

EFFECTS OF ANALYSIS AND MODEL ERRORS ON ROUTINE WEATHER FORECASTS

E. N. Lorenz

Center for Meteorology and Physical Oceanography
Massachusetts Institute of Technology
Cambridge, U.S.A.

Summary: Since 1981 we have evaluated root-mean-square differences $E_{k\ell}$ between k -day and ℓ -day ECMWF forecasts made for the same day. A curve connecting values of $E_{0\ell}$ measures the over-all performance, while one connecting values of $E_{\ell-1,\ell}$ measures the performance that would occur if the model could be made perfect after the first day. For forecasts of the global 500-mb height, the $E_{0\ell}$ curve drops from 1981 to 1987, indicating improvement, but the $E_{\ell-1,\ell}$ curve drops by an equal amount, implying as much room for improvement as before.

We suggest that this apparent paradox may be resolved by assuming that while improvements in the model move the curves closer together, improvements in the analysis may move them farther apart. To test this suggestion we perform two simulations of the ECMWF study, one with a single difference equation as the exact equation, and one with a system of three ordinary differential equations, and with solutions of these equations as the exact data. We add analysis and model errors, and then construct curves connecting values of $E_{k\ell}$. We construct similar curves after reducing the analysis error alone, the model error alone, and both errors together. The results are consistent with our suggestion.

1. INTRODUCTION

The atmosphere is continually evolving from one state to another according to a set of physical laws. Routine weather forecasting consists of attempting to extrapolate from a known present state to presently unknown future states according to these laws. In reality we cannot know the present state perfectly; even if our measurements were perfect, they would not completely determine the weather between observing sites. Likewise, we cannot make perfect forward extrapolations; even if we knew the physical laws perfectly, we could not formulate them to apply perfectly to the smoothed fields with which we must work. Thus a typical weather forecast is subject to *analysis errors*—errors that are introduced in analyzing the observations before any forward extrapolation is performed—and *model errors*—additional errors introduced by the

procedure used for extrapolation. Presumably one cannot express the total error in a particular forecast as the sum of an analysis error and a model error; one can, for example, visualize instances where the model errors that accumulate during the first day will add to the analysis error and other instances where they will almost cancel it.

Identification of the way in which analysis and model errors contribute to the total prediction error has, in fact, been a problem of major concern at the European Centre for Medium Range Weather Forecasts (see Arpe et al., 1985). A solution to this problem would aid in determining where to concentrate the efforts toward further improvement. In what follows, for consistency with our definitions of analysis and model errors, we shall refer to what is sometimes called the ECMWF model as the ECMWF *system*, reserving the term *model* to refer only to the part of the system used in forward extrapolation. By the ECMWF *analysis* we shall mean everything that routinely precedes the extrapolation, including data assimilation, data quality control, and initialization.

With a good analysis and a good model we might be able to make fairly good forecasts far into the future, if it were not for another factor—the sensitive dependence of the variations of the atmosphere upon the precise state. That is, if at some time the state of the atmosphere should be perturbed by a small amount, the subsequent development would diverge from the development that would have occurred if there had been no perturbation, and the divergence might even continue until little resemblance remained between the states that would evolve from the perturbed and unperturbed earlier states. In recent years this sensitive dependence has been popularly called “chaos.” In particular, the unperturbed and perturbed states may be the exact state and the state as it has been analyzed or predicted to be. The total error at any time will thus be due not only to the introduction of analysis and model errors but also to the continual growth of whatever error has already accumulated.

In most studies of the rate at which initial errors will amplify, pairs or larger sets of numerical solutions of an atmospheric model, originating from rather similar but not identical states, have been obtained and compared. Recent studies have tended to yield an average doubling time of between two and three days for small root-mean-square errors in the temperature and wind fields. As the errors become larger their growth rate subsides, and finally vanishes as “saturation” is reached. Ideally one should work with as large and realistic a model as is available, and apply perturbations to each of a representative sample of initial states, but this inevitably requires vast amounts of computation.

By 1981 we were able to circumvent the problem of excessive computation by noting that the then recently formulated ECMWF operational system, in contrast to many earlier-generation systems, was performing well enough for a one-day forecast,

say for day 1, to be treated as the analysis for day 1 plus a reasonably small and readily evaluated error. The growth of this error during the following day could be found by evaluating the difference between the two-day and the one-day forecasts for day 2. More generally, the average growth, during k days, of errors having various initial sizes could be obtained by comparing the average difference between k -day and $(m+k)$ -day forecasts for the same day with the average difference between analyses and m -day forecasts for the same day, for various values of k and m . These comparisons could all be made by performing a few additional computations with already archived analysis and forecast data produced during routine operations. The study that we performed (Lorenz, 1982) will serve as a starting point for this presentation.

Fig. 1a summarizes the principal findings of that study. It is based on the global analyses and forecasts of the 500-mb height field for each day of the "1981 winter," defined here as the 100-day period beginning 1 December 1980. It shows root-mean-square differences $E_{k\ell}$ between k -day and ℓ -day forecasts for the same day, for $\ell = 1, \dots, 10$ and $k = 0, \dots, \ell - 1$, the averaging being performed over the globe and over the 100 days. Here a 0-day forecast means an analysis. The top curve connects values of $E_{0\ell}$, and represents the mean performance of the model. The thin curves connect values of $E_{\ell-m,\ell}$ for fixed values of m , and represent the growth of differences between separate solutions of the model equations. Leftward extrapolation of these curves suggests that small errors will double in about 2.4 days. The bottom curve may also be interpreted as the manner in which the system would perform if the model could be made perfect after the first day. The difference between this curve and the heavy curve therefore appears to be some sort of measure of the room for further improvement in the model.

