Tellus (1984), *36A*, 98–110

Irregularity: a fundamental property of the atmosphere*

By EDWARD N. LORENZ, *Center for Meteorology and Physical Oceanography, Massachusetts Institute of Technology, Cambridge, Massachusetts, U.S.A.*

ABSTRACT

Some early ideas concerning the general circulation of the atmosphere are reviewed. A model of the general circulation, consisting of three ordinary differential equations, is introduced. For different intensities of the axially symmetric and asymmetric thermal forcing, the equations may possess one or two stable steady-state solutions, one or two stable periodic solutions, or irregular (aperiodic) solutions. Qualitative reasoning which has been applied to the real atmosphere may sometimes be applied to the model, and checked for soundness by comparing the conclusions with numerical solutions. The implications of irregularity for the atmosphere and for atmospheric science are discussed.

It is certainly as great an honor as I can visualize to have been chosen by the Royal Swedish Academy of Sciences to receive the award which has recently been established through the generosity and farsightedness of Anna-Greta and Holger Crafoord. It is likewise a great privilege to be able to address the Academy, knowing that I am speaking directly to a nation which has produced a wealth of outstanding scientists. Just in my own field, during my lifetime, there have been Professor Tor Bergeron, who identified the process by which the minute water droplets which are suspended as clouds eventually become larger drops which fall as rain, and Professor Carl-Gustaf Rossby, who recognized the field of motion as the fundamental variable in day-to-day weather changes. These meteorological giants were among those whose names I most frequently encountered during my days as a student, and I never imagined then that I would some day become acquainted with both of them.

I do not think that I can properly account for my present attitude toward meteorological research without describing my first formal contact with meteorology. I had originally planned to be a mathematician, and a rather pure one at that, and it was only the entry of the United States into World War II which suddenly found me enrolled in a special training course for weather forecasters.

The course took place at the Massachusetts Institute of Technology, and it differed only slightly from MIT's regular graduate course in meteorology, but, because of the special circumstances of my enrollment, I began with the impression that "meteorology" meant the science of weather forecasting. As new concepts were introduced in the classroom, their relevance to the forecasting problem usually became evident, and my original impression did not change rapidly. Even today, when I regard the atmosphere as something worth studying for its own sake, I look upon weather forecasting as one of the important meteorological problems. I do so with the knowledge that some of my contemporaries do not think of forecasting as a good scientific problem at all.

Not unnaturally, the subdiscipline of meteorology to which I was most strongly attracted was dynamic meteorology, with its many differential equations, even though differential-equation theory was not the branch of mathematics which had formerly appealed to me the most. As the course passed its midpoint and approached its end, it became apparent that the many applications of dynamic meteorology which we were being taught were not going to include weather forecasting. In principle the weather forecasting problem is a standard initial-value problem. With a knowledge

* Crafoord Prize Lecture presented at the Royal Swedish Academy of Sciences, Stockholm, on September 28, 1983.

of the present state of the atmosphere, together with the equations which express the governing physical laws, one should be able to say what the future states will be. But, far from being taught how to use the equations to extrapolate the weather, we were not even told whether it was possible to do so. It was only somewhat later that I realized that at that time nobody was sure that it was possible, while some meteorologists believed that it was not.

This is not to say that the course as a whole neglected forecasting. A fair fraction of our time was devoted to analyzing real weather maps and making forecasts based upon them. But the applications of dynamic meteorology were restricted to such problems as the manner in which the wind typically varies through the lowest kilometer of the atmosphere. The attention paid to these applications had the beneficial effect of making me realize that there was much more to meteorology than the forecasting problem.

As I began to learn meteorology I found it necessary to unlearn some mathematics. I was at first appalled by some of the approximate procedures which were used in manipulating the equations. Sometimes the same quantity would be considered a constant in one term in an equation and a variable in another term, or a variable would be expanded in a power series with no attempt to determine whether the series converged. It took me some time to recognize that rigor, while essential in pure mathematics, can be stifling in some applications. Of course, even the purest of mathematicians use intuition to obtain preliminary estimates, but in meteorology these estimates are often the final products.

The conflict between mathematical and meteorological procedures has been at least partly resolved in the ensuing decades, in a manner which might not have been anticipated. The new element is the widespread use of models.

These models consist of systems of equations. Although the equations may be handled by classical methods, it is often assumed as a matter of course that they are to be solved by digital computers, and the finite differences which often replace the derivatives, and even the computer's own round-off procedure, are sometimes regarded as part of a model.

Models are commonplace in most of the sciences today. Sometimes the equations are based on empirical relationships or even intelligent guesses. Meteorological models differ from these in that their equations are generally simplifications of basic physical laws.

To a considerable extent the adoption of models in meteorology represents only a change in attitude. What was once regarded as an approximation to the system of equations governing the atmosphere is now considered to be the exact system for a model of the atmosphere. The mathematically minded investigator often finds this arrangement more satisfying, since, once he has settled upon a model, he can retain full mathematical rigor in studying its behavior.

Common simplifications are of many sorts. Some replace the atmosphere and its surroundings by another physically plausible system; thus, in a model, the atmosphere may be composed of an ideal gas, with no water in any phase, the earth's surface may consist entirely of land, with no mountains or other topographic features, and the solar input may depend only upon latitude, with no summers and winters or days and nights. Less realistic, but often more beneficial, is the omission of terms which are small compared to other terms in the same equation. Of special interest is the case where the omitted terms include a time derivative, so that the new equation expresses a contemporary relation between dependent variables, and the effective number of variables is reduced. The hydrostatic approximation, expressing a balance between the density and pressure fields, and removing the bothersome vertically traveling sound waves, and the geostrophic approximation, expressing a balance between the wind and pressure fields, and removing inertia-gravity oscillations, are simplifications of this sort.

Finally, for practical reasons we must express an atmospheric state by a finite number of numbers, whose variations will be governed by a similar number of equations. Ordinarily we do this by specifying the atmospheric variables at a grid of points, or expanding the variables in truncated Fourier series. Inevitably this suppresses most of the finer structure.

