copies
available

The Growth of Errors in Prediction.

E. N. LORENZ

Massachusetts Institute of Technology - Cambridge, MA

PART I

General Aspects of Error Growth.

1. – Introductory remarks.

Among the innumerable systems which exist in Nature, in the laboratory, or as mathematical abstractions, some are convergent, or stable, while others are divergent, or unstable. By a stable system we mean one whose future succession of states, if the present state should be slightly disturbed, will converge toward the succession of states which would have occurred if there had been no disturbance. By an unstable system we mean just the opposite—a system whose future following a slight disturbance will diverge from what its future would have been without a disturbance.

Stable and unstable systems may be very simple. Consider, for example, a smooth plain in which there is a single bowl-shaped depression. Consider the motion of a ball, which is placed somewhere in the depression and allowed to roll until friction stops it. It will ultimately come to rest at the bottom of the depression. If it had been placed in a slightly different location in the depression, it would still have come to rest at the bottom. Equivalently, if two identical balls had been placed in slightly different locations, they would have come to rest at the same place. The system is stable.

Consider next a smooth plain from which there rises a single dome-shaped hill. If a ball is placed somewhere on the hill and is allowed to roll, it will come to rest somewhere on the plain. If a second ball is placed in a slightly different location on the hill, it may come to rest at a considerably different location on the plain, particularly if the initial locations are near the top of the hill. The system is unstable.

In the former example we do not need a detailed knowledge of the laws of motion to predict the final location of the ball. It is sufficient to know that the ball will seek the lowest point. Moreover, we do not need to know the initial location. It is sufficient to know that it is in the depresssion rather than on the plain. The accuracy of our prediction will be limited only by the degree of precision with which we can locate the bottom of the depression.

In the latter example our prediction of the final location depends upon the details of the laws of motion, including the manner in which friction acts. This does not mean that we must know the laws; we could, for example, have previously placed many balls in different locations on the hill and have learned by experience what final location corresponds to what initial location. We must, however, know the initial location of the ball. If the precision of our measurements allows us to say only that the ball is initially in some small region on the hill, we can say only that the final location will be somewhere in a much larger region on the plain. It is thus apparent that stability favors predictability, while instability opposes it.

Most systems are much more complicated than the two which we have described. Many of them are somewhat analogous to a ball rolling on an undulating surface with numerous depressions and hills of different sizes. Here there are additional possibilities for instability; for example, two balls rolling down a hill along slightly different paths may subsequently encounter another hill and be deflected by its curvature into widely diverging paths. This increases the likelihood that they will ultimately come to rest in the bottoms of different depressions.

2. – The concept of error growth.

In the above examples we may define the state of the system at any particular time as the combined position and velocity of the ball at that time. We may represent the state by a set of four numbers—two position components and two velocity components. Likewise, we may represent the laws of motion, applied to the system, by a set of four first-order ordinary differential equations. Strictly speaking, our systems are more complicated; in addition to rolling, the ball may be spinning about an axis perpendicular to the ground, and its spin may influence its future path. More general systems often require hundreds or thousands of numbers to represent their states, even approximately, and an equal number of ordinary differential equations to represent the governing laws. Alternatively, the laws may sometimes be represented by a relatively small number of partial differential equations.

We shall define an error as the difference between two possible states of a system. In our examples an error may be represented by four numbers, obtained by subtracting the numbers representing one state from those representing the other. The logic of this definition becomes apparent when we consider the case where one state is the true state, and the other is the state which has been observed to exist or is believed to exist, with the inevitable lack of perfect precision in the observations. If each state subsequently varies according to the physical laws, the future error becomes the error in prediction which we would make, using an optimal prediction procedure. The decay or growth of errors, therefore, influences the possible accuracy of predictions. In our first example the ultimate error is merely the error in determining the lowest point in the depression; in the second example it is the difference between two points in a possibly extensive region of the plain.

It is sometimes useful to extend the definition of an error to apply to the difference between states of two different systems. The systems must, of course, be enough alike for the state of one to be subtractable from the state of the other. In our examples the second system could be one where a ball of a different mass or radius is allowed to roll. For practical purposes two systems are different only if the equations governing their states are different. The usefulness of the extended definition becomes evident when one initial state is the true state of a real system, and its equations are the true equations, while the other is the assumed state, and its equations are approximations to the true equations.

To study the predictability of a system, we may investigate the growth or decay of errors. We need not know in advance whether the most feasible method of predicting the behavior of the system actually involves starting out from an initial state. If our investigation indicates that errors will decay, it will tell us that some simpler method is probably available. If it indicates that errors will grow, it will imply that we cannot predict the future without knowing the present or some recent past state.

The systems in our examples are somewhat specialized in that the motion of the ball is a strictly transient phenomenon. In each example the system eventually acquires a steady state. Geophysical fluid systems where prediction is of interest, such as the atmosphere, undergo fluctuations which never cease. It is systems of this sort whose predictability will be our primary concern. Despite the differences, the concepts of stability and instability, and decay and growth of errors, are still relevant.

3. – Simple numerical examples.

Before presenting a general treatment of error growth in systems whose states continue to vary, we shall consider some simple numerical examples. In each of these the state is defined by a single variable $x(t)$, and its variation with time t is given by the quadratic difference equation

$$x_{n+1} = x_n^2 - c \tag{1}$$

rather than by a differential equation, where $x_n = x(t_n)$ and $t_0, t_1, t_2, \dots$ is a sequence of times. Once the constant c has been specified, an initial value x_0 determines a sequence $x_0, x_1, x_2, \ldots$. If c lies between 0 and 2, and x_0 lies between $-c$ and $+c$, x_n will lie between $-c$ and $+c$. Because a state is defined by a single number, the example cannot illustrate all aspects of error growth. Equation (1) has received much recent attention from mathematicians because of the many types of solutions which it exhibits [1, 2].

