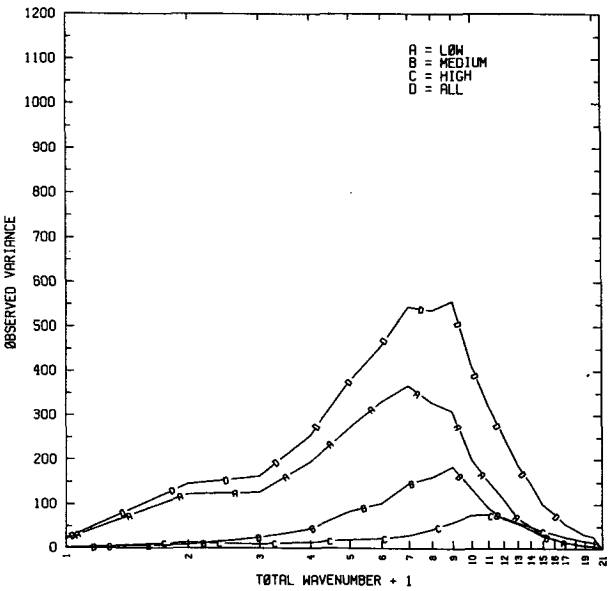

FIG. 5b. As in Fig. 5a but for two-dimensional wavenumber.

