ESTIMATES OF ATMOSPHERIC PREDICTABILITY AT MEDIUM RANGE

Edward N. Lorenz
Massachusetts Institute of Technology

ABSTRACT

Recent studies based upon the output of the ECMWF operational forecasting model indicates that if, after the first day of a forecast, a perfect model could be substituted for the present model, forecasts as good as those presently produced at seven days would be realized at ten days. These studies do not reveal how much improvement in one-day forecasting is possible.

We hypothesize that if all other imperfections in the forecasting procedure could be removed, the inevitable initial uncertainties in observing the small-scale features would, after D days, lead to error fields with amplitudes and spectra resembling those of the errors in present one-day forecasts. The appropriate value of D is highly dependent upon the spectrum of actual atmospheric motions. Estimates with a crude model place D at about four days, thereby implying that the present forecasting success at one week may some day be realized at nearly two weeks.

INTRODUCTION

Many studies which have addressed the problem of the predictability of the atmosphere or some other fluctuating system have been investigations of error growth. The basic question posed in these studies is the following: if at some time t_o we could alter the state of the atmosphere by a certain amount, and subsequently allow the atmosphere to be governed again by the correct dynamics, how greatly would the state at some later time t_1 differ from the state which would have occurred at time t_1 if no alteration at t_o had been made? The relevance of this question to predictability becomes evident when the altered state at time t_o is identified with the observed state, taking into account the inevitable shortcomings of the observations. Since the difference between the states at time t_1 is the error which an optimal extrapolatory prediction scheme would make, the expression "error growth" is apt.

Error growth studies have many ramifications; the initial error may be systematic or random, it may be restricted geographically or distributed over the globe, it may be limited to certain scales of motion or spread over the spectrum, and it may be confined to certain atmospheric properties or allocated to all. The resulting error at a later time possesses similar possibilities.

Since it is not feasible to make deliberate alterations of the atmospheric state resembling typical errors of observation, and, in any event, if we did alter the state we could never observe the evolution of the unaltered state, most studies of error growth have been based upon numerical models of the atmosphere. The wide variety of results obtained reflects the fact that the growth rates are model-dependent. It is often taken as an article of faith that the more closely the model duplicates the readily observed features of the atmosphere, the more reliably it will reveal the growth rate.

Since it is unlikely that the very best possible prediction procedure will ever be formulated, an error-growth study should yield an upper bound to the accuracy with which prediction can be made at any range, or to the range at which prediction can meet a chosen measure of acceptability. Such an estimate will, of course, depend upon the assumed magnitude and nature of the errors of observation, and one obvious way to extend the range of acceptable prediction would appear to be to improve the observing system. As we shall see, however, there is reason to believe that the range of predictability cannot be made to approach infinity by making the observational error approach zero, so that there should be an intrinsic upper bound to predictability. A lower bound can be obtained by noting how well the best currently used prediction procedures perform. As further refinements are made in operational prediction and in error-growth studies, it may be expected that these bounds will approach one another; if they should ever be made to coincide, the technique of weather forecasting will have been perfected.

The purpose of this study is to obtain up-to-date estimates of upper and lower bounds to medium-range atmospheric predictability. For our purposes the medium range will extend from about one-half to about two weeks. We shall be concerned mainly with the extratropical troposphere, and with those atmospheric properties which would characterize a "dry" atmosphere, namely wind, pressure, and temperature. We shall consider how well these quantities may be predicted on the average, rather than in individual situations or at individual locations. We shall not undertake any new computations, and our conclusions will be drawn from the results of studies which have already been performed.

THE MIDDLE AND LATE STAGES OF ERROR GROWTH

We begin by turning to the results of a predictability study[1] which we recently performed with the output of the operational model at the European Centre for Medium Range Weather Forecasts (ECMWF). We shall describe the study briefly; for further details the reader is referred to the cited paper.

Operational forecasts j days in advance, for $j = 0, 1, ..., 10$, are prepared daily at ECMWF. (By a zero-day forecast we mean simply an analysis.) The forecasts are made with a 15-level global primitive equation model with moisture and orography. The model is a grid-point model, but, before being archived, each analyzed or predicted field is represented by a series of global spherical harmonics, truncated triangularly at wave number 40, and it is the 1722 coefficients in each of these sequencies which are stored. In our study we have used only the 500 mb height fields, analyzed on or predicted *for* each of the 100 consecutive days beginning 1 December 1981. Our data thus consists of $100 \times 11 \times 1722 = 1,894,200$ numbers.