In 1987 we examined data that did not exist when the original study was performed, and constructed diagrams similar to Fig. 1a; Fig. 1b is for the 1987 winter. The downward displacement of the top curve indicates definite improvement during the intervening years, but an equally large downward displacement of the bottom curve leads to the paradoxical conclusion that there is still as much room for improvement as before.

Soon after the oral presentation of this work, we acquired access to two more years of data. Curves similar to those in Fig. 1, not shown here, suggest continuing improvement, and reveal that half of the gap between the top and bottom curves has been eliminated. Nevertheless, the paradox presented by Fig. 1 remains. In this presentation we shall suggest that the paradox may be resolved by recognizing that the observed improvement in forecasting between 1981 and 1987 has resulted partly from improvement in the analysis. There is no obvious reason why reducing the analysis errors cannot push the top and bottom curves farther apart.

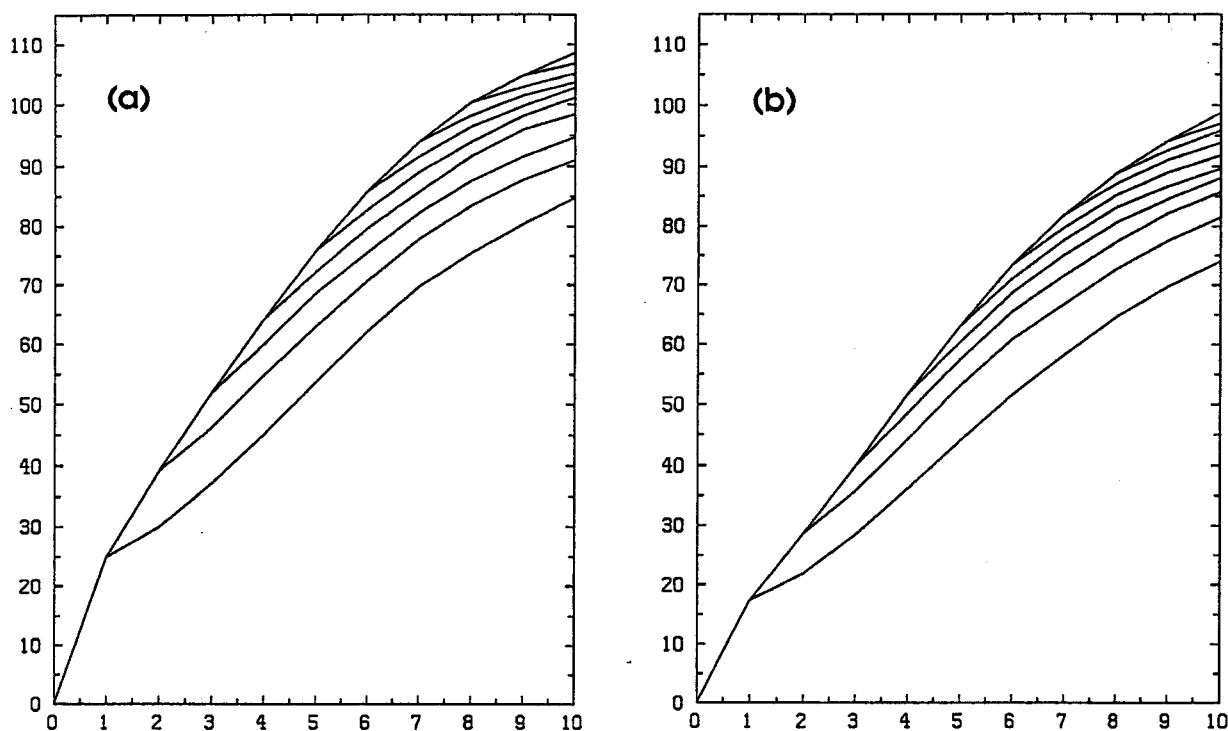


Fig. 1. a) Curves connecting values of the root-mean-square difference $E_{k\ell}$ between k -day and ℓ -day forecasts for the same day, by the ECMWF system, for the global 500-mb height field, averaged over the 100-day period beginning 1 December 1980. The scale at the left indicates the value of $E_{k\ell}$, while the scale at the base indicates the value of ℓ . For the top curve $k = 0$, while for the bottom curve $k = \ell - 1$, for the next curve upward $k = \ell - 2$, etc. b) The same, but for the 100-day period beginning 1 December 1986.

2. THE FIRST SIMULATION

To test our suggestion we shall perform some simulations of the ECMWF study, using models which are much simpler than the ECMWF model. In each simulation we shall choose a system of equations to be the "exact" equations, and let a time-dependent solution of the exact equations be the "exact" data, but then we shall add analysis errors to the data and model errors to the equations, and seek to construct curves similar to those in Fig. 1.

The rationale for a simulation is that we can alter the magnitude of the analysis or model errors at will, and then repeat the computations. We can thus determine how, at least in the simulations, the curves should respond to improvements in the analysis alone or the model alone or both. We shall not attempt to make the simulation models look much like an operational weather forecasting model, but an essential feature will be that they behave chaotically, whence the data sequences that they produce will be aperiodic.