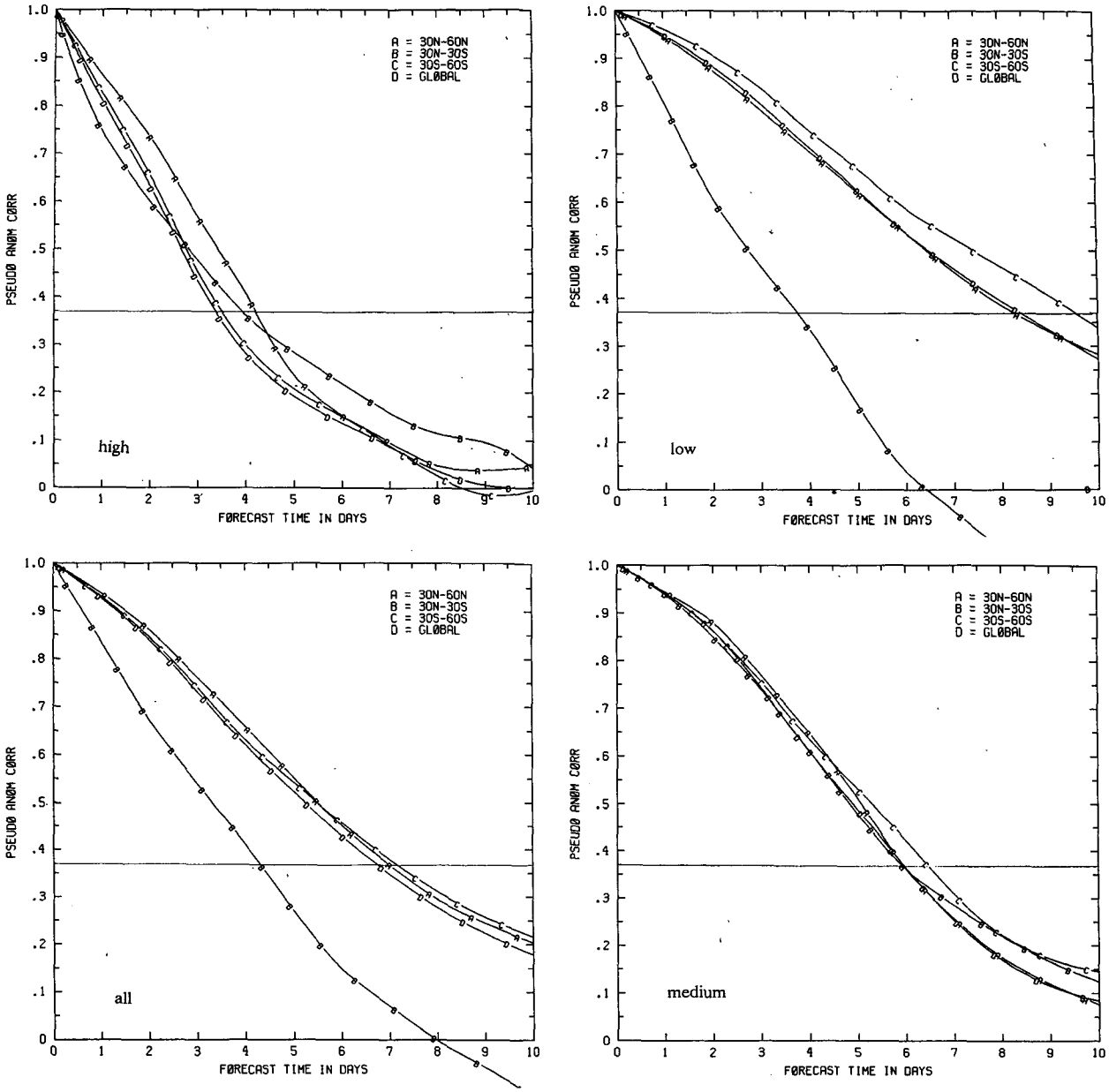

FIG. 6. The decrease of AC* with increasing lead time for low (upper left), high (upper right), all (lower left) and medium (lower right) frequencies. In each panel curves are given for the globe (D), 30°–60°N (A), 30°–60°S (C) and the tropics (B: 30°S–30°N).

4. Conclusion and discussion

We propose a method to calculate forecast skill as a function of temporal frequency. This can, for such measures as the anomaly correlation (AC), be achieved by time filtering the verifying observations only, and by a harmless alteration of the definition of AC (into a pseudo AC*). One thus avoids the difficult problem of time filtering the forecasts directly. To apply the method a large set of consecutive forecasts plus verifying analyses should be available.

We applied the method to a set of 128 consecutive 1–10 day global 500 mb height forecasts produced operationally at NMC with their MRF model during May–September 1988. Characterizing the prediction time scale by the forecast time at which AC* has *e*-folded we reach the following conclusions. (i) Low frequencies are predicted over a longer time (8.6 days) than high frequencies (3.8 days). (ii) In the high frequencies we find prediction skill over the full life cycle of events in that band, whereas in the low frequencies prediction skill is restricted to the early one-fifth of

TABLE 3. As in Table 2 but for 10 December 1988–16 April 1989.

	High	Medium	Low
(1) AC* = 0.6 (days)	3.4	5.0	8.3
(2) AC* *e*-folding (days)	4.8	6.7	12.0
(3) T_c (days)	3.6	9.4	45.5
(4) (2)/(3)	1.33	0.71	0.26
(5) Persistence *e*-folding (days)	0.7	1.8	8.6
(6) (2)/(5)	6.9	3.8	1.4

events only. (iii) Major gains in forecast skills over easily available controls (such as persistence) are made primarily in the high frequencies.

We have investigated 1000 and 250 mb heights along with the 500 mb heights on which the discussion has been based so far. As it turns out the results are nearly the same for these three levels. The ratio of AC* *e*-folding to the central period is identical at the three levels with perhaps one notable exception. At 1000 mb the AC* *e*-folding for low frequencies is about 11 days, which is 2.5 days longer than at 500 and 250 mb. We speculate that at low levels the forecasts benefit from the specification of *observed* lower boundary conditions (sea surface temperature, etc.) which are held fixed during the 10 day forecast run, but are updated at each initial time. Not surprisingly this benefit is visible only in the low-frequency skill. A parallel experiment with prescribed climatological lower boundary conditions being unavailable, we can offer the above only as a likely explanation. The seemingly better prediction skill at 1000 mb may also be an artifact of the analysis/forecast/climatology of a surface that is below the ground in many areas.

The results presented thus far are for a particular 128 day period. Some dependence on the sample should be present, especially in the low frequencies. To check the stability of our results we repeated all calculations for an independent 128 day period, 5 December 1988–16 April 1989. The results are summarized in Table 3, which is as Table 2 except for the more recent 128 day period. Because the annual cycle in the AC calculated for 500 mb height over a global domain is dominated by the annual cycle in the Northern Hemisphere, the *e*-folding times for AC* in Table 3 are all higher than those in Table 2. Nevertheless it is evident that the major conclusions are unchanged. In the low frequencies the ratio of prediction time to lifetime is 0.26 in (NH) winter, while it is 0.21 in (NH) summer.

The most optimistic interpretation of Table 2, line (4), is that if a ratio of about 1 is feasible for all time scales, there is tremendous potential for improvement in forecast skill ahead of us, especially because the low frequencies carry the bulk of the variance. On the realistic side one can ask: Why is the skill below potential in the low frequencies? Also, can we expect to do any better than we currently do in the high frequency transient eddy time scales, which has been the primary focus of NWP for about 40 years? After all, here the ratio is already about 1.

