A Measure of the Practical Limit of Predictability

SURANJANA SAHA AND H. M. VAN DEN DOOL

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

(Manuscript received 18 February 1988, in final form 15 June 1988)

ABSTRACT

An objective and practical limit of predictability for NWP models is proposed. The time T_0 is said to be the limit of predictability if a model forecast beyond T_0 has no extra skill over persisting the T_0 forecast. The "skill" is measured here in terms of standard rms and anomaly correlation scores. For the NMC medium-range forecast model, T_0 is found to be 5–6 days for 250, 500 and 1000 mb height forecasts for the period 5 May–25 July 1987. The T_0 can also be interpreted as the time at which there is no longer skill in the prediction of the time derivative of the quantity under consideration.

1. Introduction

The role of numerical weather prediction (NWP) models in short and medium range weather forecasting has increased tremendously over the past 30 years. A logical extension would be to apply the method of NWP to long-range weather forecasting. Some of the optimism regarding the use of NWP in long-range forecasting is generated by curves such as in Fig. 1. In this figure, just a recent example, we have plotted the anomaly correlation (AC) of 1–10 day forecasts of the 500 mb height field produced by the Medium Range Forecasting (MRF) model in operational use at the NMC (National Meteorological Center, Washington, DC). The correlations are calculated for the global domain, and averaged over the 82 forecasts in the period 5 May to 25 July 1987.

We are very accustomed now to the monotonic decrease of the AC with increasing forecast time (Miyakoda et al. 1972), as seen in Fig. 1. The area between the curve labeled F and the X-axis represents our current average knowledge of future 500 mb heights; above the curve F we enter "uncharted territory."

Even at day 10 the average AC for NWP forecasts is not quite zero and many efforts are therefore directed towards getting the signal more clearly out of the noise by time averaging, averaging lagged forecasts, or ensemble forecasting. But we will argue that just looking at the simple AC may overstate the true ability of NWP models to forecast the flow 10 days ahead.

Here we will develop a straightforward criterion to determine whether an integration of the NWP model out to M days actually improves the M-day forecast over an easily obtainable control forecast. The first step is to understand the role of the curve labeled P (see Fig. 1) which is determined by verifying the initial condition as if it were the M-day forecast ($0 \leq M \leq 10$ days)—the "persistence method." Out to 10 days the AC for curve F is considerably higher than that for P. And so it is not uncommon to conclude from Fig. 1 that it makes good sense to run the model out to 10 days (or even beyond). After all, we are gaining "something" relative to persistence, not just relative to the zero level (the X-axis).

However, let us now persist the 1-day forecast instead of the initial condition. Since the 1-day forecast is rather accurately known, the new curve will be virtually coincident with the P curve but translated to the right by 1 day. Obviously, using this as the new control to be beaten, we gain much less in, say, the 4-day forecast than we thought we did.

We can continue this procedure by persisting the N day forecast out to M days ($0 \leq N \leq M \leq 10$) and compare the result of these controls to the actual M day forecast. So the question to be answered now is: Does an extension of the model integration beyond N days yield a forecast that is better, by some measure, than persistence of the N day forecast? Here we address that question using a long series of height forecasts made by the MRF model. Results for various domains (global, 20°–80°N), levels (250, 500, 1000 mb), various spectral wavenumber bands, all in terms of root-mean-square (rms) error as well as anomaly correlation, will be given in section 3.

2. Data and analysis

The data consist of 1–10 day forecasts of heights at 250, 500 and 1000 mb, verifying at 5 May–25 July 1987. The forecasts (F) are made by NMC's MRF model. The verifying initial conditions (O) are also

Corresponding author address: Dr. Huug van den Dool, Dept. of Meteorology, The University of Maryland, College Park, MD 20742.

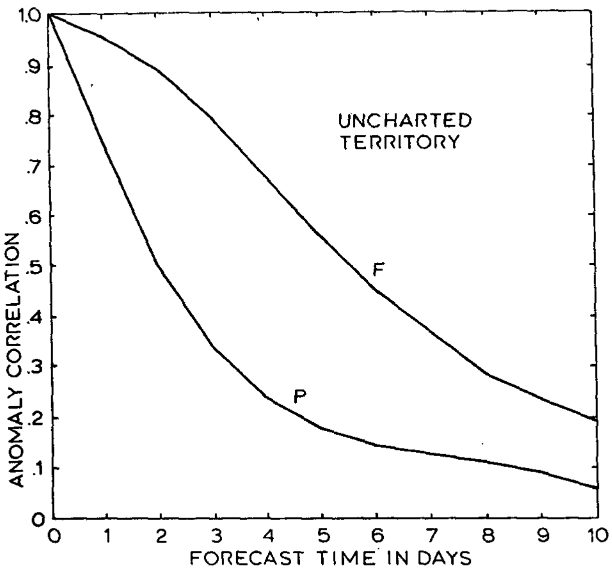

FIG. 1. The anomaly correlation of eighty-two 500 mb height forecasts as a function of forecast lead time (in days). The curve labeled F refers to forecasts made by NMC's MRF model, while P refers to persistence of the initial state. The period of verification is 5 May–25 July 1987; the domain is global.

