64

copies
available

The problem of deducing the climate from the governing equations

By EDWARD N. LORENZ, *Massachusetts Institute of Technology*[1]

(Manuscript received January 22, 1964)

ABSTRACT

The climate of a system is identified with the set of long-term statistical properties. Methods of deducing the climate from the equations which govern the system are enumerated. These methods are illustrated by choosing a first-order quadratic difference equation in one variable as a governing equation. The equation contains a single parameter. Particular attention is given to the climatic mean of the single variable.

Analytic methods yield the climate in some cases where the system varies periodically, but generally fail when the system varies nonperiodically. Numerical integration yields a value of the climatic mean for any individual value of the parameter. Additional analytic reasoning is needed to determine the nature of the climatic mean as a function of the parameter.

The progression from steady-state to periodic to nonperiodic behavior, as the parameter increases, is compared to the progression from steady-state to periodic to irregular flow in the rotating-basin experiments, as the rate of rotation increases.

1. Introduction

The continual variations of the state of the earth's atmosphere are presumably governed by a set of physical laws. These laws are frequently expressed as a system of partial differential equations, accompanied by appropriate boundary conditions. It is often assumed that the climate, i.e., the set of long-term statistical properties of the atmosphere, is determined by the same system of equations. A fundamental problem in theoretical climatology is that of deducing the climate from the equations which determine it. This problem may be viewed as a special case of the more general problem of deducing the statistics of solutions of closed systems of equations from the equations themselves.

It should be observed that climate is not universally identified with averages over infinite time intervals. The fact that such expressions as "change of climate" are in common use indicates that many investigators are concerned with averages over long but finite time intervals. In this work, however, we shall be exclusively concerned with averages (or other statistics) over infinite intervals, i.e., with the limits of averages over finite intervals, as the lengths of the intervals approach infinity.

From the beginning there are certain complicating factors which must be recognized. Obviously there are some systems of equations whose solutions become infinite as time increases, and therefore possess no long-term statistical properties. But even among those systems of equations whose solutions remain bounded for all time, there are some systems which have the property that the average of a solution, between two times t_0 and t_1, fails to approach any limit as t_1 becomes infinite while t_0 remains fixed. In other words, there is no *a priori* reason why a climate need exist.

In the special case of the atmosphere, theory alone does not tell us that a climate exists. Recourse to observations is not much more enlightening; the weather of the current century does seem to resemble the weather of the past century, but the weather of the past 12,000 years presumably does not resemble that of the previous 12,000 years, when an ice age flourished.

Next, even if a climate does exist, there is

[1] The research reported in this work has been sponsored in part by the Air Force Cambridge Research Laboratories, under Contract No. AF 19(628)-2409. A portion of this work was performed while the writer was at the National Center for Atmospheric Research, Boulder, Colorado.

no *a priori* reason why this climate should be unique. There are systems of equations which have the property that different solutions, originating from different initial conditions, possess different long-period averages.

In some instances, however, if the initial conditions are subjected to small but otherwise random modifications, while the governing equations are not altered, there is a positive probability that the resulting climate will be unchanged. In such cases the climate will be called *stable*. In other instances, if the initial conditions are similarly modified, there is zero probability that the resulting climate will be unchanged. In these cases the climate will be called *unstable*.

Again, when a stable climate exists, there is no *a priori* reason why this stable climate should be unique. A system possessing a single stable climate (and perhaps many unstable climates) will be called *transitive*. A system possessing more than one stable climate will be called *intransitive*.

In the case of the atmosphere, theory does not tell us whether a stable climate, if one exists, is unique. Observations also are of no avail; the atmosphere is essentially a one-shot experiment, and we cannot introduce new initial conditions and perform the experiment again.

Assuming, however, that we are dealing with a transitive system, let us examine the methods by which we may deduce the climate from a knowledge of the governing equations. One of the most natural approaches consists of deriving from the original set of equations a new set of equations in which the dependent variables are the desired climatological statistics. If the new equations can be solved, it is unnecessary to obtain the superfluous time-dependent solutions of the original equations.

When the original equations are linear, the derived equations are in general linear also. In such instances, this approach has often proved highly fruitful. When the original equations are nonlinear, however, the number of variables in the new equations inevitably exceeds the number of equations, and a finite closed system cannot be obtained. The derived equations may determine important constraints upon the statistics, but usually no complete solution is possible.

At this point the method may be modified by the introduction of additional hypothetical relations connecting the statistics, in order to render the new system closed. These new relations may be primarily statistical; for example, a variable may be assumed to be normally distributed, so that average values of higher powers are expressible in terms of means and variances. On other occasions the new relations may be largely physical; for example, the convective transport of momentum or heat across a given surface may be assumed proportional to the gradient of momentum or heat across the same surface, the factor of proportionality being the coefficient of eddy viscosity or eddy conductivity.