TABLE I. – *Particular solutions x_n and y_n of quadratic difference equation $x_{n+1} = x_n^2 - c$, with $c = 1.2$, difference $y_n - x_n$, particular solution z_n with $c = 1.201$, and difference $z_n - x_n$.* All values have been multiplied by 10 000.

n	x_n	y_n	$y_n - x_n$	z_n	$z_n - x_n$
0	5 000	5 010	10	5 000	0
1	− 9 500	− 9 490	10	− 9 510	−10
2	− 2 975	− 2 994	−19	− 2 966	9
3	−11 115	−11 104	11	−11 130	−15
4	354	329	−25	378	24
5	−11 987	−11 989	− 2	−11 996	− 9
6	2 370	2 374	4	2 380	10
7	−11 438	−11 436	2	−11 444	− 6
8	1 084	1 079	− 5	1 086	2
9	−11 883	−11 884	− 1	−11 892	− 9
10	2 120	2 122	2	2 132	12
11	−11 551	−11 550	1	−11 555	− 4
12	1 342	1 340	− 2	1 343	1
13	−11 820	−11 821	− 1	−11 830	−10
14	1 971	1 972	1	1 984	13
15	−11 612	−11 611	1	−11 616	− 4
16	1 483	1 481	− 2	1 484	1
77	−11 708	−11 708	0	−11 716	− 8
78	1 708	1 708	0	1 716	8
79	−11 708	−11 708	0	−11 716	− 8
80	1 708	1 708	0	1 716	8

In our first example $c = 1.2$. Table I compares a basic or « true » solution x_n where $x_0 = 0.5$ with a perturbed or « predicted » solution y_n where y_0 has been « observed » to be 0.501. The error $y_n - x_n$ is also shown. If the example represented a real physical system whose state could be measured, the « observational » error 0.001 might not be unreasonably large.

We see that the error amplifies several fold in the first few steps, but subsequently undergoes damped oscillations. Before step 80 the solutions are alike to four decimal places. Evidently the system is stable.

Table I also compares x_n with a solution z_n of (1), with $c = 1.201$. In a real system involving physical constants the error 0.001 in determining one constant c might not be unreasonable. Again, the initial error growth soon ceases, and the error ultimately oscillates between -0.0008 and $+0.0008$.

It is apparent that with this stable system we do not need to know the governing law exactly to make reasonably good predictions far in advance. Neither do we need to know the initial state very closely; it is sufficient to know that x_0 lies between -0.6 and $+0.6$.

In our second example $c = 1.8$. Table II is similar in format to table I. We see that the initially small error $y_n - x_n$ proceeds to grow rather irregularly, generally gaining an order of magnitude in about five steps, until it becomes comparable to x_n itself. At this point the prediction y_n for x_n has become

TABLE II. – *Same as table I, with $c = 1.8$ for x_n and y_n, and $c = 1.801$ for z_n.*

n	x_n	y_n	$y_n - x_n$	z_n	$z_n - x_n$
0	5 000	5 010	10	5 000	0
1	−15 500	−15 490	10	−15 510	− 10
2	6 025	5 994	− 31	6 046	21
3	−14 370	−14 407	− 37	−14 355	15
4	2 650	2 757	107	2 595	− 55
5	−17 298	−17 240	58	−17 336	− 38
6	11 922	11 722	− 200	12 045	123
7	− 3 786	− 4 260	− 474	− 3 502	284
8	−16 566	−16 185	381	−16 784	− 218
9	9 444	8 196	− 1 248	10 160	716
10	− 9 080	−11 282	− 2 202	− 7 688	1 392
11	− 9 755	− 5 271	4 484	−12 100	− 2 345
12	− 8 484	−15 222	− 6 738	− 3 370	5 114
13	−10 802	5 170	15 972	−16 874	− 6 072
14	− 6 331	−15 328	− 8 997	10 465	16 796
15	−13 992	5 493	19 485	− 7 059	6 933
16	1 578	−14 982	− 16 560	−13 227	−14 805
17	−17 751	4 447	22 198	− 1 041	16 710
18	13 510	−16 023	−29 533	−17 902	−31 412
19	253	7 672	7 419	14 037	13 785
20	−17 994	−12 114	5 880	1 693	19 687

worthless. The system is patently unstable. Of course, the error cannot amplify forever, since both y_n and x_n remain bounded.

If $z_0 = x_0$, but c is taken to be 1.801 instead of 1.8, the situation is similar. The error amplifies at roughly the same rate and levels off similarly. Evidently it makes little difference whether the original uncertainty is in the observed state or in the governing law.

It is evident that the temporary growth rate of $y_n - x_n$ or $z_n - x_n$ is highly dependent on the true state x_n. In a large system with many components, some components might undergo their more rapid growth, while others undergo their slower growth, and the overall growth might be smoother. With the present system we can produce smooth growth by averaging a large ensemble of solutions.

To obtain a representative ensemble, we have extended the solution x_n in table II to 10 000 steps, subsequently using each of the 10 000 values x_n as an

TABLE III. – *Geometric mean ε_n of absolute values of differences $y_n - x_n$ between* 10 000 *particular solutions x_n and corresponding solutions y_n, with $y_0 = x_0 + 0.001$, of quadratic difference equation $x_{n+1} = x_n^2 - 1.8$.* All values have been multiplied by 10 000.

n	ε_n	n	ε_n
0	10	20	5 534
1	15	21	5 845
2	23	22	6 130
3	34	23	6 380
4	51	24	6 639
5	76	25	6 752
6	114	26	6 901
7	172	27	6 816
8	258	28	6 850
9	385	29	7 034
10	578	30	7 101
11	853	31	7 134
12	1 273	32	7 244
13	1 794	33	7 258
14	2 358	34	7 277
15	2 904	35	7 226
16	3 430	36	7 282
17	4 138	37	7 407
18	4 918	38	7 540
19	5 406	39	7 526

initial value in a new solution. Each new initial value is perturbed by adding an error 0.001. Table III shows the behavior of the average error. At first the growth is almost perfectly exponential, with an amplification factor of about 1.5 per step. Later the growth subsides, and by step 40 it has nearly ceased. The average is actually a geometric mean; the arithmetic mean would grow somewhat less smoothly.