To use these data for a predictability study, we note first that the forecasts one day in advance are reasonably good; hence the 1-day forecast for day i, prepared on day $i - 1$, may be treated as the analysis or 0-day forecast on day i, plus a reasonably small superposed error. By comparing the 2-day and 1-day forecasts for day $i + 1$, we can observe how much this error grows in one day, when both fields are governed by the operational model. Likewise we can obtain the growth during j days, for $j \leqslant 9$, by comparing the $(j + 1)$-day and j-day forecasts for day $i + j$. We also note that forecasts two or more days ahead possess some skill, so that by comparing the $(j + k)$-day and j-day forecasts for day $i + j$, with $j + k \leqslant 10$, we can observe the growth of errors of various initial magnitudes.

As a measure of the difference between two 500 mb height fields we have chosen the root-mean-square difference in height. Figure 1, which is based on a figure in Reference (1), contains the principal results. It shows the differences, in meters, between j-day and k-day forecasts for the same day, averaged over the 100 days of the study, for all pairs (j,k) with $j < k$ and $k \leqslant 10$; these are plotted against k. A heavy curve connects the points where $j = 0$, i.e., where an analysis is compared with a forecast, and it therefore summarizes the performance of the model. The indicated growth rate is the rate at which solutions of two different systems of equations — those of the model and the real atmosphere — diverge from one another. Thin curves connect points having equal values of $k - j$, and indicate the rate at which separate solutions of a single system of equations — those of the model — diverge. This is the rate which is ordinarily evaluated in predictability studies. The dashed curves are extrapolations of the thin curves; we shall presently describe the basis for extrapolating.

The latter rate is supposed to approximate the rate at which separate solutions of the real atmospheric equations diverge. If indeed it does, and if, after the first day the model could be replaced by a perfect model, the heavy curve in Figure 1 would coincide with the lowest thin curve. The actual difference between the slopes of the curves should therefore be a measure of the amount of improvement which may still be realized. In particular, 10-day forecasts should ultimately become better than present 6-day forecasts, even if the 1-day forecast is not improved at all.

To a fair approximation the separate thin curves differ only by horizontal displacements, i.e., the error growth during one day is a function of the magnitude of the error. We may therefore extrapolate the lower thin curves beyond 10 days, by displacing the higher thin curves horizontally. We conclude that 14-day forecasts should become as good as present 8-day forecasts.

This conclusion may be overly optimistic. Since the model is not perfect, it does not necessarily yield the correct growth rate, and it may give an underestimate. In that event, as the model is continually improved, and the heavy curve moves down toward the lowest thin curve, the latter curve may move up to meet it. Improvement in forecasting will then be less spectacular.

Better and better models may be anticipated in the coming years, but some improvements may be introduced immediately. First of all, the ECMWF model produces some systematic errors[2]; these may be subtracted from the forecast. Second, the model is for practical purposes a better model in the northern than in the southern hemisphere, so that we may study the performance of a better model by evaluating root-mean-square height differences for the northern hemisphere only. Introducing these "improvements", we find that the slope of the heavy curve has been reduced, but the slope of the lowest thin curve has been steepened somewhat. Our revised conclusion is that, with no further improvement at one day, 10-day forecasts should ultimately become as good as the 7-day forecasts, and 14-day forecasts should become as good as the 10-day forecasts, which can presently be made by the *improved* ECMWF model.

A familiar measure of error growth is the doubling time for small errors. The smallest error in Figure 1, about 25 m, doubles in about 3.5 days. Larger errors grow less rapidly and ultimately level off. To obtain the doubling time for truly small errors we need to extrapolate the thin curves to the left. It is obviously impossible to do this in any unique manner unless we introduce some auxiliary hypothesis. We postulated[1] that the nonlinear terms in the equation for the growth of the root-mean-square error were essentially quadratic, after which we found that small errors would double in about 2.5 days. Repeating the computation with the "improved" ECMWF model reduced the time to about 2.0 days. This doubling time is consistent with the times obtained from earlier studies[3], although somewhat shorter. It presumably approximates the doubling time for errors in other atmospheric properties which are closely coupled with height, such as wind and temperature, but it may bear little relation to the doubling time for precipitation errors.