It may be a bit too optimistic to assume that the ratio of prediction time to lifetime could ideally become 1 for all frequencies. We tested this assumption by performing a predictability experiment of the type described in Lorenz (1982). We essentially try to estimate here the ratio of predictability time and lifetime. Denoting $F(N)$ as the N-day forecast verifying against A we calculated, see (1b), AC* = $\mathrm{cov}_{F(N)F(N+1)}/\mathrm{sd}_A^2$ for N = 1 to 9. We now had to time filter time series of 128 N-day forecasts $F'(t)$. This is not an ideal predictability experiment because of the rather severe climate drift in the MRF. To mimic a situation without climate drift we removed the time-mean forecast anomaly. From a plot like Fig. 4 we obtained as *e*-folding predictability times 4.5 days for the high frequencies, and about 15 days (well outside the range of 1 to 10 days) for both low and medium frequencies. To the extent we can trust predictability estimates obtained from a state-of-the-art NWP model, it seems that the ratio may ultimately not reach 1 in the low frequencies. Nevertheless there is room for improvement of the prediction time scale in the low and (especially) medium frequencies. Note also that the ratio may be somewhat higher than 1 in the high as well as medium frequencies.

In this paper we have not addressed the problem of the notion "scale." Both in time/space, a perceived short/small scale phenomenon may project onto a broad spectrum, including long/large scales, when using Fourier modal series. Some of the transient variance found in zonal wavenumber 2 may reflect the inadequacy of the modal representation, and a given cyclone may, over its lifetime, deposit variance in different frequency bands, which would make the linkage of lifetime and prediction time complicated.

Another difficulty in interpretation concerns the distinction of Eulerian and Lagrangian time scales. For large-scale clouds a fine analysis of that distinction was presented by Cahalan et al. (1983). If a NWP model were a perfect "advector," then a presentation of skill, such as in Fig. 1, seems to refer to the Lagrangian time scale of 500 mb height anomalies. Also the notion lifetime seems to apply to a Lagrangian time scale. A Lagrangian time scale is always longer than an Eulerian time scale. But our temporal spectral analysis, which is essentially local in physical space, deals by definition with Eulerian time scales. Also the time scales used in the arguments by Lorenz (1969) and Lilly (1973) are Eulerian. Especially in the high frequencies this distinction makes a difference. It may well be that when comparing lifetime and prediction time an analysis in a Lagrangian framework would imply that improvements can still be made in the high frequencies (as-

suming that the ratio should be order 1). However, the Lagrangian versus Eulerian distinction probably has little impact on the low frequencies and would not alter the conclusion that skill is below its potential in the low frequencies.

Perhaps for LRWP a new kind of NWP model is needed that excludes the transient eddies altogether (or gradually after 3 days) so as to allow the low frequencies to evolve on their own. The studies by Madden (1979) and Branstator (1987) clearly indicate that certain energy-rich modes exist for a long time without too much visible interference by the transients. Kushnir and Wallace (1989) found that zonal waves 1–3 in a general circulation model run were relatively insensitive to the inclusion or exclusion of high frequency transients in the initial state. On the pessimistic side it has been argued by many that nonlinear interaction between various scales of motion will limit predictability in the low frequencies (or large scales) simply because the high frequencies become unpredictable much sooner. A very specific and quantified version of the scale interaction problem is that "weather regimes" constitute equilibrium states between the transient eddies and larger scales (Reinhold 1987), and hence, the lack of absolute accuracy in forecasts of transient eddies is the roadblock in the way of successful LRWP. In that case there would not be much hope, especially because there has been no shred of evidence offered so far in the literature that the collective effect of transient eddies is of any prognostic use after they have, individually, become unpredictable. Also, if we could have afforded NWP models with resolution sufficient to resolve clouds in 1950 we might have concluded that cyclones cannot be predicted over their entire life cycle unless we have absolute precision in the cloud forecasts. By the same token we may have to use models that target low-frequency events that are governed by their own dynamics with at most a (minor?) hindering effect of transient eddies.

Other obvious candidates that could explain the lack of skill in the low frequencies are (i) deficient interaction with the lower boundary, which is often thought to provide the long time scales; (ii) incorrect propagation of long waves into the stratosphere; (iii) an incorrect time-mean state (climate drift) relative to which the low frequencies develop as slow instabilities; (iv) inconsistent vertical/horizontal resolution; and (v) uncertainty about orographic forcing which, certainly in barotropic models, is a major source of low-frequency variability. Also a simple accuracy argument points at increasing difficulties in forecasting the low frequencies. For an amplitude spectrum $\phi(k)$ the time derivative is $k\phi(k)$. So unless $\phi(k)$ drops off faster than k^{-1} [which is not the case except perhaps in very high frequencies (Van den Dool 1975)] the time derivative decreases with decreasing frequency thereby increasing the demand for observational accuracy.

Could it be that something is missing altogether in current NWP models? Graham (1988) has shown that NMC's extended forecast can be improved by a statistical scheme that takes into account a multitude of observed states antecedent to the initial state: in essence we must have some sense of what is going on in the low frequencies. Maybe we need more than one snapshot of the atmosphere (the initial state) to initialize the NWP model. Maybe the true equations are really higher order in time so that more than one time level of data from which to start the time integration can be used. In the early days of NWP the synopticians were sceptical because, by starting the integration from $t = 0$ and ignoring the past, NWP showed complete disregard for "historical development," an issue considered to be of great importance until 1950. Maybe the synopticians were right after all.