A noteworthy feature of table I is that each solution asymptotically approaches a periodic sequence. This behavior is, in fact, demanded by the stability of the system. Since x_n must be between $-c$ and $+c$, a value x_M closely approximating a previous value x_L must occur in due time. Because

TABLE IV. – *Same as table I, with $c = 1.5$ for x_n and y_n, and $c = 1.501$ for z_n.*

n	x_n	y_n	$y_n - x_n$	z_n	$z_n - x_n$
0	5 000	5 010	10	5 000	0
1	−12 500	−12 490	10	−12 510	− 10
2	625	600	− 25	640	15
3	−14 961	−14 964	− 3	−14 969	− 8
4	7 383	7 392	9	7 397	14
5	− 9 549	− 9 536	13	− 9 538	11
6	− 5 881	− 5 907	− 26	− 5 912	− 31
7	−11 541	−11 511	30	−11 514	27
8	− 1 680	− 1 751	− 71	− 1 752	− 72
9	−14 718	−14 693	25	−14 703	15
10	6 661	6 590	− 71	6 608	− 53
11	−10 563	−10 637	− 74	−10 643	− 80
12	− 3 841	− 3 642	199	− 3 682	159
13	−13 524	−13 674	− 150	−13 655	− 131
14	3 291	3 697	406	3 635	344
15	−13 917	−13 633	284	−13 689	228
16	4 368	3 587	− 781	3 728	− 640
17	−13 092	−13 713	− 621	−13 620	− 528
18	2 139	3 806	1 667	3 540	1 401
19	−14 542	−13 552	990	−13 757	785
20	6 148	3 365	−2 783	3 915	−2 233
21	−11 220	−13 868	−2 648	−13 477	−2 257
22	− 2 411	4 231	6 642	3 154	5 565
23	−14 419	−13 210	1 209	−14 016	403
24	5 790	2 449	−3 341	4 633	−1 157

of the stability the solutions u_n with $u_0 = x_L$ and v_n with $v_0 = x_M$ will approach one another, *i.e.* the sequence x_n will eventually be almost unchanged by replacing n by $n + M - L$, and the solution will be periodic, with period $M - L$. In our example $M - L = 2$. In the unstable solutions in table II there is no evidence of periodicity.

The most general behavior may be a superposition of periodicity and aperiodicity. To illustrate this possibility, we present a third example, with $c = 1.5$. Table IV shows basic and perturbed solutions x_n and y_n and an additional solution z_n with $c = 1.501$. Again we observe an irregular but unmistakable amplification, implying that the system is unstable, but the amplification ceases well before $y_n - x_n$ or $z_n - x_n$ is comparable to x_n. There is a continual oscillation between small negative or positive values of x_n when n is even and large negative values when n is odd. To someone unaware of this periodicity, y_n and z_n might seem to be moderately good predictions for x_n, for all values of n. If, however, the periodicity is removed from x_n, y_n and z_n by subtracting the average value for all the even, or odd, numbered steps from each value at an even, or odd, step, there remain three aperiodic sequences, the second and third of which do not constitute good predictions for the first in the distant future.

4. – More general systems.

We now consider error growth in more general systems. First of all, the result that stability creates periodicity still holds, provided that the system is one in which analogues, *i.e.* close approaches to previous states, must occur [3, 4]. A sufficient but not necessary condition for the occurrence of analogues is that the system may be represented by a finite number of numbers, each having a finite range. An equivalent result is that any system which is observed to very aperiodically must be unstable.

In general an unstable system will possess some periodic solutions, but, if these are even slightly disturbed, the periodicity will disappear. The unstable system defined by eq. (1) with $c = 1.8$ possesses many periodic solutions; one of these is an oscillation between $-0.5 + \sqrt{1.05}$ and $-0.5 - \sqrt{1.05}$.

The mere presence of aperiodicity does not reveal the rate at which small errors grow. For a quantitative treatment, let the state of the system at time t be given by M numbers $X_1(t), \ldots, X_M(t)$, which may be treated as elements of a matrix X with M rows and one column, or a M-dimensional vector. Let the equation governing the system be

$$\mathrm{d}X/\mathrm{d}t = F(X)\,, \tag{2}$$

and let the elements of F be $F_1, \ldots, F_M$: Let a basic solution and a perturbed solution be given by X and $Y = X + x$, so that x is the error. If x is sufficiently small, it is approximately governed by the homogeneous linear equation

$$\mathrm{d}x/\mathrm{d}t = Gx\,, \tag{3}$$

where the elements G_{ij} of the square matrix G are the partial derivatives $\partial F_i/\partial X_j$. Between a time t_0 and a later time t_1 eq. (3) may be integrated to yield the solution

$$x(t_1) = Ax(t_0)\,, \tag{4}$$

where $A = A(t_1, t_0)$ is a square matrix which depends upon the behavior of X between t_0 and t_1.

If we define the magnitude of the error as the magnitude of the vector x, small errors of a given magnitude at time t_0 satisfy the equation

$$\tilde{x}x = \varepsilon^2 I\,, \tag{5}$$

where I is the identity of order one.

At time t_1 these errors, therefore, satisfy the equation

$$\tilde{x}(A\tilde{A})^{-1}x = \varepsilon^2 I\,. \tag{6}$$

Equations (5) and (6) define a sphere and an ellipsoid in M-dimensional space. Whether or not any small errors grow between t_0 and t_1 depends upon whether any semi-axis of the ellipsoid exceeds the radius ε of the sphere. The semi-axes of the ellipsoid are the quantities $\varepsilon\lambda_i$, where $\lambda_1, \ldots, \lambda_M$ are the singular values of A, *i.e.* $\lambda_1^2, \ldots, \lambda_M^2$ are the eigenvalues of $A\tilde{A}$. These may be numbered in order of decreasing magnitude. Error growth, therefore, depends upon whether the singular value λ_1, or the eigenvalue λ_1^2, exceeds unity.