THE EARLY STAGE FOR ERROR GROWTH

Having seen that considerable improvement in medium-range prediction is potentially realizable even without altering the one-day forecasts, it is natural to ask how much improvement is possible at one day, and how much effect any such improvement might have upon the medium range. Unlike the growth of errors represented by the lowest curve in Figure 1, which assumes an optimal prediction proceedure, the error at one day is the error produced by a currently used procedure, and results from imperfections in both the forward extrapolation and the initial analysis. Perfecting the extrapolation procedure ought to improve the one-day forecast, but improving the analysis, perhaps by establishing a superior observing system, might have a considerably greater effect. Nevertheless, we feel that

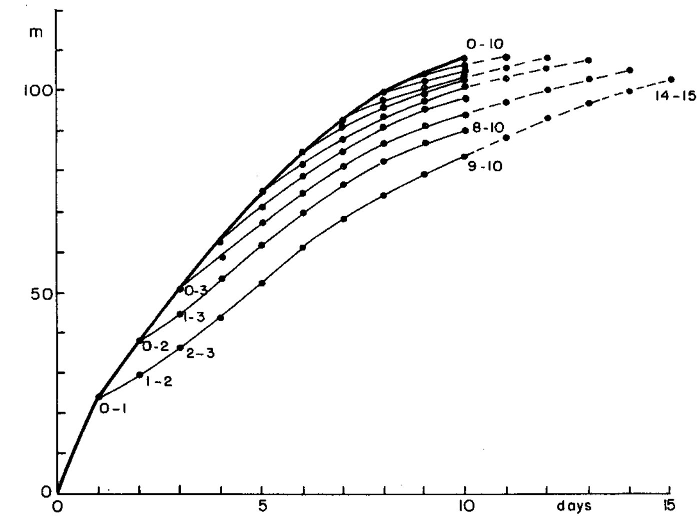

Fig. 1. Average global root-mean-square height differences, in meters, between j-day and k-day operational forecasts with the ECMWF model, for $j < k$, plotted against k. Values of (j,k) are shown beside some points. Heavy curve connects points where $j = 0$. Thin curves connect points with equal values of k- j. Dashed curves are extrapolations of thin curves; see text.

any assumption that continual reductions in the analysis error would lead to proportional reductions in the one-day prediction error is overoptimistic.

The errors represented by the points in Figure 1 are errors in the scales of motion which are resolved by the network of grid points. The *actual* error fields also include errors made by completely omitting the scales too small to be resolved. At present the actual one-day errors and very likely the analysis errors are dominated by the larger scales. However, if the observing system, and so presumably the analysis, are to be subjected to continual improvement, we can anticipate a day when this situation will no longer prevail, and the principal remaining errors will be in the unresolved or poorly resolved scales.

The two-day doubling time deduced from the ECMWF model is presumably an average over the resolved scales, with each scale weighted according to its contribution to the total error. The doubling therefore results from the self-amplification of errors in these scales. A small part of the *actual* error growth results from the influence of errors in the unresolved scales, which induce errors in the resolved scales through nonlinear interactions. Under an ideal observing system, if errors in the resolved scales are greatly reduced, the augmentation of these errors due to self-amplification will be reduced, in proportion, but the acquisition of errors from the unresolved scales will not. The error growth in the resolved scales will then be dominated by the transfer from the unresolved scales, and, until the resolved-scale errors become large, their proportional growth will be much greater.

The smaller scales themselves will amplify quite rapidly, until they approach their maximum size. At very small scales, for example, errors in the structure of a thunderstorm should amplify at least as rapidly as the thunderstorm itself, doubling in an hour or less rather than two days. It follows that any reduction in the transfer of errors from smaller to larger scales, which might be realized by reducing the initial errors in the smaller scales, will be short-lived, since small errors in the smaller scales will not remain small. In any event, developing an observing system which would resolve the mesoscale features, let alone the thunderstorms, would be a difficult and costly undertaking.

We therefore hypothesize that, when the best foreseeable observing system is put to use, the initial growth of errors in the larger scales will be dominated by the influence of the smaller scales. We further hypothesize that after D days the larger-scale errors will have grown to the point where the total error field resembles a one-day error field made by present procedures, in amplitude and spectrum. The range, beyond one day, at which predictions meeting any given measure of acceptability can be made will then be increased by D - 1 days. Our problem is to make a reasonable estimate of D.

Completed works which will lead us to a definitive estimate are hard to discover. The large global circulation models cannot be used, since they do not contain the smaller scales. Models of mesoscale or smaller-scale motions are generally too limited in an areal extent to contain the larger scales. The study[4] to which we shall turn is one in which we derived a system of second-order linear ordinary differential equations, whose dependent variables were the squared amplitudes of the wind errors in separate bands of the spectrum, each band spanning a single octave. Solutions of these equations depict the spread of errors from one scale to another.