Atmospheric features of widely differing scales, including globe-encircling currents, migratory storms of continental and sub-continental size, and local showers, are dynamically coupled, and a model will be more realistic if the expected influence of the suppressed scales on the represented scales has somehow been included. Here meteorology is more like some of the other sciences in relying heavily on empiricism and intelligent guesswork.

Even a million numbers will not allow 100-km horizontal resolution over the globe together with two-kilometer vertical resolution through the troposphere. Some of today's operational forecasting models contain more than 100,000 variables, but the computational effort needed to handle them is prohibitive for most pure research. Faced with the necessity of further reductions in resolution, some meteorologists have turned to low-order models, where the number of equations has been reduced below 100 or even below ten, and nothing beyond fair qualitative agreement with reality is to be expected.

Low-order models, and indeed many larger models, are examples of dynamical systems. The theory of dynamical systems, or the qualitative theory of ordinary differential equations, dates back at least to Poincaré (1881), but it has enjoyed a surge of interest in the past decade. A dynamical system may consist of any finite number of equations, but, for obvious reasons, the most thoroughly examined systems have been small ones. Low-order models of the atmosphere and other fluid systems have provided pure mathematicians with many of their specific examples.

Dynamical systems are commonly described in geometrical terms. The dependent variables are treated as coordinates in a multidimensional space. A particular state then becomes a point in the space. As the state varies in accordance with the equations, it traces out a trajectory, or orbit. An exactly periodic solution becomes a closed curve, while a steady-state solution becomes a fixed point, which is treated as a special type of closed orbit. Two distinct orbits cannot intersect, although they may approach one another asymptotically.

Of particular interest are bounded systems, where each orbit eventually enters a fixed region of space and subsequently remains there. For such systems a basic subset of space is the attractor. Given any particular orbit, there are certain points which the orbit will repeatedly approach, arbitrarily closely, at regular or irregular intervals; these are the limit points, or attracting points, for that orbit. A point having a greater-than-zero probability of being an attracting point for a randomly selected orbit lies in the attractor.

An attractor may consist of a stable fixed point, a stable closed curve, or something more complicated. The orbit passing through any point of the attractor lies entirely in the attractor. Orbits not contained in the attractor are likely to approach it asymptotically; exceptions include unstable fixed points and unstable closed orbits. When a solution of a system is determined by numerical integration, it is usually safe to assume that, after an initial transient segment has been discarded, the solution describes a part of the attractor as closely as the round-off error will permit.

When a dynamical system is an atmospheric model, the points on the attractor represent those states which are compatible with the climate. States where, for example, the winds generally blow the wrong way about the high and low pressure centers, or the average wind speed is of hurricane strength, are easily introduced into a model, but they presumably correspond to points which are not on the attractor.

Although all scales of atmospheric motion are interconnected, it is standard practice in investigating the dynamics of a particular type of system, such as a thunderstorm, to disregard the presence of systems of larger or smaller scale, or else prespecify the relevant properties of the latter systems. Such an approach can yield only partial explanations, but attempts to deal with all scales at once often yield nothing at all. I wish to turn to the general circulation of the atmosphere, which is sometimes considered synonymous with the motions of largest scale, and which formed the subject of much of my early research.

Eighteenth and nineteenth century meteorologists commonly regarded the general circulation as being symmetric with respect to the earth's axis; departures from symmetry, when acknowledged, were thought to be independent phenomena. The simplest picture of the general circulation, proposed by Hadley (1735), appeared to be dynamically consistent, i.e., it would have constituted a solution of the not yet formulated dynamical equations, but, after standing for nearly a century, it proved to disagree with newly available observations. A somewhat more complicated picture, arrived at independently by Thomson (1857) and Ferrel (1859), eventually replaced Hadley's picture, but, before the end of the nineteenth century, it met the same fate. Despite its apparent dynamic consistency, it evidently did not represent the way in which the atmosphere chose to circulate.

As the twentieth century progressed it was seriously suggested that the general circulation, as it actually existed, could not be explained independently of the superposed large-scale asymmetric eddies. Especially significant was the work of Defant (1921), who maintained that the eddies transported a considerable amount of heat poleward, and Jeffreys (1926), who proposed that they also transported angular momentum. Observations sufficient to confirm their ideas were lacking, but detailed computations with modern data collections have subsequently shown that they were right; they also show that the eddies transport significant amounts of water. Since in the actual atmosphere there can be no net long-term accumulation or depletion of energy, angular momentum, or water at any latitude, removal of the asymmetries, with no compensating adjustment of the symmetries, would leave an unbalanced circulation, which could not continue to prevail. In other words, the symmetric component of the circulation, i.e., what one obtains by averaging everything about latitude circles, is not by itself a solution of the dynamic equations.

With this new view, general-circulation theory acquired a new problem: why do asymmetries exist at all? It is this problem which I wish to consider in detail. Some theoreticians took the attitude that a circulation without asymmetries, like the one pictured by Hadley, or Thomson and Ferrel, was dynamically impossible. Strictly speaking they were correct, since the asymmetric distribution of oceans and continents, and the topographic features on the continents, would disturb any circulation which might temporarily be symmetric. However, it was claimed that even without the surface asymmetries a symmetric circulation would be impossible. Today, with our tendency to turn to models and computers, we would reject this idea, noting that if we took a model with no external asymmetries, chose symmetric initial conditions, and solved the equations numerically, the solution would remain symmetric forever.

Other investigators claimed that the symmetric component of the circulation was unstable with respect to asymmetric disturbances of small amplitude, and, since there was no shortage of disturbing influences, the ultimate total circulation would possess asymmetric eddies. I find this argument somewhat irrelevant, since the actual symmetric component by itself could not persist even if it were stable.

The explanation which I prefer was clearly stated by V. Bjerknes (1937). He argued that there should be a dynamically consistent symmetric circulation, looking much like the one proposed by Thomson and Ferrel, and differing from the symmetric component of the actual circulation, and that this circulation should be unstable with respect to small asymmetric disturbances. No symmetric circulation could therefore persist. The unstable symmetric circulation which Bjerknes visualized, and, more generally, any symmetric circulation satisfying the dynamic equations, is usually called a Hadley circulation, regardless of whether it looks like the one originally proposed by Hadley.