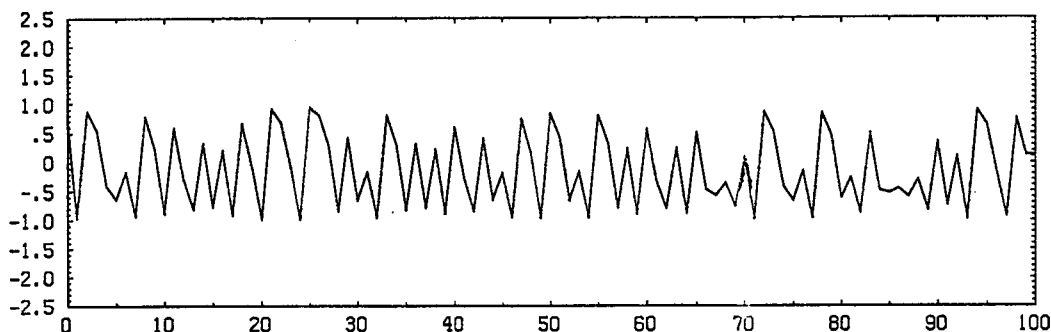


Fig. 2. Consecutive values of X_n up to $n = 100$ in a computed solution of eq. (1), with $X_0 = 3/4$. Because of the amplification of round-off errors the computed values presumably do not even approximate the true values beyond $n = 50$, but they are very close to the true values for another solution with X_0 very near $3/4$. The straight lines connecting the values have been introduced to make the chronological order easier to observe, and do not imply linear or any other type of variation of X between integral values of n .

In the first simulation we shall use what is probably the simplest and also the most thoroughly studied of all dynamical systems that behave chaotically. It is also one of the relatively few chaotic systems whose general solution can be expressed analytically, so that numerical solutions are not needed. It consists of the single difference equation

$$X_{n+1} = 2X_n^2 - 1 \quad (1)$$

in the single dependent variable X , the iteration subscript denoting the number of time units since some "initial" time. We shall take eq. (1) to be the exact equation.

To solve eq. (1) we let $X_n = \cos \theta_n$, and observe that to satisfy (1) it is sufficient to let

$$\theta_{n+1} = 2\theta_n, \quad (2)$$

whence, by iteration

$$\theta_n = 2^n \theta_0, \quad (3)$$

so that

$$X_n = \cos(2^n \cos^{-1} X_0). \quad (4)$$

Fig. 2 shows a typical solution of eq. (1), extending over 100 iterations. Short segments of the solution seem to fall into a few preferred patterns, but these do not follow one another in any regular order. The attractor of the system is the entire interval from -1 to $+1$.

We now let the analyzed value of X_n , with an analysis error, be $X_{n,0} = \cos \theta_{n,0}$, where

$$\theta_{n,0} = \theta_n + \delta_n, \quad (5)$$

θ_n is the exact value at time n , and the successive values of δ_n are chosen independently and randomly from a Gaussian distribution with mean 0 and standard deviation α . Likewise, we tentatively let the k -day prediction for X_n by the model, with a model error, be $X_{n,k} = \cos \theta_{n,k}$, where

$$\theta_{n+1,k+1} = 2\theta_{n,k} + \varepsilon, \quad (3)$$

and ε is a constant. From (6), (5), and (3) it follows that

$$\theta_{n,k} = \theta_n + 2^k \delta_{n-k} + (2^k - 1)\varepsilon. \quad (7)$$

For $\ell > k$, we define the mean-square difference between k -day and ℓ -day predictions as

$$E_{k\ell}^2 = \langle (X_{n,\ell} - X_{n,k})^2 \rangle = \langle (\cos \theta_{n,\ell} - \cos \theta_{n,k})^2 \rangle, \quad (8)$$

where pointed brackets denote an average over n .

One might object that choosing the analysis error δ in θ to be symmetrically distributed produces decidedly asymmetric distributions for the error in the basic variable X , except when $X = 0$, and, indeed, the distributions are entirely one sided when $X = 1$ or -1 . Our choice greatly simplifies the mathematics, but we can also defend it fairly well, if it needs justification, by the following scenario.

Consider a situation where a forecaster notes that observed values of a variable X are symmetrically distributed between 1.1 and -1.1, and where he is aware that there is an observational error, presumably independent of the true value of X and symmetrically distributed, that may reach 0.1 or -0.1. He will conclude that the true values of X lie between 1.0 and -1.0, since otherwise an error close to 0.1 (or -0.1) would sometimes accompany a true value surpassing 1.0 (or -1.0), producing an observed value not actually encountered. Before performing a forward extrapolation, he may then "initialize" the observed values of X by moving them closer to 0. Errors will still remain, but they will tend to be smaller, and for most values of X their distribution will be asymmetric.

The model as it stands has a more serious drawback. The accumulated model error $(2^k - 1)\varepsilon$ in $\theta_{n,k}$ will become large as n increases, but for some values of k it may be close to a multiple of 2π , so that the accompanying error in $X_{n,k}$ will be small. To eliminate this possibility we replace the tentative model by an ensemble of models, each the same as the original one, with a constant value of ε , but with different values of ε for different models in the ensemble; we let the values of ε have a Gaussian distribution with mean

0 and standard deviation β . The pointed brackets in (8) will now denote averages over the ensemble and over n . Substituting (7) into (8), and noting in performing the averaging that averages of sines vanish, we find that

$$E_{k\ell}^2 = 1.0 - \exp \left[-\frac{1}{2}(2^{2\ell} + 2^{2k})\alpha^2 - \frac{1}{2}(2^\ell - 2^k)^2\beta^2 \right]. \quad (9)$$

A further problem is that, as with any difference equation, the variables are defined only at discrete times. This would be of little consequence if the times were close together, but, in this system, small errors double on the average in one iteration, so that the time unit is one doubling time, which corresponds to 2 or 3 days for the ECMWF model. We shall assume that the successive values of X_n occur at 3-day intervals.