If t is a later time than t_1, it is evident from (4) that

$$A(t, t_0) = A(t, t_1)A(t_1, t_0)\,. \tag{7}$$

This does not mean that the singular values of $A(t, t_0)$ are products of singular values of $A(t, t_1)$ and $A(t_1, t_0)$. It is even possible that each of the latter two matrices possesses a singular value exceeding unity, while the product matrix does not. To investigate the ultimate growth or decay of small errors, as opposed to temporary growth or decay, we should make $t - t_0$ large. The limiting values

$$l_i = \lim_{t\to\infty} \lambda_i^{1/(t-t_0)} \tag{8}$$

are called the Liapunov numbers of the system, while their logarithms

$$a_i = \lim_{t\to\infty} \log \lambda_i/(t - t_0) \tag{9}$$

are called the characteristic exponents. For many well-behaved systems these are independent of the choice of the state at t_0. For eq. (1), with $c = 1.8$, the single Liapunov number is 1.5; with $c = 1.2$ it is 0.89. (There are other definitions of Liapunov numbers and characteristic exponents, in which the quantities λ_i are eigenvalues of A instead of square roots of eigenvalues of $A\tilde{A}$.)

To evaluate $A(t_1, t_0)$, we may first choose the initial state $x(t_0)$ and then integrate (2) numerically from t_0 to t_1, obtaining $x(t_1)$. We then perturb $x(t_0)$ with M separate one-column matrices $y_1, \ldots, y_M$, where the j-th component y_{ij} of y_i is $\varepsilon\delta_{ij}$ and ε is very small, and integrate numerically M times from t_0 to t_1: We obtain $x(t_1) + z_i$ for each i, from which we may subtract $x(t_1)$. The M columns of $A(t_1, t_0)$ are the M one-column matrices z_i/ε.

5. – Approximate formulae.

Unless a_2 and a_1 are nearly equal, a randomly chosen small error will eventually behave as if the only characteristic exponent were a_1: The magnitude E of the error will then be governed approximately by the equation

$$\mathrm{d}E/\mathrm{d}t = a_1 E \,. \tag{10}$$

A popular measure of the growth of an error is the doubling time, which in this case is given by $\log 2/a_1$.

We have noted that errors do not grow forever. The processes which limit the error growth must be represented by nonlinearities in (2), which have been purposely omitted from (3). A simple assumption, which often gives fairly realistic results, is that these processes are quadratic in E, so that (10) may be replaced by

$$\mathrm{d}E/\mathrm{d}t = -\, c_1 E^2 + a_1 E \,, \tag{11}$$

where c_1 is chosen so that the limiting value of E is a_1/c_1. Equation (11) cannot be rigorously justified even in the frequent cases in which eq. (2) is quadratic in X, but it often affords a useful means for dealing approximately with the life history of an error. The numbers in table III fit (11) rather well, with $a_1 = 1.5$ and $c_1 = 2$.

If the governing equation (2) is not perfectly known, the prediction $X + x$ for X will obey an equation

$$\mathrm{d}(X + x)/\mathrm{d}t = F(X + x) + f(X + x) \,, \tag{12}$$

where f as well as x is small. In this event eq. (3) will be replaced by

$$\mathrm{d}x/\mathrm{d}t = Gx + f(X) \,. \tag{13}$$

The term Gx, depending on x, implies a possible exponential amplification, while the term $f(X)$, which is independent of x, implies a possible additional linear accumulation. Integrating (13) between t_0 and t_1, we obtain

$$z = Ay + B , \tag{14}$$

where A is the same as in (4), and B is a one-column matrix depending only upon the behavior of X between t_0 and t_1. Equation (10), when the approximations leading to it are justified, is then replaced by

$$\mathrm{d}E/\mathrm{d}t = a_1 E + b_1 , \tag{15}$$

where b_1 is a constant. Integration of (15) shows that the eventual exponential growth rate is independent of b_1, *i.e.* independent of the fact that the wrong governing equation is used. This result is consistent with the behavior observed in table II. Likewise, when E becomes large, (11) is replaced by

$$\mathrm{d}E/\mathrm{d}t = -\, c_1 E^2 + a_1 E + b_1 . \tag{16}$$

Equation (16) includes the case in which b_1 is large, whence the solutions of (11) and (16) are not alike. This indicates that the proper exponential growth will not be revealed with a completely erroneous governing equation.

The case in which (11) and (16) fail occurs when the most rapidly amplifying mode grows very rapidly to a small limiting value, after which another mode continues to grow more slowly to a larger limiting value. A better approximation than (11) might then be the pair of equations

$$\mathrm{d}E_i/\mathrm{d}t = -\, c_i E_i^2 + a_i E_i , \tag{17}$$

for $i = 1$ and 2, where $a_1 > a_2$ and $a_1/c_1 < a_2/c_2$ and $E_1 + E_2 = E$. More generally there may be many modes $E_1, E_2, \ldots$ which grow at different rates while small and level off at different amplitudes. This is sometimes the case in geophysical fluid systems which possess a wide variety of scales of motion.

PART II

Error Growth in the Atmosphere.

6. – The general atmospheric problem.

Investigations of error growth acquire special significance when the system being studied is one whose behavior has been the object of numerous attempts at prediction. Probably no system fits this description better than the atmosphere.

There is no question but what the atmosphere is an unstable system. Evidence is its lack of perfect or nearly perfect periodicity [3, 4]. The prominent periods in temperature, wind, moisture and other weather elements are the diurnal and annual periods and their principal overtones (semidiurnal, semiannual, etc.). Other periodic variations such as a lunar tide are detectable with careful measurements. Nevertheless, when all known periods are subtracted out, the remaining variations are still large. These include the weather changes associated with the passage of migratory cyclones and anticyclones across the continents and oceans.

The periodic variations can easily be predicted without a detailed knowledge of the governing laws or the present weather pattern; it takes little skill, for example, to predict that during the next century the winters will continue to be colder than the summers. On the other hand, there is a range beyond which the locations of the migratory storms cannot be predicted with detectable accuracy; if this were not the case, these storms would also occur periodically.

As we have noted, the fact that a system is varying aperiodically does not indicate the rate at which errors will grow. The most direct way to determine this rate would be to disturb the system, *i.e.* to introduce an error at some « initial » time, and observe what happens. For some simple systems this procedure might work, but, if our system is the atmosphere, we would never know what would have happened if we had not disturbed it. We might attempt to predict what would have happened and subtract this prediction from what did happen, but, unless we have somehow introduced a disturbance which is as large as our uncertainty in observing the initial state, the uncertainty in the prediction, and hence in evaluating the error, will continue to be as large as the error itself.