At this point we might ask why we need to invoke instability to explain the atmosphere's asymmetries, when the earth's surface is asymmetric in any case. Noting that the earth's asymmetries are fixed, we might, instead of asking why atmospheric asymmetries exist, ask why moving asymmetries, such as migratory storms, exist. I have argued (Lorenz, 1969) that there should be a dynamically consistent modified Hadley circulation, differing not too greatly from the unmodified Hadley circulation proposed by Bjerknes, and also steady except for forced periodic seasonal and diurnal variations, and that this circulation, like the unmodified Hadley circulation, should be unstable with respect to inevitable small disturbances. Again the ultimate total circulation must possess large eddies.

How can we verify hypotheses? The atmosphere will not spontaneously acquire the modified Hadley circulation, so we cannot observe what would subsequently happen. Our most promising approach seems to be to turn to models.

Even the most elaborate operational forecasting model can be used to find a Hadley circulation if the earth's surface features and the solar input are first altered so as to be independent of longitude. There should then be a set of initial states whose corresponding time-dependent solutions will converge to the Hadley solution, even if that solution is unstable. This set includes the set of symmetric states. The procedure will fail only if the Hadley circulation is unstable with respect to *symmetric* disturbances, in which case symmetric oscillations will persist forever. Once the solution has been found, its stability is easily tested by perturbing it. To the best of my knowledge the procedure has never been carried out with a large model.

Determining the modified Hadley circulation is another matter. Again there should be a set of appropriate initial states, but they are no longer symmetric, nor do they possess any other readily observed features by which they may be identified. The procedure will therefore fail. There are, of course, numerous algorithms for determining unstable steady states of dynamical systems, but it is not certain how efficient they are when there are 100,000 variables.

Accordingly, I shall turn to a low-order model, which may well be the simplest possible model capable of representing an unmodified or modified Hadley circulation, determining its stability, and, if it is unstable, representing a stationary or migratory disturbance. The model is defined by the three ordinary differential equations

$$\mathrm{d}X/\mathrm{d}t = -Y^2 - Z^2 - aX + aF, \quad (1)$$

$$\mathrm{d}Y/\mathrm{d}t = XY - bXZ - Y + G, \quad (2)$$

$$\mathrm{d}Z/\mathrm{d}t = bXY + XZ - Z. \quad (3)$$

The independent variable t represents time, while X represents the intensity of the symmetric globe-encircling westerly wind current, and also the poleward temperature gradient, which is assumed to be in permanent equilibrium with it. The variables Y and Z represent the cosine and sine phases of a chain of superposed large-scale eddies, which transport heat poleward at a rate proportional to the square of their amplitude, and transport no angular momentum at all. The terms XY and XZ in eqs. (2) and (3) represent amplification of the eddies through interaction with the westerly current; this occurs at the expense of the westerly current, as indicated by the terms $-Y^2$ and $-Z^2$ in eq. (1). The variables have been scaled so that the coefficients are unity. The terms $-bXZ$ and bXY represent displacement of the eddies by the westerly current, and the coefficient b, if greater than unity, allows the displacement to occur more rapidly than the amplification. The linear terms represent mechanical and thermal damping; the damping time for the eddies has been chosen as the time unit, while the coefficient a, if less than unity, allows the westerly current to damp less rapidly than the eddies. The constant terms aF and G represent symmetric and asymmetric thermal forcing; F and G are the values to which X and Y would be driven if the westerly current and the eddies were not coupled. I shall identify the eddies with Rossby waves, even though a prominent mechanism for wave propagation identified by Rossby (1939) is missing from the model. The equations constitute a bounded dynamical system, since it is easily shown that the total energy $(X^2 + Y^2 + Z^2)/2$ will decrease if it exceeds a certain value.

What can such a simple model possibly tell us about the real atmosphere? Certainly it cannot yield much quantitative information. It may serve principally in examining existing hypotheses and formulating new ones. Often we draw conclusions about the general circulation on the basis of qualitative reasoning. Sometimes we can apply similar reasoning to the model, and, in addition, we can solve the equations of the model. If the solution fits the reasoning, it will give us added confidence in our reasoning regarding the real atmosphere. If it reveals a flaw in the reasoning, it will indicate where our reasoning in the atmospheric case need reexamination.

Let us first consider steady-state solutions of eqs. (1)–(3). Equating the time derivatives to zero, we find that

$$Y = (1 - X)G/(1 - 2X + (1 + b^2)X^2), \quad (4)$$

$$Z = bXG/(1 - 2X + (1 + b^2)X^2), \quad (5)$$

while X satisfies the cubic equation

$$a(F - X)(1 - 2X + (1 + b^2)X^2) - G^2 = 0. \quad (6)$$

When $G = 0$, there is a single steady solution, given by $X = F$, $Y = Z = 0$. This represents the Hadley circulation. The solution is clearly stable if $F < 1$ and unstable if $F > 1$, since, if Y and Z are slightly altered, $Y^2 + Z^2$ will decrease in the former case and increase in the latter. The hypothesis proposed by Bjerknes for the atmosphere is therefor applicable to the model, when $F > 1$.

I have argued that a modified Hadley circulation should closely resemble the Hadley circulation which would be found if the external asymmetries could be removed. One kind of external asymmetry is the contrast between the thermal influences of the oceans and continents, represented in the model by

G. To be consistent with my argument, a steady solution with $G \neq 0$ should closely resemble the steady solution with the same values of a, b and F and with $G = 0$.

When $G \neq 0$ the possibility of three steady states arises. We find, in fact, that there is a range of values of G for which eq. (6) possesses three solutions, provided that F exceeds a critical value depending upon b. Unless the three solutions all resemble the Hadley solution, in which case they would also resemble one another, there must be steady solutions which do not qualify as modified Hadley solutions.

Fig. 1, constructed with $a = 0.25$, $b = 4.0$, and $F = 2.0$, is typical. No computer is required; the curve of X as a function of G is readily drawn without even solving the cubic, by choosing values of X and calculating G from eq. (6). The intersection of this curve with the vertical axis, indicated by "H" in Fig. 1, marks the Hadley solution.