We cannot interpolate values of X between iterations uniquely; eq. (4) might seem capable of defining X_n for nonintegral values of n , but a particular value of X_0 has many inverse cosines differing by multiples of 2π , and hence producing different values of X_n except when 2^n is an integer. This difficulty does not, however, enter the derived equation (9), which defines $E_{k\ell}$ as a smooth function of k and ℓ , and therefore offers a logical means of interpolating to fractional values of k and ℓ . In constructing diagrams like Fig. 1, we shall therefore evaluate $E_{k\ell}$ for values of k and ℓ that are multiples of $1/3$, up to $10/3$.

Suitable values of α and β are best found by trial and error; we need values that produce top and bottom curves separated somewhat as those in Fig. 1. We have chosen $\alpha = 1/8$ and $\beta = 1/4$ for our basic simulation.

Fig. 3a shows the resulting curves. There is a fair resemblance to Fig. 1a, with some obvious differences. Saturation occurs at $E = 1.0$. The failure of $E_{0\ell}$ to approach zero as $\ell \rightarrow 0$ may be attributed to the implicit assumption that all analysis errors are independent.

Fig. 3b has been constructed with the same analysis error, but with the model error cut in half, i.e., $\alpha = 1/8$, $\beta = 1/8$. The top and bottom curves now come closer to coinciding, and, in fact, the $E_{0\ell}$ curve is on the bottom. Qualitatively this is what we had anticipated would result from a noticeably improved model. In Fig. 3c, on the other hand, we have cut the analysis error in half while keeping the model error as in Fig. 1a; thus $\alpha = 1/16$ and $\beta = 1/4$. The separation between the top and bottom curves has now decidedly increased. Finally, in Fig. 3d, we let $\alpha = 1/16$ and $\beta = 1/8$. The separation is again more like that in Fig. 3a, although somewhat greater.

It thus appears that improvements in the analysis and the model have opposite effects on the separation between the curves. A lowering of both curves with little change in the separation, such as occurred with the ECMWF system between 1981 and 1987, would therefore seem to imply some improvement in both the analysis and the