As an alternative to introducing an error, we can search through the many years of past atmospheric states which have been recorded and archived. If we find two states which are analogues, *i.e.* which are very much alike, we may regard the second state as equal to the first plus a reasonably small error. We can then determine the growth of the error by subtracting states following the first state from states following the second. We shall presently examine what we have learned about error growth from studying analogue pairs; for the moment we simply mention that, among all states which are presently archived, there appear to be no pairs which would qualify as good analogues [5].

There remains the possibility of using the equations governing the atmosphere, and this is just what has been done in most investigations of atmospheric predictability. Since we cannot formulate the equations exactly, and since, even if we could do so, we could not solve them exactly, we must introduce some simplifying approximations. Systems of equations which to some degree approximate the true atmospheric equations are generally called « models ». In most instances the solutions have been obtained numerically.

It is to be expected that the least simplified models can yield the most realistic results. However, they also require the greatest computational effort. Since in general the growth rate of an error depends upon both the nature of the error and the state on which it is superposed, the labor involved in performing comprehensive investigations with the best available models is prohibitive. Very simple models had to be used some years ago, when computers were slower, or else the scope of the investigations had to be limited.

7. – A simple model.

The first systematic study of error growth was made with a model in which the state of the atmosphere was represented by only 28 numbers, and its behavior was governed by 28 ordinary differential equations [4, 6]. The study was concerned mainly with the growth of very small errors. The model was derived from the familiar two-level quasi-geostrophic equations, with thermal and mechanical damping and thermal forcing, by expanding the streamfunction at each level in a double Fourier series, and then retaining only 14 terms in each series. Two of these terms represented a zonally symmetric flow with a variable profile, while the remaining twelve represented three interacting waves with variable shapes and longitudinal phases.

The simplicity of the model permitted a special treatment of random errors. We may denote the dependent variables of the model by $X_1, \dots, X_M$, where $M = 28$, and let the governing equation be eq. (2). Small errors then obey (3) and (4), and small errors of a given magnitude at time t_0 satisfy (5). The squared magnitude of an error at time t_1 is then

$$\tilde{x}(t_1)x(t_1) = \tilde{x}(t_0)\tilde{A}Ax(t_0)\,. \tag{18}$$

If an ensemble of errors is random at time t_0, aside from the existence of a common magnitude ε,

$$\langle x_i(t_0)x_j(t_0)\rangle = \varepsilon^2\delta_{ij}/M\,, \tag{19}$$

where the pointed brackets denote an ensemble average. If follows, since the eigenvalues of $\tilde{A}A$ equal those of $A\tilde{A}$, that

$$\langle\tilde{x}(t_1)x(t_1)\rangle = \varepsilon^2\sum_{i=1}^{M}\lambda_1^2/M\,. \tag{20}$$

That is, the amplification of a random error is simply the root-mean-square singular value of A.

We first performed a basic run of 64 simulated days, using three-hour time steps in the integration, and retained the values of x at times $t_0, t_2, \ldots, t_{62}$, the subscript denoting the number of days since t_0. At each time t_j for $j = 0, 2, \ldots, 62$ we then determined the matrix $A(t_{j+2}, t_j)$; this required 28 two-day runs for each value of j. Finally we determined the singular values λ_i of $A(t_{j+2}, t_j)$ and, using (20), found the average two-day amplification factor $\alpha(t_{j+2}, t_j)$ of small errors which were random at time t_j.

We found that α varied greatly with X. On four of the 32 occasions $\alpha(t_{j+2}, t_j)$ was actually less than unity; on three occasions it exceeded 3.0.

We then determined the matrices $A(t_{j+4}, t_j)$ for $j = 0, 4, \ldots, 60$, $A(t_{j+8}, t_j)$ for $j = 0, 8, \ldots, 56$, etc., using (7). Three of the 16 amplifications $\alpha(t_{j+4}, t_j)$ were less than 2.0; two exceeded 10.0. Both 32-day amplifications were greater than 100-fold, while the single amplification $\alpha(t_{64}, t_0)$ was $2 \cdot 10^4$; this implied an average doubling time of 4.5 days. From these results, and from estimates of the accuracy with which real atmospheric states could be determined, we concluded that, pending further results from larger models more closely resembling the real atmosphere, the prospect was rather favorable for forecasting a week in advance and unfavorable for one-month forecasts.

8. – Early experiments with global circulation models.

At the time when the results of the 28-variable study appeared, the meteorological community was in the process of planning the Global Atmospheric Research Program (GARP). This was to be an international program involving many related observational and theoretical investigations. One of its stated aims was to extend the range at which useful predictions could be achieved [7].

As the 4.5-day doubling time indicated by the 28-variable model, and also the general result that an aperiodically varying system could not be predicted at sufficiently long range, became generally known, the question arose as to whether the goal of extended-range prediction might be unattainable. It soon became obvious that error growth studies ought to be performed with the most realistic models possible.

At that time the suitable models in existence were those which had been developed by SMAGORINSKY [8], LEITH [9] and MINTZ [10], who, at a special meeting devoted to the planning of GARP, consented to performing some error growth experiments with their respective models. To the extent possible, the experiments were to have a common format.

Each model had its individuality. Smagorinsky's model, which was the oldest, was a two-layer model covering most of the northern hemisphere. Heating was Newtonian in form, and surface friction and lateral diffusion provided the mechanical damping. The underlying surface was uniform, and the atmosphere was dry. Leith's model had six layers and was global in extent. The underlying surface was again flat, but the model included moisture and an accompanying release of latent heat. Radiation and small-scale convective heating were present, and a somewhat unrealistically large horizontal eddy diffusion coefficient was used to control the computational stability. Mintz's model had only two layers, but was also global. Moisture was not present explicitly, but the underlying surface possessed oceans and continents, and the continents had mountains. The model contained radiative and convective heating and a reasonably small eddy diffusion coefficient; computational stability was controlled by a differencing scheme designed by ARAKAWA [11]. It should be noted that none of these model was constructed for the purpose of investigating predictability.

With each model a basic run was performed and was followed by one or more perturbed runs in which the initial error was confined to the temperature field. In Leith's model, the error decayed rapidly during the first few days and then partially recovered, but leveled off while still small. Once the result was obtained, it became apparent that the large diffusion coefficient which prevented computational instability also suppressed much of the real atmospheric instability and that the migrating disturbances were passing by in essentially periodic sequence.