For small values of G the single steady solution, typified by "A" in Fig. 1, resembles the Hadley solution very closely; X is close to F, and Y and Z, calculated from eqs. (4) and (5), are close to 0. This solution is also unstable if F is large enough, and small disturbances superposed on it will again amplify and become Rossby waves.

For intermediate values of G there are three steady solutions, typified by "B", "C", and "D". Solution B resembles A and H, but solutions C and D do not, and only B can qualify as a modified Hadley circulation. Solution D might qualify as an atmospheric "blocking" situation. Solution C is always unstable, but D may be stable while B is unstable, in which case small disturbances superposed on B will ultimately approach D. In other instances B and D may both be unstable, and small disturbances superposed on B or D will develop into oscillatory waves, or B and D may both be stable. The latter situation resembles the case of multiple stable equilibria found by Charney and Devore (1979), who used another simple model in which the external asymmetries were orographic rather than thermal.

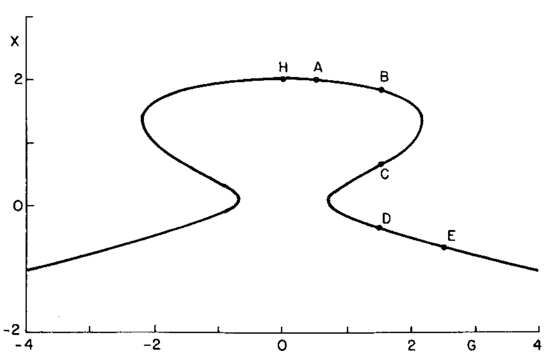

Fig. 1. Variations of X with G, for steady solutions of eqs. (1)–(3) with $a = 0.25$, $b = 4.0$, and $F = 2.0$. H represents Hadley solution: A, B, C, D, E represent solutions discussed in text.

When G is large there is again a single steady solution, typified by "E". It resembles solution D, and cannot logically be called a modified Hadley solution unless D is also. There are therefore values of a, b, F, and G for which a modified Hadley solution cannot be said to exist.

How then does the hypothesis which I have proposed fare, when applied to the model? We cannot invaribly attribute the waves to the instability of the modified Hadley circulation, since, when G is large, there is no such circulation. It would seem that in this instance we must attribute the waves to external asymmetries. Yet we cannot invariably attribute the waves to external asymmetries, since, when $G = 0$, there are no such asymmetries. The remaining possibility is to attribute the waves to a combination of instability and asymmetry, acknowledging that the relative importance of these factors varies from case to case. Perhaps it is reasonable to regard instability as the principal cause, and external asymmetry as only a modifying factor, when G is so small that solution A is the only steady solution.

Returning to the atmosphere, we see that it cannot be taken for granted that the migratory storms owe their existence to the instability of a more regular circulation. Careful quantitative investigation is needed to determine whether or not the atmosphere is the analogue of the model with large F and small G, when the lone steady solution B is unstable. We might add that it is very unlikely that present or past asymmetries in the real atmosphere did develop as a result of the instability of a nearly symmetric circulation; probably such a circulation never occurred in the history of the atmosphere.

Let us see next what the model has to say about the waves which should develop when the steady states are unstable. We shall consider only the case where G is too small for solution D or E to exist.

When there is no asymmetry, eqs. (1)–(3) possess the solution $X = 1$, $Y = \sqrt{a(F-1)} \cos b(t - t_0)$, $Z = \sqrt{a(F-1)} \sin b(t - t_0)$, for some suitably chosen time t_0. This solution proves to be stable with respect to further disturbances, and the circular orbit which represents it forms the attractor. The waves thus behave much more simply than their atmospheric counterparts.

Qualitative reasoning suggests that when $G \neq 0$, the effect of the asymmetric heating will be to intensify the waves when Y and G have the same sign, i.e., when the waves have reached a favorable longitude, and to weaken the waves again when the longitude is unfavorable. The waves will therefore fluctuate in amplitude as they migrate, so that the heat which they transport will fluctuate, whereupon, according to eq. (1), the westerly current will also fluctuate. The waves will move more rapidly when the current is strong, and will show more tendency to amplify.

Numerical solutions indicate that this is often precisely what happens, particularly when the external asymmetry is weak. In our numerical integrations we have used a fourth-order Taylor-series procedure to advance through each time step; our step has generally been 1/30 unit, or four hours, assuming that the time unit, which is the damping time for the waves, is five days. Fig. 2 shows the behavior of X, Y, and Z during a 90-day period, when $a = 0.25$, $b = 4.0$, $F = 8.0$, and $G = 0.2$. Again the behavior is much more regular than anything seen in the atmosphere.

When G is increased to 0.3 or larger, while a, b, and F are kept the same, an entirely different behavior sets in. This is shown in Fig. 3, for $G = 0.8$. The oscillations of the westerly current are more intense and the period has doubled, and during each oscillation a weaker wave followed by a stronger wave passes each longitude. The variations in the wave speed have become more noticeable. The solution is still periodic, with a period of 14.2 days, and the attractor is still a closed curve, but it is no longer a simple circle.

Examination of the numerical solution, before the attractor has been reached, indicates that when the initial behavior is like that in Fig. 2, the even-numbered maxima in X become successively stronger, while the odd-numbered ones become successively weaker and finally disappear altogether. Thus the simpler migratory-wave solution suggested by qualitative reasoning is still dynamically consistent, but it is unstable with respect to another mode of oscillation.

When G becomes still larger another change occurs. In fact, there is an interval of values of G for which either of two types of behavior may prevail. Fig. 4 shows the new type of behavior, again for $G = 0.8$. The maxima and minima of X have become more pronounced and more widely separated, and the successive maxima or minima are unequal. The waves exhibit greater variation in speed, and at times become nearly stationary. There is close repetition after two maxima of X, and exact repetition after four maxima, making the period 80.5 days. The oscillations of X resemble atmospheric "index cycles".