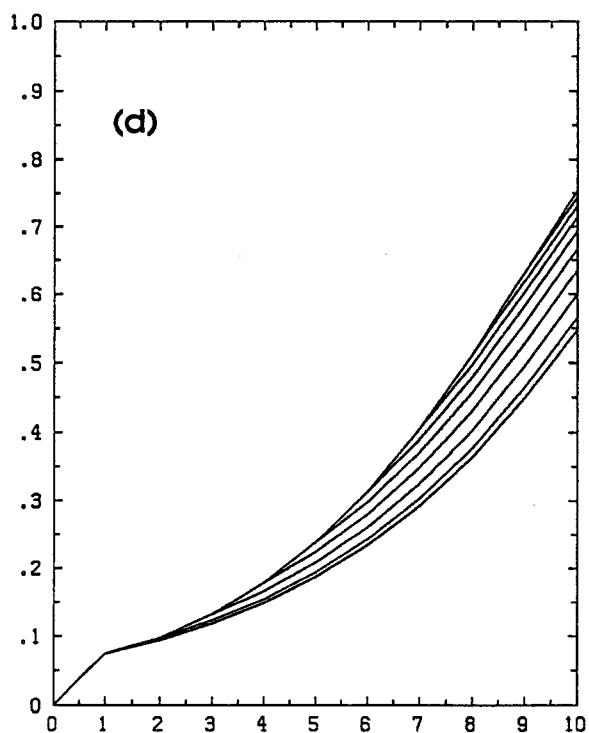
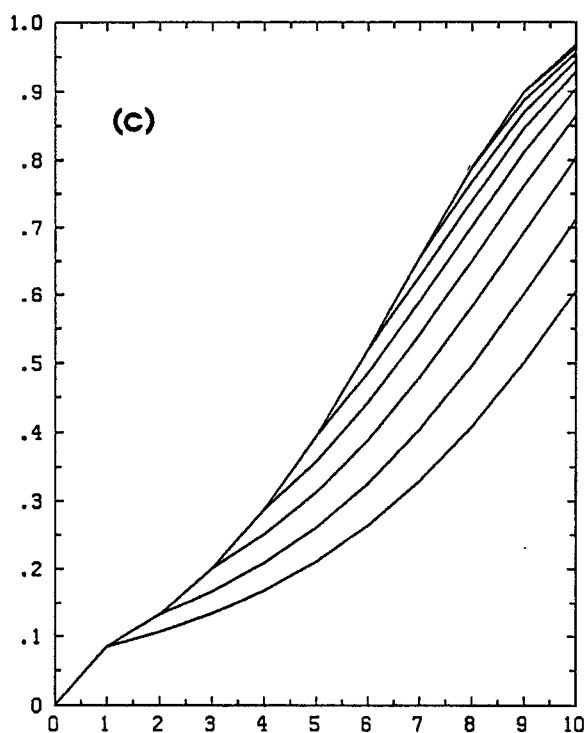
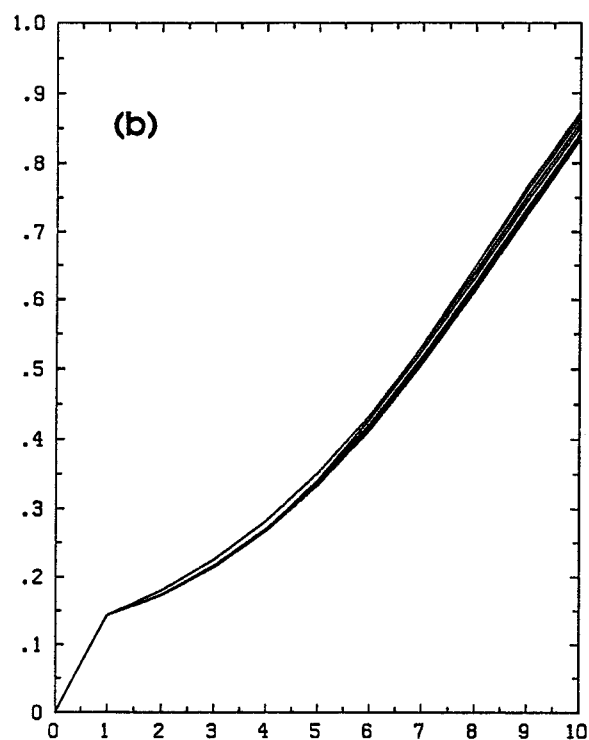
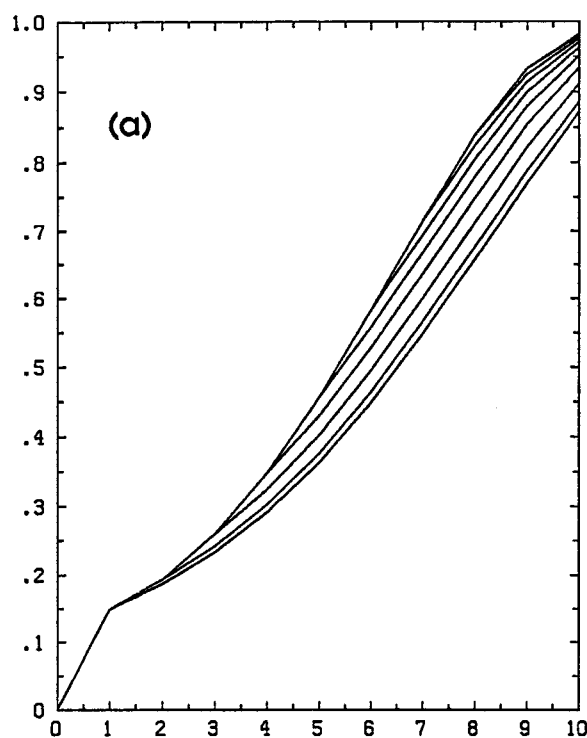


Fig. 3. The same as Fig. 1, but for the first simulation, with mean analysis error α and mean model error β . a) The basic simulation; $\alpha = 1/8$, $\beta = 1/4$. b) $\alpha = 1/8$, $\beta = 1/8$. c) $\alpha = 1/16$, $\beta = 1/4$. d) $\alpha = 1/16$, $\beta = 1/8$.

model. A reduction of the separation, such as occurred from 1987 to 1989, might imply that the principal improvements were in the model.

We observe also, by examining the top curves in Figs. 3c and 3d and the bottom curve in Fig. 3b, that improving the analysis is much more effective than improving the model for reducing the prediction error $E_{0,1}$ at 1 day, while the opposite is true for reducing the error $E_{0,10}$ at 10 days. In addition, the improvement at 10 days that results from reducing the analysis and model errors together is decidedly greater than the sum of the the improvements that result from reducing these errors separately. This would remain true if we had used the mean-square error instead of the root-mean-square error as a measure.

3. THE SECOND SIMULATION

In order to achieve the foregoing results from a simulation based on an equation as simple as eq. (1), we were forced to replace the tentative model, with its model error, by an ensemble of models, and afterward to interpolate between computed values of E_{kt} . In the next simulation we avoid these drawbacks by using differential instead of difference equations, and letting these equations govern several dependent variables instead of one. We let the exact equations be

$$dX/dt = -Y^2 - Z^2 - aX + aF, \quad (10a)$$

$$dY/dt = XY - bXZ - Y + G, \quad (10b)$$

$$dZ/dt = bXY + XZ - Z, \quad (10c)$$

governing the exact values X , Y , Z .

These equations were originally designed as a minimal general-circulation model (Lorenz, 1984). The variable X represents the strength of a circumpolar westerly wind current, while Y and Z are cosine and sine phases of a superposed large-scale wave. The variables have been scaled so that the time unit is the dissipation time for Y and Z , which we shall take to be 4 days. For certain values of the parameters the general solution is chaotic. With $a = 1/4$, $b = 6$, $F = 16$, and $G = 3$, typical variations of X are like those shown in Fig. 4, which was produced by numerically integrating eqs. (10), using a fourth-order Taylor-series scheme with a time step of 0.025 units, or 2.4 hours. There is a noticeable tendency for successive maxima, or minima, to be about 10 days apart, but there are also stretches, sometimes lasting several months, where the oscillations between maxima and minima have a large amplitude; these are separated by generally shorter stretches where the oscillations are much weaker. Transitions between large-amplitude and small-amplitude oscillation regimes do not appear to occur at regularly spaced intervals.