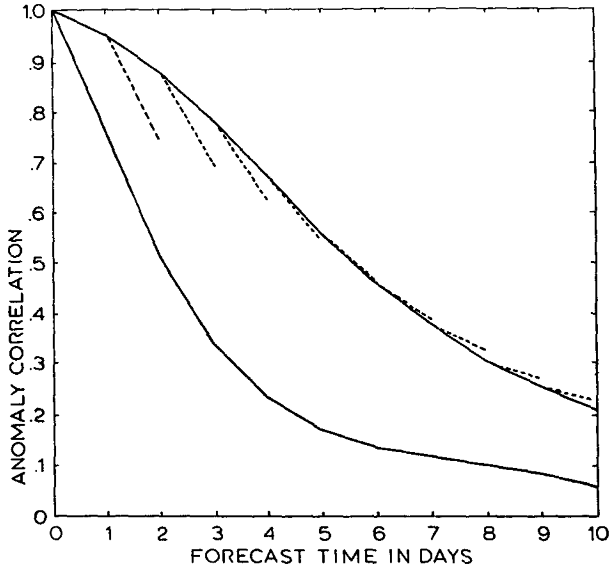

FIG. 2. The anomaly correlation of eighty-two 500 mb height forecasts as a function of forecast lead time (in days). The solid curves are identical to those in Fig. 1. The nine dashed segments represent verification of the day N forecast persisted out to $N + 1$ days; $1 < N < 9$ days. The period is 5 May–25 July 1987; the domain is global.

available. The domain is global and the height fields are represented in spherical harmonics with a Rhomboidal truncation at zonal wavenumber 30.

The verification tools are 1) rms error, and 2) anomaly correlation (AC). The rms is standard i.e., root mean square of (F − O). While calculating the AC we subtracted the same climatology (C) from both forecasts and observations, obtained from the Climate Analysis Center, which is based on NMC's initial states for their global model runs during the years 1978–85. The climatology is a monthly mean, interpolated from the nearest two calendar monthly means to the date in question. For the AC (on any domain) we used Miyakoda et al.'s (1986) Eq. (8), carrying the products in time as well, i.e.,

$$AC = \frac{\sum_t \sum_s (F - C)(O - C)}{[\sum_t \sum_s (F - C)^2 \sum_t \sum_s (O - C)^2]^{1/2}} \quad (1)$$

where t is time (5 May–25 July) and s is a space index, either wavenumber space or gridpoint space.

3. Results

In Fig. 2 we have repeated the F and P curves from Fig. 1 (the full lines) along with the first segment of lines representing the AC if one were to persist the N-day forecast ($N = 1, \cdots, 9$). Only the first segments of the nine curves are shown to avoid overcrowding at the right-hand-side of the figure near day 10 and AC = 0.2. The model outperforms persistence out through 4 days. At 5 days it is better (in terms of AC) to persist the 5-day forecast than to continue the model integration, therefore T_0 = 5 days. This result is reasonably insensitive to the method of verification. Figure 3 is as Fig. 2 but for the rms error instead of AC. From Fig. 3 we also conclude T_0 = 5 days. Both Figs. 2 and 3 display a systematic rotation of the dashed segments relative to the F curve. So while at 5 days there is little

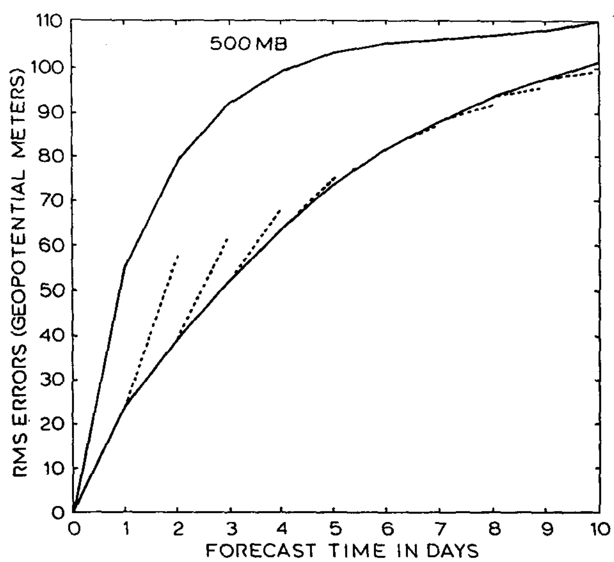

FIG. 3. As Fig. 2, but for the root-mean-square error (in gpm) instead of anomaly correlation.