In Smagorinsky's models the errors grew rather slowly, but after two months acquired a reasonable amplitude before leveling off. The more rapid growth did not commence until the second month. Examination after the fact indicated that during the first month the variations were mainly periodic, subsequently becoming more irregular. During the second month the doubling times was about eight days.

In Mintz's model the error initially decayed, but then grew regularly, with a five-day doubling time. It leveled off with a reasonable amplitude. The visible outcome of these experiments was a report on the feasibility of a global meteorological experiment [7], which concluded that only Mintz's model displayed a reasonable absence of periodicity and that a five-day doubling time represented the best available estimate.

There were some interesting additional results. In the experiments common to all models the initial error field was sinusoidal. SMAGORINSKY and MINTZ also performed experiments with random and localized initial errors and found that the form of the error had little effect on the growth rate. MINTZ also compared the northern and southern hemispheres and found that the errors tended to be smaller in the southern hemisphere (where it was summer), particularly when the initial error was confined to the northern hemisphere, but that the growth rates in the two hemispheres were similar. We have already noted that SMAGORINSKY obtained different growth rates during different months, when the weather patterns were different.

9. – Later studies with global circulation models.

A significant advance in predictability studies with large models came with a new study by SMAGORINSKY [12]. His new model was a nine-level primitive-equation model covering the northern hemisphere. It contained all the advanced physical features of Leith's and Mintz's models; in addition, the underlying surface contained sea ice and land ice and snow, and the absorbers of radiation included water vapor, carbon dioxide and ozone. Actually the model had been in existence when the earlier experiments were performed, but it could not be used then because it would have required too much computation.

In format Smagorinsky's experiment was like the earlier ones. He found that after the first day small errors were doubling in about three days. As they became larger, the growth rate subsided, but the errors had not reached their ultimate size by three weeks, when the computations were terminated. Interesting additional results were that the temperature errors grew most rapidly in the lower troposphere, while, when a spectral analysis was performed, the smaller scales were found to grow most rapidly.

Following these studies, error growth experiments using newly developed or improved global circulation models made frequent appearances [13, 14]. We shall confine our description to one of the most recent ones, which we performed with the operational model of the European Centre for Medium Range Weather Forecats (ECMWF) [15]. This is a 15-level primitive-equation model which contains most of the refinements developed in the decade or more since Smagorinsky's experiment, including a nearly complete hydrological cycle.

As with some of the earlier studies, the amount of computation required for a comprehensive study would have been prohibitive. However, it turned out that nearly all of the computations needed for a more limited study had already been performed in the course of preparing the operational forecasts.

Forecasts from one to ten days in advance are issued daily at ECMWF. The feature which makes it possible to use these forecasts in an error growth study is the rather high quality of the one-day forecasts; thus the one-day forecast for today's weather pattern may be regarded as equal to today's pattern plus a moderately small error. To see how much this error grows in one day, when both patterns are governed by the equations of the model, it is sufficient to compare the two-day forecast for tomorrow with the one-day forecast for tomorrow. Likewise, we can determine the error growth during the next day by comparing the three-day with the two-day forecast for the day after tomorrow, and we may, in fact, continue the process for nine days. Additional estimates of the growth rate of somewhat larger errors may be obtained by comparing K-day with J-day forecasts, for various values of J and K.

As a measure of the error we have used the root-mean-square difference of the 500 mb height fields; this choice effectively gives greater weight to middle and higher latitudes. We have used the analyses and forecasts for each day of the 100-day period beginning 1 December 1980. Table V shows the average error when J-day and K-day forecasts are compared. A zero-day forecast is simply an analysis.

TABLE V. – *Average global root-mean-square* 500 mb *height differences* E_{JK}, *in meters, between J-day and K-day forecasts for the same day, during the period 1 December 1980 - 10 March 1981, made by the* ECMWF *operational model.*

J	E_{J1}	E_{J2}	E_{J3}	E_{J4}	E_{J5}	E_{J6}	E_{J7}	E_{J8}	E_{J9}	E_{J10}
0	24	38	51	63	75	85	93	99	104	108
1		29	45	59	71	82	90	97	102	106
2			36	53	67	78	88	95	100	104
3				44	62	74	85	93	99	103
4					52	70	81	91	97	102
5						61	77	86	95	100
6							68	82	91	97
7								74	87	94
8									79	90
9										84

The top row, comparing forecasts with analyses, reveals the average rate of error growth when two states are governed by two different systems of equations—the true atmospheric equations and the model. It, therefore, indicates the average performance of the model. The diagonal rows, where $K - J$ is constant, reveal the rate of growth when both states are governed by the model equations; this is the rate usually sought in predictability studies. Comparison of the diagonals for different values of $K - J$ indicates that, to a reasonable approximation, the amplification rate $\mathrm{d}E/\mathrm{d}t$ is a function of the error E, so that the numbers in the lower diagonal may be extrapolated for several more days.

The smallest error, 24 m, doubles in about 3.5 days. The usually quoted doubling time is, however, the doubling time for very small errors. We have extrapolated the numbers on the lowest diagonal to smaller errors, using eq. (11). We find that very small errors double in 2.4 days.

Comparing the results of predictability studies performed with hemispheric or global circulation models, we find that the estimates of the doubling time have continually decreased. The first and simplest model, Smagorinsky's original model, suggested about eight days, when it was behaving aperiodically [7]. Mintz's model, with such features as oceans and continents, indicated five days [7]. Smagorinsky's more recent model, with nine levels, reduced the time to three days [12], while the most refined model, the ECMWF model, gave 2.4 days [15].

It seems unlikely that this continual decrease is coincidental. We suspect that the most appropriate doubling time is rather short, perhaps two days. The earlier models were necessarily the simplest, and one result of the simplifications appears to have been an underestimate of the atmosphere's instability. The ECMWF model, whether because of additional physical features, higher spatial resolution, or superior numerical techniques, appears to give the closest estimate.