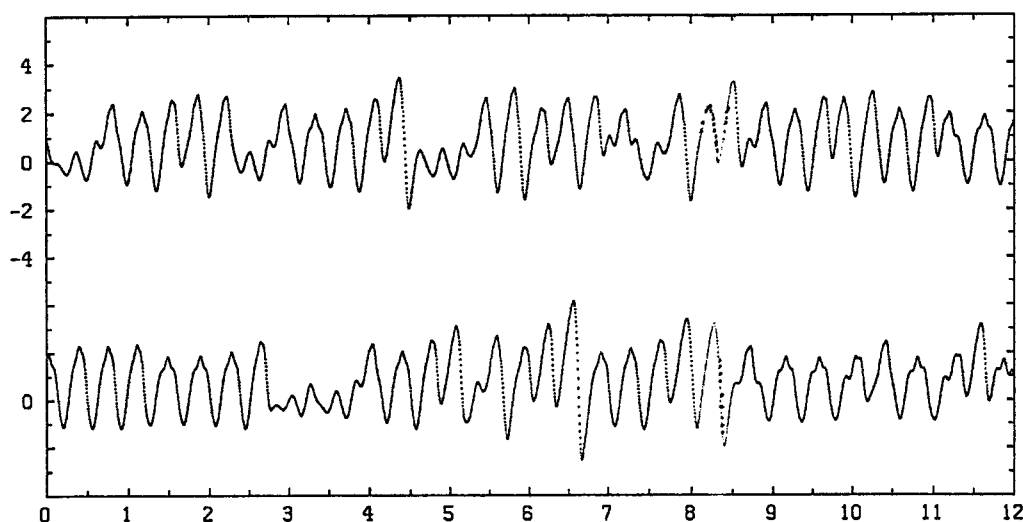


Fig. 4. Values of X in a numerical solution of eqs. (10), with $a = 1/4$, $b = 4$, $F = 16$, and $G = 3$ extending over two years. The lower curve is a continuation of the upper curve. The scale at the base indicates time in months.

Fig. 5 shows a cross-section of the attractor of eqs. (10). It was constructed by plotting (X, Y) at 8000 consecutive zero-crossings of Z , occurring during 45 years. The Cantor-set structure typical of strange attractors is apparent, while the marked variations in shading reveal that some regions are visited far more often than others.

To simulate the analyzed values, with their errors, we first make a 10-year run with eqs. (10), and denote the values of X , Y , Z at the end of the n^{th} day by X_n , Y_n , Z_n . For each day n , we then choose three numbers randomly from a uniform distribution between -1 and $+1$, and let x_n , y_n , z_n be the ratios of these numbers to the square root of the sum of their squares. We then let the analyzed values of the variables be

$$X_{n,0} = X_n + \alpha x_n, \quad (11a)$$

$$Y_{n,0} = Y_n + \alpha y_n, \quad (11b)$$

$$Z_{n,0} = Z_n + \alpha z_n, \quad (11c)$$

where α is a constant, independent of n .

To simulate the model error we simply replace F in (10a) by $F - \beta$, where β is a constant. We then obtain the k -day predictions $X_{n,k}$, $Y_{n,k}$, $Z_{n,k}$ for X_n , Y_n , Z_n by numerically integrating eqs. (10), starting from $X_{n-k,0}$, $Y_{n-k,0}$, $Z_{n-k,0}$.

For the basic simulation we let $\alpha = 0.224$ (1/40 times the root-mean-square distance from the centroid of the attractor) and $\beta = -3$. Fig. 6a, analogous to Fig. 3a, shows the curves. There is a decidedly closer resemblance to Figs. 1a and 1b than with the first simulation, suggesting that the second simulation may be generally superior.

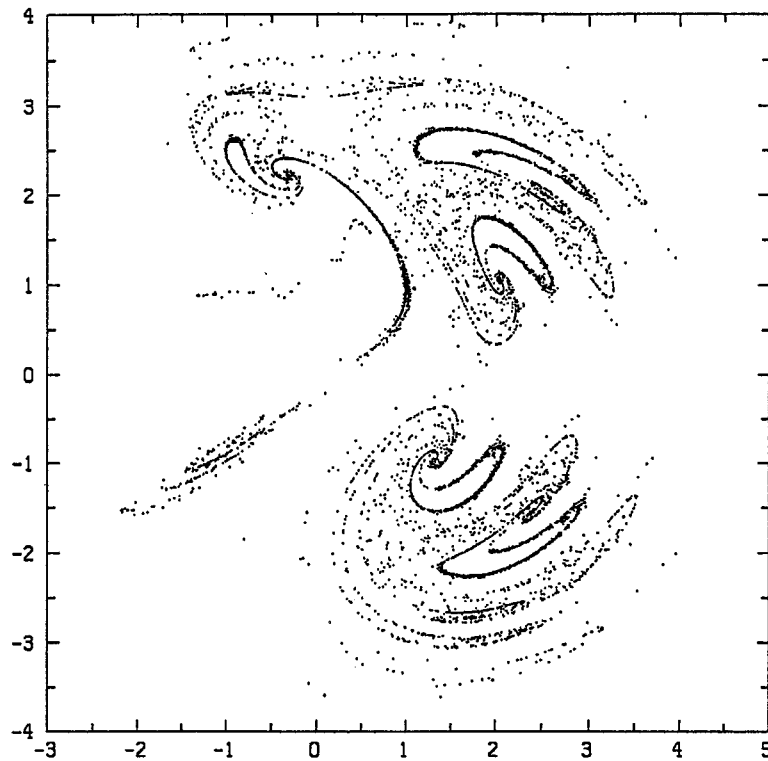


Fig. 5. A cross-section of the attractor of eqs. (10), with $a = 1/4$, $b = 6$, $F = 16$, and $G = 3$, as represented by values of X and Y at 8000 successive zero-crossings of Z , in a numerical solution extending over 45 years.