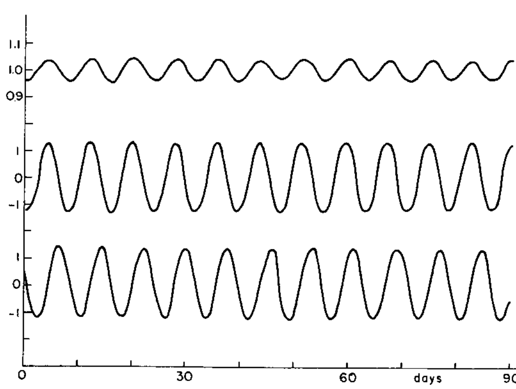

Fig. 2. Variations of X, Y, Z during 90 day interval, in solution of eqs. (1)–(3) with $a = 0.25$, $b = 4.0$, and $F = 8.0$; stable periodic solution with $G = 0.2$. Note expanded scale for X, relative to Y and Z.

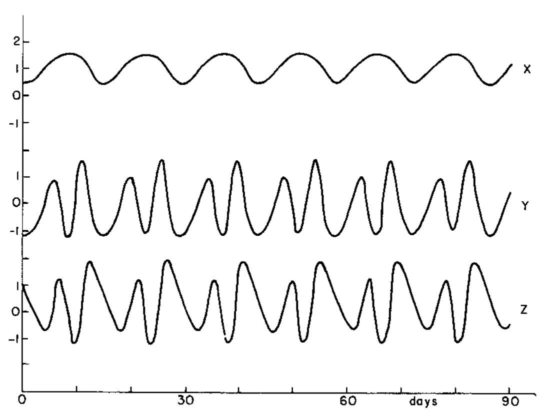

Fig. 3. Same as Fig. 2; stable periodic solution with $G = 0.8$. Normal scale for X.

What we have encountered is a case of intransitivity. There are two distinct sets of long-term statistics, and almost every solution conforms to one or the other. The attractor consists of two disjoint pieces; in this case each one is a closed curve. In meteorological language we would say that there are two climates. We should note that the system where steady solutions B and D are both stable is also intransitive; there the attractor consists of two points.

We cannot tell whether the real atmosphere possesses two possible modes of variation, and is destined forever to conform to just one of these modes, which it acquired by chance. However, it behaves more nearly as if it possessed two or more temporary modes, switching at intervals from one mode to another. It certainly does not repeat its behavior after 14 days, or 80 days, and indeed its behavior during the past summer, for example, does not appear to be a repetition of its behavior during any previous summer. With a further increase in external asymmetry the model also behaves aperiodically, or irregularly. Before examining this case I wish to say something about irregular or chaotic dynamical systems in general.

One might wonder why irregularity should be one of the atmospheric properties worth reproducing with a model. My own interest arose some 25 years ago in connection with statistical weather forecasting, i.e., forecasting by means of empirically established formulas. The idea had circulated among the statistical forecasting community, apparently as a misinterpretation of a paper by Wiener (1956), that suitable linear prediction formulas, where predicted future values of observable quantities are expressed as linear combinations of present and past values, would perform as well as any other attainable formulas. I found the idea implausible, and proposed to test it by choosing a system of nonlinear equations, solving it numerically, and treating the numerical output as observational data, to which standard methods of determining optimum linear prediction formulas could be applied.

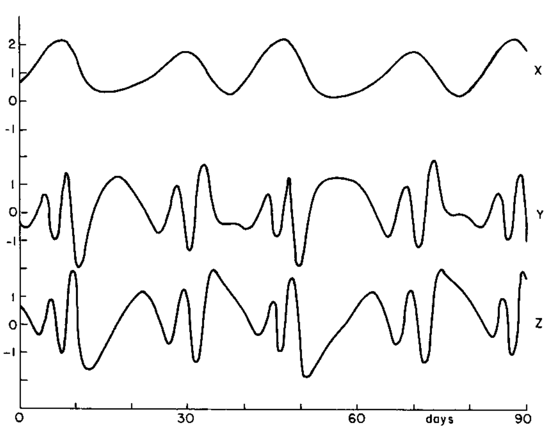

Fig. 4. Same as Fig. 2; second stable periodic solution with $G = 0.8$.

The first task was to choose a system of equations for study. Although a wide variety might have sufficed, it appeared that a model of the general circulation might offer valuable side benefits. It would have to be a low-order model to fit the small computer which resided in my office, and it would have to include sources and sinks of energy, since the statistics of a model which conserved energy would be too highly dependent upon the choice of initial conditions. At that time no model meeting these conditions was in existence.

It soon became apparent that the model would also have to possess an aperiodic general solution, since the linear prediction of a periodic variable by simple extrapolation is a trivial matter. I eventually discovered a 13-variable model whose solutions were unmistakably aperiodic. The three-variable model defined by eqs. (1)–(3) can in fact be derived from it by introducing still more simplifications. The model clearly revealed that linear forecasting had its limitations.

In the course of the work I decided to analyze one solution in greater detail. Every minute the computer would generate one more day of data, and print out the 13 numbers representing the new state. I stopped the computer, typed in the old state, and set the computer running again. Upon returning after about an hour, I found that the solution was quite different from the one which the computer had previously produced. At first I suspected computer trouble, which was fairly common, but the true explanation soon became obvious. The numbers were carried to about six decimal places in the computer, but only three places were printed out. The new initial values were therefore not identical to the old ones, but contained small errors. In the course of about two months these errors had amplified to the point where they drowned the signal. In other words, the time-variable solution was unstable with respect to small disturbances.

This result had far-reaching implications. If the atmosphere and the model should be similar in this respect, long-range weather forecasting would be impossible, since atmospheric observations are certainly not accurate to three decimal places.

Subsequent study revealed that in general aperiodicity implies instability. Provided that the system is bounded, an orbit in the attractor must eventually pass arbitrarily close to a point through which it has previously passed. If the orbit is stable, it will from then on closely repeat its past history, passing close to the point at regular intervals, and it will in a certain sense be periodic, even though it may never pass exactly through the point again. It follows that an aperiodic orbit, with no tendency to repeat at regular intervals, even approximately, must be unstable.