On occasions this procedure may yield gratifying results. However, if the assumptions are not well justified, the results may be entirely unrealistic.

When it is impractical to derive closed systems with statistics as variables, the alternative procedure consists of solving the original equations, and then computing statistics from the solutions. In the most favorable cases, analytic time-dependent solutions may be found, and the climate may be readily evaluated. A familiar example of an analytic solution of a highly simplified system is ROSSBY's (1939 solution of the vorticity equation; Rossby's well-known expression for the speed of a wave is a special statistic of a climate determined by his equation.

Equations whose general solutions oscillate irregularly, however, often possess special steady-state or periodic solutions which are unstable. It is precisely these solutions which are most readily found analytically. Thus there is a very real danger of deducing an unstable climate rather than the desired stable climate. The failure of Rossby's formula in actual day-to-day weather forecasting (where Rossby did not intend to apply it anyway) is easily ascribed to the over-simplification of the equation, but the formula may also be a statistic of an unstable climate. Real weather patterns never contain waves of only one length, and any solutions of the real atmospheric equations exhibiting a single wave length would presumably be unstable.

In the remaining cases, which include those where the stable climates are associated with irregularly oscillating solutions, numerical procedures seem to be indicated. If the equations can be handled numerically at all, the solutions may be treated as data, and the climate may be estimated by processing the data. Nevertheless, some dangers still remain. Although the probability of encountering an unstable climate is very small, if the initial conditions are chosen randomly, the climate is necessarily computed from a finite segment of a time-dependent solution. The possibility that this segment will be unrepresentative of the total solution is just as real as it is when a climate is computed from actual weather data.

In this work we shall illustrate these alternative approaches, using an extremely simple governing equation. From the results we shall draw further conclusions concerning the appropriateness of the different approaches.

2. The governing equation

The usual procedure for solving a system of partial differential equations numerically involves first replacing the system by a system of ordinary differential equations. The new dependent variables, which are functions of time alone, may, for example, be the values of the original dependent variables at a chosen set of points. These equations must in turn be approximated by a system of difference equations. Both these approximations can alter the statistical properties of the solutions.

Wholly apart from these considerations, however, the exact integration of a system of differential equations over a chosen interval of time determines a system of difference relations which is exactly equivalent to the original equations. When the original equations are nonlinear, the equivalent difference equations generally cannot be written in finite form in terms of familiar analytic functions. The existence of the difference equations is assured, however, by the existence of solutions of the differential equations.

We therefore lose no generality, in choosing an arbitrary system of equations to illustrate the problem of deducing the climate, if we choose a system of difference equations instead of differential equations. The alternative methods of attack are still available, and they still possess their distinctive characteristics.

In the interests of economy, we shall seek the simplest possible system of nonlinear difference equations, among those systems capable of generating a stable climate. The simplest system is a system consisting of one equation in one variable, say

$$X_{n+1} = f(X_n) \tag{1}$$

provided that any such equation can govern a climate. We shall require that $f(X)$ be single-valued and continuous in X. If X_0 is an arbitrarily chosen initial value, equation (1) generates the series $\{X\} = \{X_0, X_1, X_2, \ldots\}$, and the long-term statistics of this series, if they exist, constitute a climate determined by equation (1).

We observe that if (1) possesses no steady-state solution $X_0 = X_1 = \ldots$, then either $X_0 < X_1 < \ldots$ or $X_0 > X_1 > \ldots$, in view of the continuity of $f(X)$. Hence $X_n \to \infty$ or $X_n \to -\infty$ as $n \to \infty$, since any finite limit would be a steady-state solution. The series $\{X\}$ then possesses no climate. We shall therefore require that $f(X) = X$ for at least one finite value of X.

The simplest continuous nonlinear function $f(X)$ would appear to be a quadratic function. Upon replacing the dependent variable X by an appropriate linear function of X, we can reduce the most general quadratic equation (1) with at least one steady-state solution to

$$X_{n+1} = aX_n - X_n^2, \tag{2}$$

where $a \geqslant 0$.

If $a > 4$, the choice $X_0 = \frac{1}{2}a$ makes $X_1 > a$ and $X_2 < 0$, after which $X_n \to -\infty$ as $n \to \infty$. If however $0 \leqslant a \leqslant 4$, and if $0 \leqslant X_0 \leqslant a$, then $0 \leqslant X_n \leqslant a$ for all n. We shall therefore choose, as our governing equation, equation (2) with $0 \leqslant a \leqslant 4$, and require in addition that $0 \leqslant X_0 \leqslant a$.

The parabola in Fig. 1 is a plot of X_{n+1} against X_n, constructed for the case $a = 3.75$. The coordinates of some of the points to which we shall later refer are shown.