difference between persisting the 5-day forecast and continuing the NWP model run, at 9 days persistence is decisively better. At 5 days the forecast errors have become so large that continued integration is more harmful than assuming that the time derivatives are 0 (persistence).

The results, condensed into one number T_0, are quite uniform with height. Figure 4 is as Fig. 3 but now for 250 mb height. Again $T_0 = 5$ days!

Table 1 gives a summary of T_0 for height forecasts at three levels, in several spectral bands. All T_0s were determined from visual inspection of graphs such as Figs. 2–4, both for AC and rms. In all cases T_0 is somewhere between 4 and 7 days. No spectacular dependence on wavenumber can be seen, either for total (n) or zonal (m) wavenumber. It has often been speculated that the low wavenumbers are more predictable than the rest of the spectrum. In terms of zonal wavenumber m, the low wavenumbers (m = 0–3) have a T_0 similar to or even smaller than T_0 for m = 4 to 9. However, using the total wavenumber n as scale selector, T_0 is indeed larger by about 2 days for $1 \leq n \leq 5$ compared to $6 \leq n \leq 12$, at least at 500 and 250 mb. The behavior of the global mean height at all levels is somewhat odd in that it loses out against persistence right from the start: $T_0 = 0$.

The very low values of T_0 for the global mean as well as for m = (0, 3) could be caused by the "systematic" error which tends to be a large fraction of the total error at low wavenumbers. So the values in parentheses in Table 1 are T_0 recalculated for "corrected" forecasts. The correction is an honest method in that it could be applied in an operational setting. The time-mean forecast error of the 21 most recent verifiable forecasts has been removed from the next forecast at scales $n \leq 6$ (Saha and Alpert 1987). (This procedure nearly eliminates the bias-error but gives at best a modest reduction in total rms error.) After correction, T_0 changes only slightly, mostly in the global mean (T_0 > 0 at last) and for m = (0, 3) which, after correction, has a T_0 value 1 day larger than the intermediate $4 \leq m \leq 9$ band.

In summary, T_0 = 5–7 days for geopotential heights, relatively uniform in the vertical, with the largest T_0 values occurring for low n at 500 and 250 mb.

Although, by our measure, there is no skill beyond T_0 in instantaneous flow forecasts superior to persistence of a previous forecast, it has been suggested that time averaging of the forecasts could still display skill. In fact the 6–10 day averaged forecast has been used in operational forecasting at NMC under that assumption (See Guaraldi and Bosart 1985, for a critical assessment). In Table 2 we compare the AC for the averaged 6–10 day forecast to the AC of the instantaneous 5, 6, 7 and 8 day forecast verified against the same observed 5-day mean, valid for day 6–10. As can be seen the (average) AC for this period in May, June

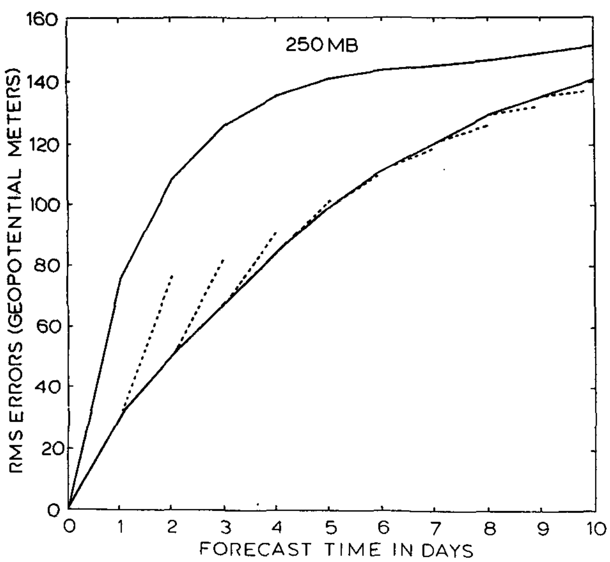

FIG. 4. As Fig. 3, but for 250 mb height forecasts.