Figs. 6b, 6c, and 6d result from multiplying just β , just α , and both α and β , respectively, by $1/2$; they are analogous to Figs. 3b–3d. Again the top and bottom curves move closer together when the model alone is improved, and farther apart when the analysis alone is improved. Again, the shorter-range predictions are improved more by reducing the analysis error, while the more extended-range predictions respond more to reductions in the model error. Again the improvement at the ten-day range realized by reducing α and β together is greater than the sum of the improvements made by reducing α and β separately.

The more-than-superficial resemblance between Figs. 6c and 6a, and between Figs. 6d and 6b, suggests that we have been working in a range of α and β where prediction errors are more sensitive to model changes than analysis changes. Whatever the shortcomings of any simulation with a simple system may be, the findings of the second simulation suggest that replacing the model by an ensemble of models in the first simulation, and interpolating between computations, did not seriously harm the qualitative results.

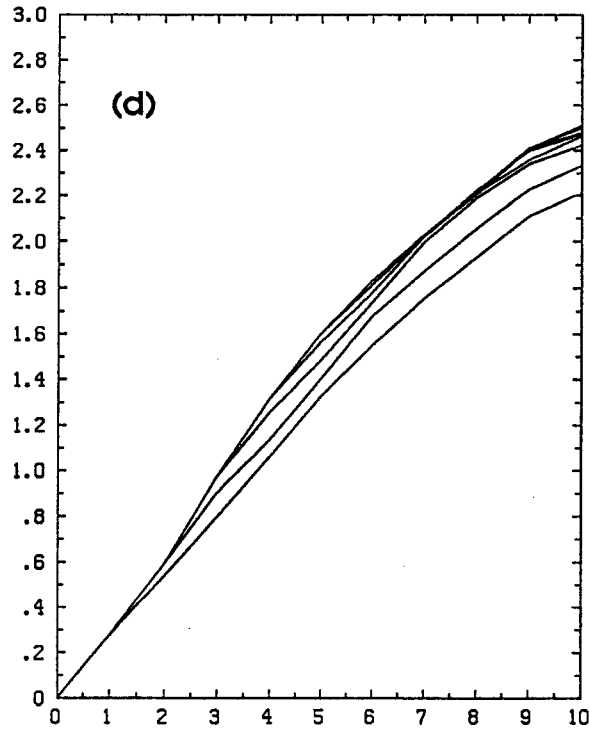
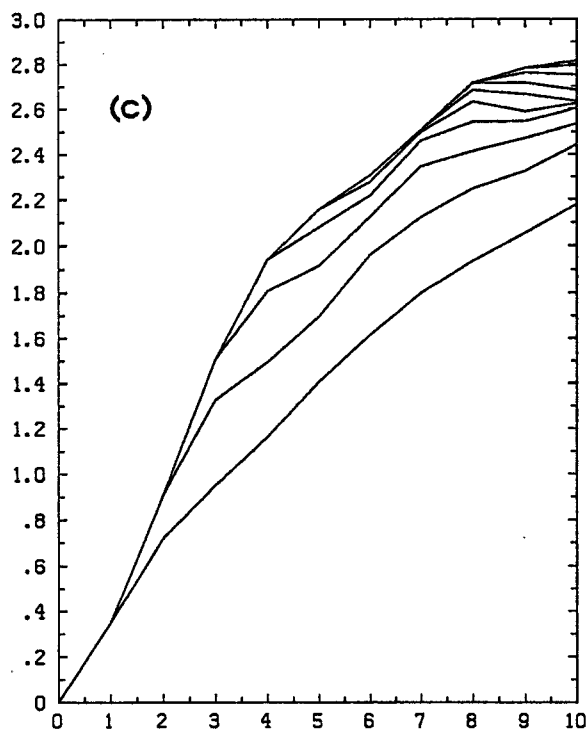
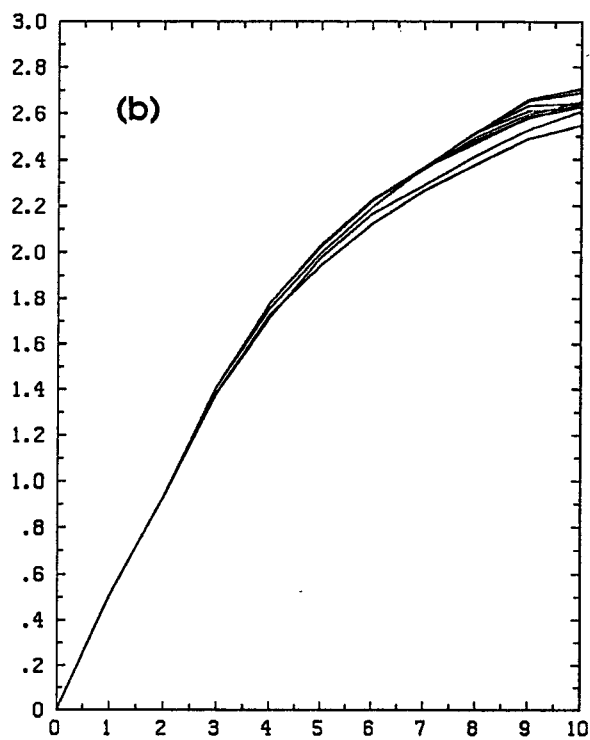
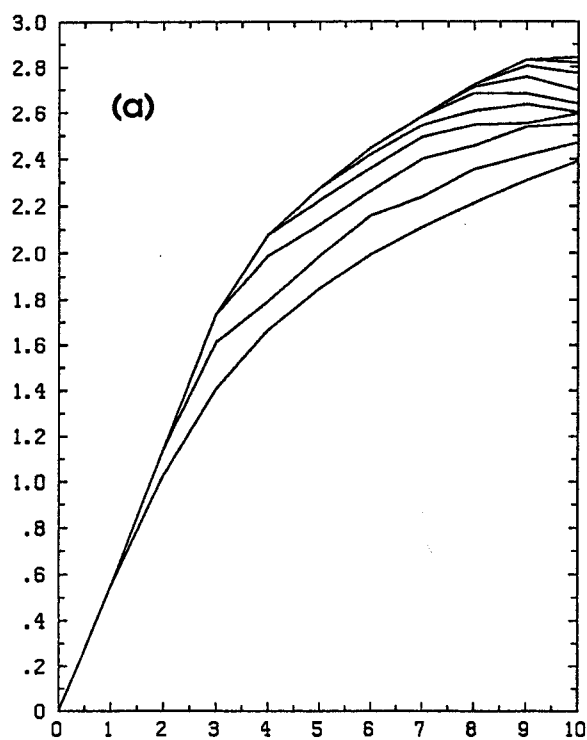