Equation (2) represents a transformation of the interval $[0, a]$ into a portion of itself. Alternative standard forms, which might be preferable for some purposes, would be

$$X_{n+1} = 4b\,(X_n - X_n^2), \tag{3}$$

transforming the interval $[0, 1]$ into a portion of itself, and

$$X_{n+1} = \tfrac{1}{2} X_n^2 - C. \tag{4}$$

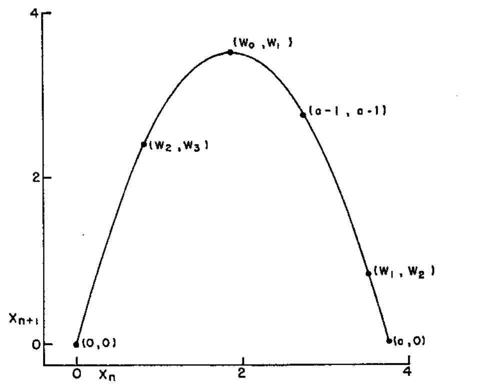

FIG. 1. Graph of the function $X_{n+1} = aX_n - X_n^2$, for the case $a = 3.75$, showing the coordinates of some of the points of interest.

where the slope of the parabola equals the abscissa.

In the following sections we shall illustrate the alternative procedures for deducing the climate, using equation (2) as the governing equation. We shall pay particular attention to the *climatic mean*

$$\bar{X} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} X_n, \tag{5}$$

noting particularly how $\bar{X}$ varies with a.

For some values of a, equation (2) may be shown to be transitive, i.e., to determine a unique stable climate. For all other values of a between 0 and 4 we shall hypothesize that equation (2) is transitive, and speak of *the* value of $\bar{X}$ corresponding to a. We offer no proof of transitivity, but the many numerical solutions of (2) which we have studied do not offer the slightest suggestion of intransitivity.

3. Analytic methods

Consider first the procedure of deriving new equations whose variables are statistics. Averaging both sides of equation (2), we obtain the relation

$$(a-1)\bar{X} - \overline{X^2} = 0. \tag{6}$$

This single equation contains two statistics $\bar{X}$ and $\overline{X^2}$. Any attempt to obtain a closed system by deriving further equations containing $\bar{X}$ or $\overline{X^2}$ inevitably introduces new statistics; for example, if we square both sides of (2) and average, we obtain the relation

$$(a^2-1)\overline{X^2} - 2a\overline{X^3} + \overline{X^4} = 0. \tag{7}$$

The procedure is therefore not entirely satisfactory.

Although not a closed system by itself, equation (6) does place an important constraint upon $\bar{X}$. With the identity

$$\overline{X^2} = \bar{X}^2 + \sigma^2 \tag{8}$$

where σ is the standard deviation of X_n, equation (6) becomes

$$(a-1)\bar{X} - \bar{X}^2 = \sigma^2. \tag{9}$$

Since σ^2 is non-negative, $\bar{X}$ must lie between 0 and $a-1$. If $0 \leqslant a \leqslant 1$, $\bar{X}$ is then non-positive, and since X_n is non-negative for all n, the only remaining possibility is $\bar{X} = 0$. For this range of a, the problem is solved. If $1 < a \leqslant 4$, equation (9) imposes the upper limit $a-1$ for $\bar{X}$, but it does not determine $\bar{X}$.

Consider next the procedure of introducing new hypothetical relations to yield a closed system. We may, for example, assume that, as in a normal distribution, $\mu = 0$ and $\varkappa = 3$, where μ and $\varkappa$ are the skewness and the kurtosis of the distribution of X_n. With the identities

$$\overline{X^3} = \bar{X}^3 + 3\bar{X}\sigma^2 + \mu\sigma^3, \tag{10}$$

$$\overline{X^4} = \bar{X}^4 + 6\bar{X}^2\sigma^2 + 4\mu\bar{X}\sigma^3 + \varkappa\sigma^4, \tag{11}$$

and the identity (8), equations (6) and (7) yield the equation

$$\bar{X}[\bar{X} - (a-1)][2\bar{X}^2 - 2(a+1)\bar{X} + (a^2-1)] = 0, \tag{12}$$

possessing the four roots

$$\bar{X} = 0,\ a-1,\ \tfrac{1}{2}(a+1) \pm \tfrac{1}{2}(a+1)^{\frac{1}{2}}(3-a)^{\frac{1}{2}}. \tag{13}$$

All of these roots are real when $0 \leqslant a \leqslant 3$. As we shall see when we consider the next method, one of these roots correctly characterizes a stable climate, but the present method does not tell us which root this is. When $3 < a \leqslant 4$, only the roots 0 and $a-1$ are real; we shall see that these are both characteristic of unstable climates.

Let us now examine the possibility of solving equation (2) analytically. We shall first look for periodic solutions of period K, for which $X_K = X_0$. Among these are the steady-state solutions, where $K = 1$.