The study is by no means ideal for our present task, partly because it was not intended primarily as an atmospheric study. The basic equation from which the ordinary differential equations were derived was the barotropic vorticity equation, which certainly does not approximate the laws governing the smaller atmospheric scales. To keep the equations manageable, the motion field was assumed to be homogeneous and isotropic. The deceleration of the error growth in each scale, which should have been brought about by nonlinear effects, was simulated by allowing the error growth to grow quasi-exponentially to a prechosen scale-dependent value, and then terminating its growth altogether. Viscosity and external forcing were omitted. Finally, a formula related to the discredited quasi-normal approximation was used to close the system.

The model was atmospheric to the extent that the prechosen spectrum of the unperturbed motion was modeled after the assumed atmospheric motion spectrum; this spectrum exerted a controlling influence on the time scale associated with each spatial scale. The principal results of the study are summarized in Figure 2, which is based on a figure in Reference (4), and shows the growth of an error field confined initially to the smallest scales. The upper curve is the assumed atmospheric spectrum, which also serves as an upper bound for the spectrum of the errors. Each curve labeled with a time (8 days, etc.) actually extends from the extreme left to the extreme right of the figure: to the left it is indistinguishable from the zero line, while to the right it is indistinguishable from the upper curve. These curves are the spectra of the errors at the indicated times.

In attempting to estimate D from the results in Figure 2, we must recall that the errors there are wind-field errors, while those in Figure 1 are height-field errors. The ratio of a wind error to a height error may be estimated geostrophically, and it is highly scale-dependent. We could scale down the right-hand portion of Figure 2 so that the curves would represent height spectra instead of wind spectra, but instead we shall circumvent the scale dependence by examining a particular scale.

In making the study[1] with the ECMWF model, we performed certain additional computations which were not described in the final write-up. These included a spectral analysis of the prediction errors. We found that the smallest scales in the archived data, with wave numbers near 40, were predicted moderately well at one day, rather poorly at two days, and not at all at three. These scales correspond to the 1250-625 km band in Figure 2. This band is indicated as being reasonably predictable at 1/2 day, slightly predictable at 1 day, and unpredictable at 1.5. Combining these results, we see that motions of this scale are already being predicted better than they can be. We are therefore faced with a contradiction, and the fault is presumably in the model which produced Figure 2.

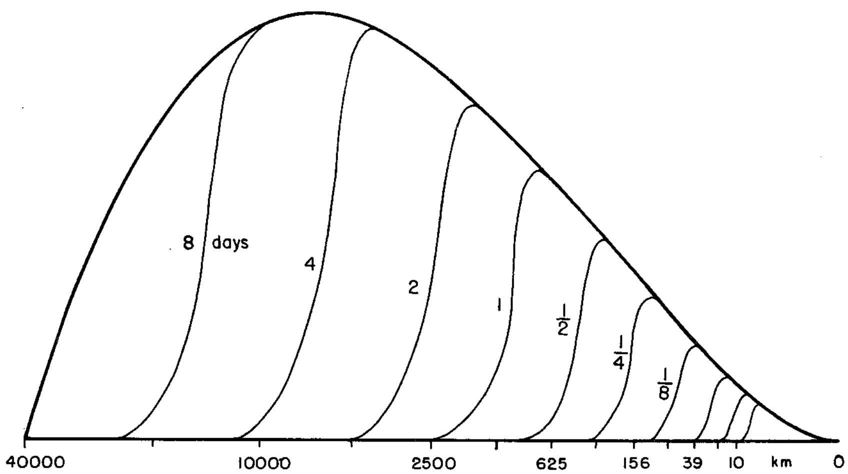

Fig. 2. Growth of errors initially confined to smallest scales, according to theoretical model. Upper heavy curve is assumed atmospheric motion spectrum; lower heavy curve is zero line. Thin curves are spectra of errors at indicated times; each thin curve coincides with lower heavy curve to the left, and upper heavy curve to the right. Areas are proportional to kinetic energy.

We have already enumerated some of the defects of the model, and it would appear possible to perform a new study with an improved model. Perhaps the model could include baroclinic effects, some inhomogeneity and anisotropy, and some forcing and damping. Perhaps the nonlinear effects could be incorporated more realistically. Perhaps a more realistic closure assumption, like those used in some subsequent works[5,6] could be introduced. However, the deduced predictability times are so highly dependent upon the assumed atmospheric spectrum that any of the above-mentioned improvements would be pointless until our estimate of the atmospheric spectrum has been made as realistic as possible.