Global circulation models are still undergoing development, and we have attempted to estimate the doubling time which some future model would yield by noting that the ECMWF model makes some systematic errors. Further refinements in the physics or mathematics will presumably remove these errors, but, in the mean time, we can subtract them out. We have done this and have repeated the study which led to table V. The new doubling time is 2.1 days [15]. This is consistent with our hypothesis that continual refinement will continue to shorten the doubling time.

10. – An analogue study.

To pursue the matter further, we turn to one of the few atmospheric predictability studies which involves no models [5]. It is based on the occurrence of analogues. As we have noted, if we can discover two weather patterns which closely resemble each other, their difference will constitute an error whose growth can be studied. If the patterns occur at the same time of year, they will be governed by effectively the same equations; at different times of year the different diabatic heating fields might cause them to behave rather differently. Ideally the values of each weather element should be nearly alike at all points of the globe.

Our study was based upon twice-daily observations for the five years 1963-1967. At that time there were no data which would reliably indicate whether two patterns were similar in the southern hemisphere, and even in the northern hemisphere we lacked large-scale cloudiness and moisture fields. We were ultimately led to using the height fields at 850, 500 and 200 mb in the northern hemisphere and defining the error as a weighted root-mean-square height difference. Patterns which are sufficiently alike in these fields should also be somewhat alike in temperature, according to the hydrostatic relation, and wind, according to the geostrophic relation. Today we could probably evaluate errors defined in terms of global fields of cloudiness, as measured by satellite.

We compared only those patterns which occurred within one month of the same time of year, but in different years (or different winters, if they occurred in December and January). This yielded a total of about 400 000 error values. These values were normalized so that two randomly chosen weather patterns at the same time of year would yield an error $E = 1$.

We had hoped to find some moderately small errors, say 0.2 or 0.3, but the smallest error encoutered was 0.62. We were thus forced to base our study on mediocre analogues. We first observed that the value of $\mathrm{d}E/\mathrm{d}t$, averaged over all cases in which E fell within a given narrow range, was a smoothly varying function of E, having the same sign as $1-E$. Upon invoking the quadratic hypothesis as expressed by eq. (11), we were able to extrapolate to small values of E, and we found a doubling time of 2.5 days.

It should be noted that this study was performed before SMAGORINSKY had obtained the three-day doubling with his newer model and that the generally accepted value, among those who accepted the reality of error growth, was Mintz's five days [7]. In seeking to explain the discrepancy, we considered the possibility that the analogue results were invalid, but decided that models were probably more likely than observations to go astray. In particular, it appeared that the Arakawa differencing scheme [11] which ensured computational stability could also render the simulated atmosphere more stable [16]. The more recent model studies agree with the analogue study to within 20 percent one way or the other; closer agreement is unlikely in view of the differences in the dates and locations of the data and the quantities chosen to define the error.

11. – The influence of smaller scales.

Error growth studies based on global circulation models tell us only about the growth of errors in the spatial scales which the models resolve. Likewise, the study based on analogues tells us only about the scales appearing in the analyses. By omitting the smaller scales, we effectively make the initial errors in these scales large enough so that they undergo no further growth. Even if the smaller scales were resolved, the error growth in these scales would constitute only a minor part of the total growth, since these scales account for only a small part of the total variability.

The influence of errors in the smaller scales upon errors in the larger scales can nevertheless be large. In most models which retain only the larger scales explicitly, the smaller scales are treated as turbulent eddies, and their effect on the larger scales is represented in terms of coefficients of eddy viscosity and eddy conductivity. In reality the small-scale features which are present at any one time constitute no more than a statistical sample, and their effect is subject to sampling fluctuations. Thus the fact that the details of the smaller scales are uncertain introduces some uncertainty into the behavior of the larger scales. To predict the larger scales perfectly, we would also have to predict the smaller scales perfectly [17].

It appears that errors in very small scales produce an almost undetectably small direct accumulation of errors in the very large scales. Nevertheless, they may cause errors to accumulate fairly rapidly in slightly larger scales. Once these have become appreciable, they will produce an accumulation in still larger scales, etc., and the eventual result will be errors in the largest scales.

Whether errors in the small scales are of practical importance depends on how rapidly the accumulation in the large scales can occur. Suppose that a weather pattern contains an incipient storm and that the initial error consists of overlooking the storm. The error will then grow just as rapidly as the storm itself. If the storm is a migratory cyclone, the error may double in two days or less. If it is a thunderstorm, it may double in less than an hour.

It thus appears that errors in the very small scales require very little time before they have attained the size at which they can appreciably affect somewhat larger scales. These scales will require somewhat longer, but not too long, to affect still larger scales, etc. The initial amplitudes of the smallest-scale errors are, therefore, of little concern, provided that they are not zero. Even an uncertainty of a factor of ten in the magnitude of the error in the thunderstorm scale will cause an uncertainty of only a few hours in the time required for the larger scales to be affected. Equivalently, if we could somehow make perfect measurements of all scales larger than thunderstorms—a task which, incidentally, would be far more expensive than any observational program so far undertaken—, we would subsequently add only a few hours to the range of useful prediction by accomplishing the equally expensive task of improving our thunderstorm observations by a factor of ten.

Let us consider the possible procedures for determining how soon a small error in a small scale will significantly affect the large scales. We can certainly accomplish nothing by physically perturbing the small scales and observing what happens, because again we would have nothing with which to compare the perturbed state. We can construct global models with one-kilometer or even finer horizontal resolution, but to solve them numerically would vastly overburden even today's fastest computers. The studies which have provided tentative answers have taken existing atmospheric models—generally rather simple ones—and derived new systems of equations whose dependent variables represent the amplitudes of the errors, in the various scales [17-19]. The number of equations is minimized by introducing such simplifications as spatial homogeneity and assuming that the spectral amplitude is a smooth function of scale, so that fairly coarse spectral resolution is allowable.

In the first study of this sort [17], our atmospheric model was the simple barotropic-vorticity equation, with neither damping nor forcing. The new variables were the spectral amplitudes in 21 consecutive bands, extending from the 40 m scale to the 40 000 km scale. Growth rates of errors depend very much upon the properties of the basic state on which they are superposed, and the variance spectrum of the basic state was specified in advance. As is generally the case when the original equations are quadratic, the derived equations for means contain covariances, the derived equations for covariances contain mean triple products, etc., and at some point an auxiliary assumption must be introduced to close the system. Our assumption was that quadratic functions of the errors and quadratic functions of the basic state were statistically independent. Subsequent studies with more realistic closure assumptions have yielded rather similar results [18, 19].