TABLE 1. The T_0 (in days) for height forecasts for the MRF model at three levels and various total (n) and zonal (m) wavenumber bands, according to rms error (left) and Anomaly Correlation (right). Numbers in parentheses are T_0 values calculated from forecasts corrected for the time-mean error (see text). The period is 5 May–25 July 1987.

Wavenumber band	Level (mb)					
	Rms			AC		
	1000	500	250	1000	500	250
all R30	6 (6)	5 (5)	5 (5)	6 (6)	5 (5)	5 (6)
(0, 0)	0 (0)	0 (2)	0 (6)	0 (0)	0 (4)	0 (≥10)
$1 \leq n \leq 5$	5 (6)	7 (7)	7 (6)	6 (6)	7 (7)	7 (6)
$6 \leq n \leq 12$	6 (5)	5 (5)	5 (5)	6 (5)	5 (5)	5 (5)
$0 \leq m \leq 3$	5 (6)	4 (6)	4 (6)	7 (6)	4 (6)	4 (6)
$4 \leq m \leq 9$	6 (5)	5 (5)	5 (5)	6 (6)	5 (5)	5 (5)

TABLE 2. The anomaly correlation (×100) for height forecasts of the MRF model at three levels for the 6–10 day mean, and for instantaneous forecasts (5, 6, 7 and 8 days). All forecasts are verified against the same 5-day mean observed during the days for which the 6–10 day forecast is valid. The period is 5 May–25 July, 1987.

Day	Level (mb)					
	Global R30			$20°N \leq \phi \leq 80°N$; $1 \leq m \leq 20$		
	1000	500	250	1000	500	250
6–10 mean	52	38	36	31	44	45
5	44	34	33	23	38	40
6	48	38	37	27	42	43
7	49	38	36	29	42	43
8	46	34	32	26	39	40

and July 1987 is already quite low, about 0.4 or so. Among the instantaneous forecasts the one for 6 and/or 7 days ahead verifies best against the 6–10 days observed average, beating the 5 and 8 day forecast by about 0.04. The forecast averaged over day 6–10 is better than the instantaneous 6 day forecast by only a small amount, ranging from −0.01 to 0.04 (see Table 2). This exercise demonstrates that there is no significant advantage in forecast skill to be gained by integrating the model out to 10 days and by averaging day 6 through 10.

4. Conclusions and discussion

We propose here a measure of the practical limit of predictability for NWP models. As soon as a continued integration starting from day N does not yield, on the average, at day $N + 1$, a higher AC (or lower rms) than persisting the N-day forecast out to day $N + 1$, the practical limit of predictability is said to be reached. This happens at $T_0 = N$. We obtained the values of T_0 for tropospheric height forecasts made by NMC's MRF model for the period of May 5 to July 25, 1987. The conclusions are:

1) T_0 is 4–7 days, depending slightly on pressure, wavenumber and whether rms or AC is the desired measure of skill.

2) T_0 tends to be increased by 1–2 days (but $T_0 \leqslant 7$ days even then) after removal of the systematic error.

3) T_0 is largest for the low wavenumbers (but $T_0 \leqslant 7$ days even then), most clearly so for low total wavenumbers at 500 and 250 mb.

4) The 6–10 day averaged forecast has equal, or only up to 0.04 better, AC than the instantaneous 6 or 7 day forecast, both compared against the same observed 6–10 day average.

In summary, under the conditions stated below, there is currently not much *operational* weather forecasting utility in integrating NWP models beyond about 5–6 days. This conclusion is true for this model (NMC's MRF model as of spring/summer 1987), on the average (all tables and graphs display 3-month mean AC and rms, calculated over the whole globe), for this particular period and for the quantities considered (height of tropospheric pressure levels). (Further research is on its way to test these results for other seasons and other models.)

Alternatively, recall the formalism of time integration, namely:

$$A^{n+1} = A^n + \left(\frac{\partial A}{\partial t}\right)\Delta t.$$

Apparently beyond $T_0 \approx 6$ days it is better to keep $\partial A/\partial t = 0$ (persistence) than to update A by using an erroneous time derivative. So apparently beyond T_0 there is no longer skill in the forecast of $\partial A/\partial t$, knowledge of which is the essence of forecasting A.

The $T_0 = 5$–7 days is certainly much less than the classical theoretical limit of predictability which is estimated to be 2–3 weeks (Lorenz 1969). The latter is sometimes defined as the time (T_s) when the rms error has reached 95% of its saturation value (Dalcher and Kalnay 1987). Our limit turns out to be more stringent than the 95% criterion. Between T_0 and T_s, the model produces forecasts with skill but that level of skill (or slightly higher) can also be obtained by persisting earlier forecasts.