Fig. 6. The same as Fig. 1, but for the second simulation, with mean analysis error α and model error β . a) The basic simulation; $\alpha = 0.224$, $\beta = -3.0$. b) $\alpha = 0.224$, $\beta = -1.5$. c) $\alpha = 0.112$, $\beta = -3.0$. d) $\alpha = 0.112$, $\beta = -1.5$.

4. CONCLUSIONS

Since 1981 we have been constructing diagrams showing root-mean-square differences $E_{k\ell}$ between k -day and ℓ -day forecasts produced for the same day by the ECMWF operational system, for $k = 0, \dots, 9$ and $\ell = k + 1, \dots, 10$. The curve connecting values of $E_{0\ell}$ indicates the performance of the system, while curves connecting values of $E_{k\ell}$ for constant values of $\ell - k$ indicate the growth rate of differences between solutions of the model equations, which, it is hoped, approximates the growth rate of differences between solutions of the exact equations. In the present study we have formally repeated the computations of $E_{k\ell}$ and the construction of the curves, using two very simple models in place of the ECMWF model, and including analysis errors and model errors. We have then repeated the procedure, using reduced values of the model error, the analysis error, or both.

The basic finding is that the curves respond quite differently to improvements in the analysis and improvements in the model. The former tend to move the $E_{0\ell}$ and $E_{k,k+1}$ curves farther apart, while the latter move them closer together. In addition, at short range, improvements in the analysis are likely to be most effective, while, at more-extended range, improvements in the model play a greater role. It appears, moreover, that the benefit to be gained by improving the analysis, or the model, by a given amount, will be greater if the model, or the analysis, has already been subjected to some improvement.

It would be of obvious value to know whether the results of the simulations are qualitatively valid for the ECMWF system. Wholly apart from the fact that the equations in the simulations cannot describe real atmospheric behavior, there are some assumptions that affect the results but may be unrealistic. For example, the analysis errors on successive days may not be independent of each other, or of the states on which they are superposed.

The ideal way to determine the response of the curves to improvements in the system would be to introduce the improvements and see what happens. This is of course impossible; the system is a state-of-the-art system, and specific alterations are continually being made as soon as they are deemed to be beneficial. Great over-all improvement has resulted, but one can not generally tell, even after the fact, whether a particular alteration has yielded a greater improvement than some other alteration, or any improvement at all.

In principle there is a straightforward method for comparing the effects of analysis and model changes, although computationally it would probably be judged prohibitively expensive. It should not be difficult to make both the analyses and the model systematically worse. Having done so, we may repeat the routine forecasting procedure, perhaps for one winter season, and construct curves like those in Fig. 1. We may then

introduce improvements, and evaluate their effects, simply by returning part way or all the way to the present state of the art.

Acknowledgments. Most of this work was supported by the Climate Dynamics Program of the Division of Atmospheric Sciences, National Science Foundation, under Grant ATM 85-15010. The work involving the second simulation was performed while the author was a summer visitor at the National Center for Atmospheric Research; NCAR is supported by the National Science Foundation.

5. REFERENCES

Arpe, K., A. Hollingsworth, M. S. Tracton, A. C. Lorenc, S. Uppala, and P. Kallberg, 1985: The response of numerical weather prediction systems to FGGE level IIb data. Part II: Forecast verifications and implications for predictability. *Quart. J. Roy. Meteor. Soc.*, **111**, 67-101.

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

Lorenz, E. N., 1984: Irregularity: a fundamental property of the atmosphere. *Tellus*, **36A**, 98-110.