The derived equations were formally linear. The nonlinear effects, which, as in eq. (1), should prevent the errors from growing indefinitely, were incorporated by replacing each variable by a constant as soon as that variable, representing the amplitude of the error in one scale, reached the prespecified amplitude of the basic state in that scale.

TABLE VI. – *Times $t_n(10)$ and $t_n(90)$ required for mean square error in spectral band n, with average wave length L_n, to attain 10 percent and 90 percent, respectively, of its limiting value, in theoretical model with no spectral gap, and similar times $t'_n(10)$ and $t'_n(90)$ in theoretical models with strong spectral gap centered near band 10, when initial errors are confined to band 20.*

n	L_n (km)	$t_n(10)$	$t_n(90)$	$t'_n(10)$	$t'_n(90)$
12	12	0.8 h	1.3 h	1.0 h	1.5 h
11	25	1.3	2.1	1.8	2.9
10	50	2.2	3.4	3.6	7.6
9	100	3.5	5.6	7.8	3.5 d
8	200	5.8	9.2	2.3 d	4.6
7	400	9.6	15.2	3.7	5.5
6	800	15.9	1.1 d	4.7	6.3
5	1 600	1.1 d	1.8	5.2	6.7
4	3 200	1.9	3.1	6.0	7.6
3	6 400	3.3	5.4	7.6	9.8
2	12 800	5.9	9.8	10.2	14.2
1	25 600	9.4	16.3	13.8	20.6

In the numerical integrations the initial error was confined to the smallest scales. The middle columns in table VI show the times required for the squares of various amplitudes to attain 10 percent, and 90 percent, of their limiting values. Predictions with a 10 percent error would generally be considered good, while those with a 90 percent error would be almost worthless.

We see that the error progresses up the scale rapidly at first and then more slowly, until, after about half a day, 10 percent errors have reached the scales commonly resolved by global circulation models. The largest scales remain predictable for a week or two.

The final columns in table VI were produced by similar computations, in which the basic state was assumed to possess a « spectral gap » centered in the mesoscale [20]. The errors progress through the smallest scales more or less as before, but they encounter considerable difficulty in crossing the gap and are delayed by several days in reaching the larger scales. Whether or not a spectral gap exists is still a topic for debate. In any event, it is apparent that a definite determination of the range at which the larger scales are predictable will require a more precise knowledge of normal atmospheric conditions than we presently possess.

12. – Concluding remarks.

We have hypothesized that the errors presently made in one-day prediction by the « improved » ECMWF operational model are similar in magnitude and spectral distribution to the errors which would be present in J-day prediction as a result of the inevitable errors in the very small scales, if the largest scales possessed no initial errors at all [20]. Assuming a somewhat weaker spectral gap than the one leading to the final columns in table VI, we have estimated that $J = 4$. It, therefore, appears reasonable to add about three days to estimates, based upon table V, of the range at which predictions of a given quality are possible.

Having seen that there is a limit to atmospheric predictability, it is relevant to ask why there should be a limit, *i.e.* why the atmosphere should be unstable. In the 28-variable atmospheric model the cause is easy to identify; it is advection, which is the only nonlinear process. Advection appears in the model as a displacement of the temperature and vorticity fields by the wind field; since the wind which does the advecting is not uniform, it produces distortion as well as displacement. This increases the variety of temperature and vorticity patterns which can occur and reduces the likelihood of periodic repetition.

In the real atmosphere and also the global circulation models the cause of instability is probably advection also. There are other important nonlinear processes, including evaporation and condensation of water, and radiation. We suspect, however, that the latter processes by themselves would not produce instability, while advection by itself would. Very likely these processes are important modifiers to the instability which advection produces.

* * *

This work has been supported by the GARP Program of the Atmospheric Sciences Section of the National Science Foundation, Grant No. 82-14582 ATM.

REFERENCES

[1] J. GUCKENHEIMER: *Invent. Math.*, **39**, 165 (1977).
[2] M. FEIGENBAUM: *J. Stat. Phys.*, **19**, 25 (1978).
[3] E. LORENZ: *J. Atmos. Sci.*, **20**, 130 (1963).
[4] E. LORENZ: *Trans. N. Y. Acad. Sci.*, II, **25**, 409 (1963).
[5] E. LORENZ: *J. Atmos. Sci.*, **26**, 636 (1969).
[6] E. LORENZ: *Tellus*, **17**, 321 (1965).
[7] J. CHARNEY, R. FLEAGLE, V. LALLY, H. RIEHL and D. WARK: *Bull. Am. Meteorol. Soc.*, **47**, 200 (1966).
[8] J. SMAGORINSKY: *Mon. Weather Rev.*, **91**, 99 (1963).
[9] C. LEITH: *Methods Comput. Phys.*, **4**, 1 (1965).
[10] Y. MINTZ: *W. M. O. Tech. Notes*, **66**, 141 (1964).
[11] A. ARAKAWA: *J. Comput. Phys.*, **1**, 119 (1966).
[12] J. SMAGORINSKY: *Bull. Am. Meteorol. Soc.*, **50**, 286 (1969).
[13] R. JASTROW and M. HALEM: *Bull. Am. Meteorol. Soc.*, **51**, 490 (1970).
[14] D. WILLIAMSON and A. KASAHARA: *J. Atmos. Sci.*, **28**, 1313 (1971).
[15] E. LORENZ: *Tellus*, **34**, 505 (1982).
[16] E. LORENZ: *Bull. Am. Meteorol. Soc.*, **50**, 345 (1969).
[17] E. LORENZ: *Tellus*, **21**, 289 (1969).
[18] C. LEITH: *J. Atmos. Sci.*, **28**, 148 (1971).
[19] C. LEITH and R. KRAICHNAN: *J. Atmos. Sci.*, **29**, 1041 (1972).
[20] E. LORENZ: *Predictability of Fluid Motion*, Am. Inst. Phys., Conf. Proc. (1983).