In keeping with our intention of basing this study on the results of previously completed works, we shall turn to a study entitled "Limits of Meteorological Predictability", which we prepared in 1972 at the request of the American Meteorological Society. To the best of our knowledge the results were for internal use, and were never published. In that study we addressed the question of the effect of a possible spectral gap in the mesoscale band. We performed three sets of computations similar to those used to produce Figure 2: one with no spectral gap, one with a "weak gap", and one with a "strong gap". Although we are not even sure of the existence of a gap, let alone its structure, our "best guess" is something between the weak and strong gaps. Figure 3, which has the same format as Figure 2, has been constructed from an interpolation between the weak-gap and strong-gap computations. We see that errors initially confined to the smallest scales begin to spread up the scale just as in Figure 2, but, upon encountering the gap, they experience considerable difficulty in crossing it. Thus, it takes nearly five days for the error spectrum in Figure 3 to acquire the same form, outside of the gap, which it acquired in one day in Figure 2. Turning to motion in the 1250-625 km band, we see that it is moderately predictable at 4 days, slightly predictable at 5 days, and unpredictable at 6 days. Comparing this result with current skill in predicting these scales, we see that about 3 days can be added to the range at which they may be predicted. From this we conclude, very tentatively, that D = 4.

We believe that this model, crude as it may be, depicts fairly realistically the qualitative influence of a spectral gap on predictability. What may be quite unrealistic is the assumed spectrum.

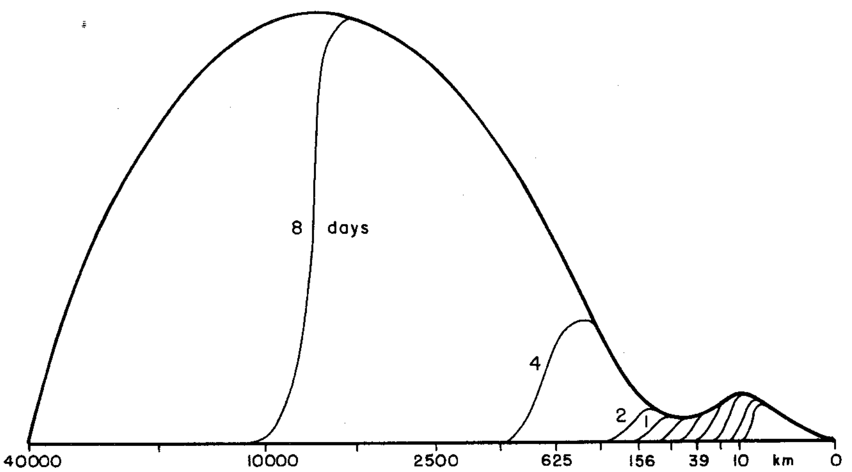

Fig. 3. Same as Figure 2, when assumed atmospheric motion spectrum has a moderately strong gap in the mesoscale band.

CONCLUDING REMARKS

We have examined some studies which together imply that major improvements in medium-range weather forecasting are possible, and,in particular, that the present forecasting success at one week may some day be realized at nearly two weeks. The studies contain estimates of the rate at which inevitable errors in the analysis will grow as the range of the forecast is extended, until they eventually render the forecast unacceptable. For the middle and later stages of error growth our estimates are on reasonably firm ground; for the early stage they are highly speculative.

The model which indicates that we might eventually forecast as well at four days as we can now forecast at one contains a rather arbitrarily chosen atmospheric spectrum, which strongly influences the numerical results. The model is also crude in other respects, and we believe that some computations with some other model, possibly a rather sophisticated mesoscale model, are in order. Nevertheless, we do not see how the final result can fail to depend upon the atmospheric spectrum which the model entails, whether it is prechosen on the basis of real observations or produced by the model itself. Since different models can produce different spectra, and since we must have confidence in the spectrum if we are to have confidence in the conclusions, it seems rather likely that the next significant refinement in our estimate of medium-range predictability will result from observations.

ACKNOWLEDGMENT

This research has been supported by the GARP Program of the Atmospheric Sciences Section, National Science Foundation, under Grant 82-14582 ATM.

REFERENCES

1. E.N. Lorenz, Tellus **34**, 505-513 (1982).
2. L. Bengtsson and A.J. Simmons, Large-scale dynamic processes in the atmosphere, B. Hoskins and R. Pearce, eds. (New York and London, Academic Press (1983).
3. J. Smagorinsky, Bull Amer. Meteor. Soc., **50**,286-311 (1969).
4. E.N. Lorenz, Tellus, **21**, 289-307 (1969).
5. C.E. Leith, J. Atmos. Sci., **28**, 148-161 (1971).
6. C.E. Leith and R.H. Kraichnan, J. Atmos. Sci., **29**, 1041-1058 (1972).