Instability appears to be the most fundamental property of irregular dynamical systems, and may be considered the cause of the irregularity. Absence of periodicity is the property by which an irregular system can most readily be recognized.

A third property which irregular dynamical systems appear to share is the existence of a strange attractor. That is, the attractor does not consist of a simple point, curve, or higher-dimensional manifold, but contains an infinite complex of manifolds. An arbitrary straight line intersects the attractor in a Cantor set, which is an uncountably infinite set whose points are all separated by continua.

While instability is a fundamental concept in meteorology, it is not obvious that strange attractors have important applications. Nevertheless, they are probably the features of chaotic systems which have attracted the most mathematicians to the field, and much of the theory of dynamical systems might not yet have been developed if the attractors had assumed a more commonplace form.

For a concrete example of an iregular system we return to eqs. (1)–(3), with $G = 1.0$, while, as before, $a = 0.25$, $b = 4.0$, and $F = 8.0$. Fig. 5 shows the variations of X, Y, and Z during a 90-day interval. Qualitatively the solution has much in common with the second solution when $G = 0.8$, including some of its regularities. The magnitude of the westerly current oscillates smoothly, with maxima of unequal strength spaced unequally,

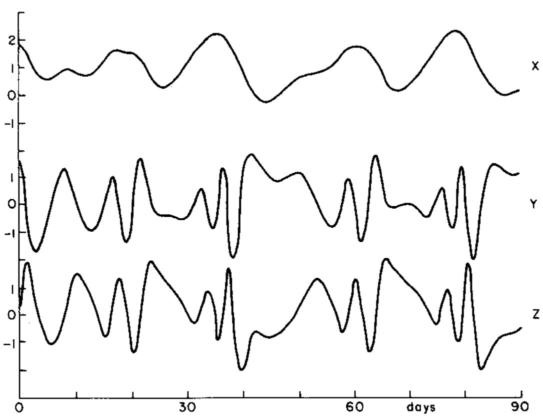

Fig. 5. Same as Fig. 2; aperiodic solution with $G = 1.0$.

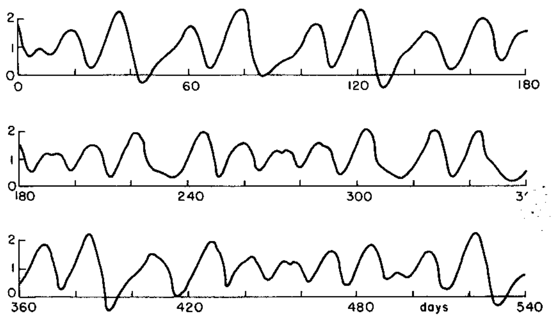

Fig. 6. Variations of X during 540-day interval, for same solution of eqs. (1)–(3) as in Fig. 5.

while the successive waves pass by at intervals as short as four days when the current is strong and longer than ten days when it is weak. The curves of all three variables from day 15 to day 39 look much like those in Fig. 4 from day 26 to day 50. Days 58 to 82 nearly repeat days 15 to 39. In fact, in view of the time needed for the curves in Fig. 4 to repeat themselves, the segment shown in Fig. 5 is too short to assure us of irregularity.

We therefore turn to Fig. 6, which shows three consecutive six-months segments for X alone; the first three months are the ones shown in Fig. 5. The indications for irregularity are considerably stronger. The approximate repetition after six weeks extends through another strong maximum and sub-zero minimum, but then disappears. During the second six months the maxima and especially the minima become weak, and for a while are spaced more closely. It would be easy at this point to mistake the strong minima for transient phenomena. Only during the third half year do they reappear. Indeed, the solution has many of the vicissitudes of the real atmosphere; it switches at irregular intervals between something like Fig. 4 and something like Fig. 3. Perhaps the chief lesson about the atmosphere to be learned from the model is that such fluctuations need not require external irregularity or even external variability.

Further confirmation of irregularity is obtained by directly examining the system's stability instead of its periodicity. If a sufficiently small sphere is placed about some point of the attractor, and if each point of the sphere is then allowed to follow its orbit for a long time, the sphere will be deformed into an approximate ellipsoid. If the system is unstable, at least one axis of the ellipsoid must exceed the diameter of the sphere.

We find that after one year the longest axis has grown by a factor of about 10^7. The second axis, which presumably corresponds to the direction of the orbit, remains essentially unchanged, while the third axis changes by a factor of about 10^{-13}. These factors, known as Liapunov numbers, or their logarithms, the characteristic exponents, tend to determine how complicated the attractor will be (cf. Farmer et al., 1983).

Theory indicates that the attractor should be strange. Even in three dimensions an attractor can be very complicated, and I shall show only the intersection of the attractor with the plane $Y = 0$. Fig. 7 contains 1000 points marking successive crossings of a single orbit with the plane; these occurred during about 10 years. Each crossing on the right, except for those at the lowest two points, is directed from the front of the page ($Y > 0$) to the rear ($Y < 0$); the orbit then proceeds to the left and penetrates the plane at a point on the left, returning afterwards to the right to begin its next circuit. The complete attractor thus resembles a distorted torus.

There is a strong suggestion of a Cantor set of curves; there are large gaps between the finger-like projections, and even within the "fingers" there is a suggestion of parallel curves separated by gaps.