Currently there is considerable attention focused on the question of forecasting forecast skill (Kalnay and Dalcher 1987). That is, can be recognize in advance those forecasts that will have a high level of skill? Since our $T_0 \approx 6$ days pertains to an average case, one might wonder whether T_0 is not much larger in some cases. We have not addressed that question, but do note that if the predictable cases are also the persistent cases it would be harder for the NWP models to beat the persistence control so that, on balance, T_0 would not increase as much as T_s.

Although we think it is useful to keep track of T_0 as a diagnostic tool for model performance we note a serious problem. Many operational NWP models suffer from a monotonic decrease in variance with increasing forecast time. Therefore the rms error of persistence is ultimately larger than the rms error of a regular forecast, thus pushing T_0 towards infinity. The decrease of amplitude is, in the MRF model, so large that T_0 can not be determined for high zonal wavenumbers. To some extent all values of T_0 reported in Table 1 suffer from that problem, and are therefore overestimated. So, in order to calculate T_0 properly one would therefore have to make the variance of forecast and observed atmospheric states equal. The same argument has to be made for determining the theoretical limit of predictability through methodologies such as in Lorenz (1982).

Although our T_0 is more stringent than T_s we have not necessarily defined the severest possible test. After all, given the initial state and the 1-day forecast, for example, there are perhaps more intelligent combinations that may serve as a better control for the 2-day forecast than persisting the 1-day forecast.

The European Centre for Medium Range Weather Forecasts (ECMWF) has popularized AC = 0.6 as one limit of useful predictability (See Hollingsworth et al. 1980, plus references therein). This value reflects the consensus of a large number of forecasters, who were asked to judge subjectively the skill of a large number of forecasts for different flow types. We note that, coincidentally or not, the T_0 value corresponds roughly to AC = 0.55 (see Fig. 2, for example). Of course the rms error and AC measures are not sacred, and for some users, it may be better to apply different verification measures leading to perhaps different values for T_0.

The idea of using T_0 as the practical limit of predictability is not entirely new. Grönaas (1985) presented graphs rather similar to Fig. 2 for ECMWF forecasts of 500 mb heights over a small European area for the summer of 1985. He made the following observation: "although the anomaly correlation is high, the model does not seem to add information to the forecast after day 7, since the 7 day persistence gives better scores than the model." Grönaas (personal communication, 1988) felt that a measure like T_0 makes sense as a limit of useful predictability, better perhaps than any particular fixed value of the AC.

Acknowledgments. We would like to thank innumerable colleagues for reading the manuscript and returning it with mostly encouraging remarks. It is our pleasure to acknowledge in particular Dr. P. Caplan for stimulating discussions on the 6–10 day forecast, Dr. A. Hollingsworth for calling our attention to the Grönaas reference, Dr. Å. Johansson for his particularly close-reading of the manuscript and Charlene Mann and Marco Rodriques for typing and drafting respectively. The work was supported by the Cooperative Institute of Climate Studies under NOAA Grant NA84-AA-H-00026.

REFERENCES

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

Grönaas, S., 1985: A pilot study of the prediction of medium range forecast quality. Tech. Memo. No. 119. [European Centre for Medium Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG29AX, England.]

Guaraldi, J., and L. F. Bosart, 1985: An assessment of the skill of the NMC 6–10 day temperature and precipitation outlook, 1978–1981. *Bull. Amer. Meteor. Soc.,* **66,** 831–839.

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

Lorenz, E. N., 1969: Three approaches to atmospheric predictability. *Bull. Amer. Meteor. Soc.,* **50,** 345–349.

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

Kalnay, E., and A. Dalcher, 1987: Forecasting forecast skill. *Mon. Wea. Rev.,* **115,** 349–356.

Miyakoda, K., G. D. Hembree, R. F. Strickler and I. Shulman, 1972: Cumulative results of extended forecast experiments. I: Model performance for winter cases. *Mon. Wea. Rev.,* **100,** 836–855.

——, J. Sirutis and J. Ploshay, 1986: One month forecast experiments—without anomaly boundary forcings. *Mon. Wea. Rev.,* **113,** 2363–2401.

Saha, S., and J. C. Alpert, 1987: Correction of systematic errors in NMC's medium range forecasts. *Proc. XIIth Annual Climate Diagnostics Workshop,* Salt Lake City, Oct. 12–16, 1987, NTIS.