Acknowledgments. We would like to thank Drs. B. Reinhold, E. Källén and Å. Johansson for their critical comments on an early version of the manuscript. The final product benefited from comments and suggestions by Drs. S. Schubert, J. Tribbia, E. Epstein, Z. Toth and R. Livezey. The work was supported by the Cooperative Institute for Climate Studies under NOAA Grants NA84-AA-H-00026 and NA89-AA-H-MC066.

REFERENCES

Arpe, K., and E. Klinker, 1986: Systematic errors of the ECMWF operational forecasting model in midlatitudes. *Quart. J. Roy. Meteor. Soc.*, **112,** 181–202.

Bengtsson, L., 1985: Medium-range forecasting—The experience of ECMWF. *Bull. Amer. Meteor. Soc.*, **66,** 1133–1146.

Blackmon, M. L., 1976: A climatological spectral study of 500 mb geopotential height of the Northern Hemisphere. *J. Atmos. Sci.*, **33,** 1607–1623.

——, Y.-H. Lee and J. M. Wallace, 1984: Horizontal structure of 500 mb height fluctuations with long, intermediate and short time scales. *J. Atmos. Sci.*, **41,** 961–979.

Boer, G. J., 1984: A spectral analysis of the predictability and error in an operational forecast system. *Mon. Wea. Rev.*, **112,** 1183–1197.

Branstator, G., 1987: A striking example of the atmosphere's leading traveling pattern. *J. Atmos. Sci.*, **44,** 2310–2323.

Cahalan, R. J., D. A. Short and G. R. North, 1983: Cloud fluctuation statistics. *Mon. Wea. Rev.*, **110,** 26–43.

Dalcher, A., and E. Kalnay, 1987: Error growth and predictability in operational ECMWF forecasts. *Tellus,* **39A,** 474–491.

Graham, N., 1988: Slowly evolving features in the winter Northern Hemisphere 500 mb field and their influence on medium range forecasting. *Proc. of 13th Climate Diagnostics Workshop,* 371–376. [NTIS PB89178115.]

Kushnir, Y., and J. M. Wallace, 1989: Interaction of low- and high-frequency transients in a forecast experiment with a general circulation model. *J. Atmos. Sci.*, **46,** 1411–1418.

Leith, C. E., 1973: The standard error of time-average estimates of climatic means. *J. Appl. Meteor.*, **12,** 1066–1069.

Lilly, D. K., 1973: Lectures in sub-synoptic scales of motion and two-dimensional turbulence. *Dynamical Meteorology,* P. Morel, Ed., pp. 622.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus,* **21,** 289–307.

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

Madden, R. A., 1979: Observations of large-scale travelling Rossby waves. *Rev. Geophys. Space. Phys.,* **17,** 1935–1949.

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

Reinhold, B. B., 1987: Weather regimes: The challenge in extended range forecasting. *Science,* **235,** 441–447.

Saha, S., and J. Alpert, 1988: Systematic errors in NMC medium range forecasts and their correction. *Eighth Conference on Numerical Weather Prediction,* Amer. Meteor. Soc., 472–477.

——, and H. M. van den Dool, 1988: A measure of the practical limit of predictability. *Mon. Wea. Rev.,* **116,** 2522–2526.

Savijarvi, H., 1984: Spectral properties of analyzed and forecast global 500 mb fields. *J. Atmos. Sci.,* **41,** 1745–1754.

Shukla, J., 1981: Dynamical predictability of monthly means. *J. Atmos. Sci.,* **38,** 2547–2572.

Straus, D. M., and J. Shukla, 1981: Space–time spectral structure of a GLAS general circulation model and a comparison with observations. *J. Atmos. Sci.,* **38,** 902–917.

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,** 2306–2317.

Van den Dool, H. M., 1975: Three studies on spectral structures of the horizontal motion in the time domain. Ph.D. dissertation, Rijksuniversiteit, Utrecht, Netherlands, 128 pp.

——, 1985: Prediction of daily and time-averaged temperature for lead times of 1 to 30 days. *Proc. Ninth Conf. on Probability and Statistics in Atmospheric Sciences,* Virginia Beach, Amer. Meteor. Soc., 149–153.

Wallace, J. M., G.-H. Lim and M. L. Blackmon, 1988: Relationship between cyclone tracks, anticyclone tracks and baroclinic waveguides. *J. Atmos. Sci.,* **45,** 439–462.

White, G. H., 1988: Systematic performance of NMC medium-range forecasts 1985–88. *Eighth Conf. on Numerical Weather Prediction,* Baltimore, Amer. Meteor. Soc., 466–471.