For further confirmation I have enlarged the upper left finger on the right "hand" by a factor of 10. Fig. 8 shows 2000 points; it required a 240-year integration. Evidence for Cantor sets is stronger; a

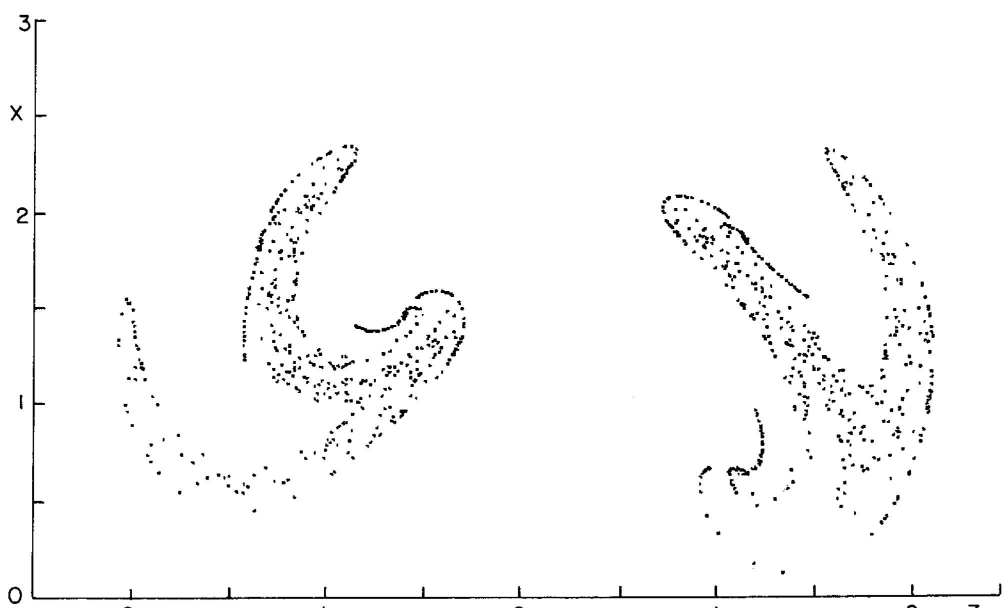

Fig. 7. Intersection of plane $Y = 0$ with attractor of system defined by eqs. (1)–(3), with $a = 0.25$, $b = 4.0$, $F = 8.0$, and $G = 1.0$, as indicated by 1000 intersection points.

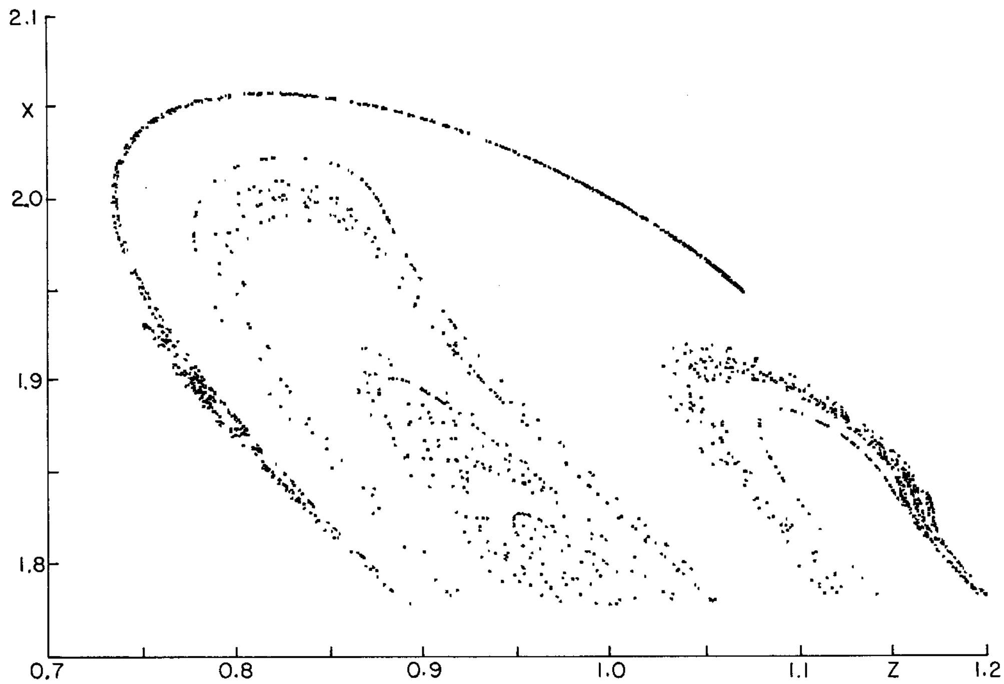

Fig. 8. Portion of intersection shown in Fig. 7, enlarged 10 times, as indicated by 2000 intersection points.

multiplicity of curves, with wide empty spaces between them, is clearly revealed. By contrast, when $G = 0.8$ the intersection of the attractor with the plane $Y = 0$ consists of just 24 points.

There is little question but what the real atmosphere is an irregular system. The principal evidence is its observed absence of periodicity. To be sure, the forced annual and diurnal periods and their overtones account for a fair fraction of the total variability. A weak lunar tidal signal is also detectable, and there has been a deluge of claims for other periodic components. However, if all known or suspected periodic oscillations are subtracted from the total signal, much variability remains. There is no evidence, for example, that migratory storms pass a given location at regularly spaced times.

Can we maintain, however, that irregularity plays an important role in the atmosphere? In some respects it probably does not. It seems rather likely, for example, that a succession of large-scale waves passing by at regular intervals might transport essentially the same amounts of heat, angular momentum, and water across each latitude as are actually transported by the irregularly spaced waves. The general circulation, as sometimes defined, would hardly be altered.

The situation would be different with regard to other climatological statistics. Whenever a particularly hot or cold day arrives, we often wonder whether a record will be broken. In dynamical-systems terminology, we wonder whether a rarely visited portion of the attractor will be encountered. With periodic behavior no records are ever broken; all records are simply tied during each period.

Record temperatures may serve mainly as a topic of conversation, but record rainfalls, with their accompanying record floods, are another matter. Even if we take care to protect our property or life against the greatest flood yet recorded, a still greater flood may momentarily develop.

Deep valleys, carved out by the rivers which now flow through them, are familiar features of the earth's topography. The process of carving never ceases, but much of it occurs when the rivers are heavily overloaded, and carry sand, stones, and rocks as carving tools. If floods of such intensity that they now occur less then once in ten years, for example, never occurred at all, much of the erosion which has shaped our present landscape might not yet have occurred, and the face of the earth would present a different appearance. More than a few cities might never have been founded on their present sites.

However, from the point of view of the atmosphere itself, the most obvious influence of irregularity is its limitation on the extent to which the weather may be predicted. Our present global observing network leaves us with an average error of more than a degree in temperature, a few meters per second in wind, and at least ten per cent in relative humidity. To reduce these errors by a factor of two would require far greater sums than the worlds's nations have seen fit to spend on even more urgent matters. It follows, since we cannot be sure which of many rather similar states is the true present atmospheric state, that we lack a basis for saying which of many widely differing states will occur at some sufficiently distant future time. In other words, accurate long-range prediction of global weather patterns is impossible, even though some features, such as continental averages, may remain reasonably predictable after other features, such as the longitudes of migratory storms, have lost all predictability.

It now behooves us to determine how long is "long range". For a tentative answer we turn again to models. The first systematic studies, carried out some twenty years ago, were interpreted as indicating that small root-mean-square errors in the temperature and wind fields would double in about five days (cf. Charney et al., 1966). Ultimately the growth of errors would subside, since two solutions will never become more different than two randomly chosen states, except temporarily. The immediate conclusion was that good two-week forecasts might eventually become a reality, while the prospect for one-month forecasts seemed dim.

As more and more refined models were developed and then applied to the predictability problem, the estimate of the doubling time continually decreased. The specific reasons for this behavior are not certain, but there are numerous physical features which can cause errors to grow. Apparently some of these were absent from the earlier models, or they entered only in attenuated form.

Very recent computations (Lorenz, 1982), made with the operational model of the European Centre for Medium Range Weather Forecasts, indicate a doubling time of slightly more than two days, for errors in the spatial scales resolved by the model. Moreover, the divergence of the model solutions from reality is close enough to the divergence of the model solutions from one another to permit an estimate of the performance of a perfect model. It appears that even without further improvement in one-day forecasting, use of a perfect model after the first day should produce two-week forecasts as good as today's ten-day forecasts, which show detectable skill, and one-week forecasts as good as today's four-day forecasts, which exhibit some usefulness.

There remains the question as to how much the one-day forecast may be improved. It seems unlikely that a nearly perfect model can produce a nearly perfect forecast. The two-day doubling time applies to horizontal scales of a few hundred kilometers or more. Errors in smaller-scale features, such as thunderstorms, presumably double in hours or minutes rather than days, so that even in the unlikely event that these scales may some day be observed on a world-wide basis with considerable accuracy, they cannot be predicted far ahead. Since the larger and smaller scales are dynamically coupled, errors in the smaller scales, once established, will induce errors in the larger scales, which will then proceed to grow as if they had been present initially. Crude estimates (Lorenz, 1983) suggest that errors comparable to those presently appearing at one day may eventually be postponed to four days. Pending further results, there is additional hope for useful two-week forecasts, while prospects for one-month forecasts are about as dim as before.

Having noted some of the implications of irregularity for the atmosphere, what can we say about the more mundane implications for atmospheric science? The most obvious is that without irregularity there would be no weather forecasting problem, and a far greater portion of meteorological research would have been inspired by pure scientific interest. Scientists capable of creating the large global circulation models would still have been active, but who would have funded their efforts? Who, indeed, would have financed today's large global observing systems, from which we have gained our present knowledge of the general circulation, if the observations had not been needed for forecasts? Progress in meteorology might have been as great as it actually is, but it would presumably have proceeded along different lines. In dynamic meteorology, analytical methods would have been favored over vast computations.

Would this situation have been superior or inferior to what has actually transpired? Eighteenth and nineteenth century meteorology abounded with analytical studies which appeared to be dynamically consistent, but which lacked abundant observations to guide them, and consequently reached incorrect conclusions. I suspect that, in a similar way, current analytical studies may need both observations and computer-generated information, inspired by irregularity, to steer them around similar pitfalls.

Acknowledgement

The development of the model introduced in this talk, and the numerical computations performed with it, were supported by the GARP Program of the Atmospheric Sciences Section, National Science Foundation, under Grant 82-14582 ATM.

REFERENCES

Bjerknes, V. 1937. Application of line integral theorems to the hydrodynamics of terrestrial and cosmic vortices. *Astrophys. Norvegica, 2*, 6, 263–339.

Charney, J. G., Fleagle, R. G., Riehl, H., Lally, V. E. and Wark, D. Q. 1966. The feasibility of a global observation and analysis experiment. *Bull. Amer. Meteorol. Soc. 47*, 200–220.

Charney, J. G. and Devore, J. G. 1979. Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci. 36*, 1205–1216.

Defant, A. 1921. Die Zirkulation der Atmosphäre in den gemässigten Breiten der Erde. *Geogr. Ann. 2*, 209–266.

Farmer, J. D., Ott, E. and Yorke, J. A. 1983. The dimension of chaotic attractors. *Physica 7D*, 153–180.

Ferrel, W. 1859. The motions of fluids and solids relative to the Earth's surface. *Math. Monthly 1*, 140–147, 210–216, 300–307, 366–372, 397–406.

Hadley, G. 1735. Concerning the cause of the general trade winds. *Phil. Trans. 39*, 58–62.

Jeffreys, H. 1926. On the dynamnics of geostrophic winds. *Q. J. Roy. Meteorol. Soc. 52*, 85–104.

Lorenz, E. N. 1969. The nature of the global circulation of the atmosphere: a present view. *The Gobal Circulation of the Atmosphere*, Roy. Meteor. Soc., London, 3–23.

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

Lorenz, E. N. 1983. Estimates of atmospheric predictability at medium range. *Predictability of Fluid Motions*, Amer. Inst. Physics, Conf. Proc.

Poincaré, H. 1881. Mémoire sur les courbes définies par une équation différentielle. *J. de Math. 7*, 375–442.

Rossby, C.-G. 1939. Relation between variations in the intensity of the zonal circulations of the atmosphere and the displacements of the semi-permanent centers of action. *J. Mar. Res. 5*, 2, 38–55.

Thomson, J. 1857. On the grand currents of atmospheric circulation. British Assoc. Meeting, Dublin. See 1892: *Phil. Trans. Roy. Soc. A, 183*, 653–684.

Wiener, N. 1956. Nonlinear prediction and dynamics. *Proc. 3rd Berkeley Sympos. Math. Stat. and Prob. 3*, 247–252.