If $\{X_0, X_1, \dots\}$ and $\{Y_0, Y_1, \dots\}$ are any two solutions of (2), and if $Y_n = X_n + \varepsilon_n$, then

$$\varepsilon_{n+1} = \lambda_n \varepsilon_n - \varepsilon_n^2, \tag{14}$$

where $\lambda_n = a - 2X_n$ is the slope of the parabola (2) at the point (X_{n+1}, X_n). If $X_K = X_0$, and if ε_0 is sufficiently small, then approximately

$$\varepsilon_n = \Lambda \varepsilon_0, \tag{15}$$

where

$$\Lambda = \prod_{n=0}^{K-1} \lambda_n \tag{16}$$

is the product of the slopes of the parabola (2) at K points. A periodic solution is therefore stable with respect to small perturbations if $|\Lambda| < 1$, and unstable if $|\Lambda| > 1$. If $|\Lambda| = 1$, further considerations must be invoked to determine the stability.

Considering first the case $K = 1$, we observe that $X_0 = X_1 = \dots = 0$ is always a solution of (2). The statistics of this solution, including the value $\bar{X} = 0$, therefore constitute a climate.

For this solution $\Lambda = a$, so that the climate is stable if $a < 1$ (and also if $a = 1$), but unstable if $a > 1$. Thus the earlier result that $\bar{X} = 0$ if $0 \leqslant a \leqslant 1$ is again obtained.

When $a > 1$, a second steady-state solution $X_0 = X_1 = \dots = a - 1$ exists. For this solution $\Lambda = 2 - a$, so that the climate is stable if $1 < a < 3$ (and also if $a = 3$), but unstable if $a > 3$. Hence $\bar{X} = a - 1$ if $1 < a \leqslant 3$.

When $a > 3$, no stable steady-state solution exists, and we consider the case $K = 2$. Letting $X_2 = X_0$, we obtain the fourth degree equation

$$X_0^4 - 2aX_0^3 + (a^2 + a)X_0^2 - (a^2 - 1)X_0 = 0. \tag{17}$$

Dividing out the steady state solutions 0 and $a - 1$, we find that X_0 and X_1 are the two roots of the equation

$$X_0^2 - (a+1)X_0 + (a+1) = 0. \tag{18}$$

For $0 \leqslant a < 3$ the roots of (18) are complex, but for $a \geqslant 3$, they are real. For this solution

$$\Lambda = -a^2 + 2a + 4, \tag{19}$$

so that $|\Lambda| < 1$ if $3 < a < 1 + \sqrt{6} = 3.449$. For this range of a, the statistics of the periodic solution of period 2 constitute a stable climate, with $\bar{X} = \frac{1}{2}(a+1)$.

Periodic solutions of higher periodicity could be treated similarly. However, X_3 is an eighth-degree polynomial in X_0; the equation $X_3 = X_0$ reduces only to sixth degree when the steady state solutions are divided out. Likewise X_4 is of sixteenth degree in X_0, and the equation $X_4 = X_0$ reduces only to twelfth degree after the solutions of periods 1 and 2 are divided out. Numerical procedures for solving these equations would probably be needed to find $\bar{X}$ as a function of a, and to determine the stability.

We therefore turn to the case $a = 4$, the one remaining case where equation (2) is readily solved analytically. Here, if

$$X_0 = 4 \sin^2 (\pi\theta), \tag{20}$$

we find by repeated application of (2) that

$$X_n = 4 \sin^2 (2^n \pi\theta). \tag{21}$$

We see that only the residual of θ modulo 1 is effective in determining X_n. In particular, if θ is rational, $2^n\theta$ eventually differs from θ by an integer, and the solution is periodic. If θ is not rational, the solution is not periodic. Moreover, for almost all the nonperiodic solutions, the values of $2^n\theta$ modulo 1 have a constant probability density. The corresponding values of X_n are symmetrically distributed about $X_n = 2$, so that $\bar{X} = 2$. The properties of this equation have been discussed by ULAM (1960, ch. 6), and in further detail by STEIN and ULAM (1963, appendix I).

Figure 2 is a graph of X_{n+1} as a function of X_n, for $0 \leqslant a \leqslant 3.449$, and for the single value $a = 4$. It will be left to numerical procedures to complete the graph, for the remaining values of a.

4. Numerical methods

It is a straightforward task to generate numerical solutions of equation (2), whose accuracy will be limited only by the necessity for round-off. Table (1) presents particular solutions, for the cases $a = 3.74$, $a = 3.75$, and $a = 3.76$. In performing these computations, the value of $\frac{1}{4}X_n$, a number between 0 and 1,

TABLE 1. *Numerically determined solutions of the equation* $X_{n+1} = aX_n - X_n^2$, *with initial values* $X_0 = \frac{1}{2}a$, *for the cases* $a = 3.74$, $a = 3.75$, *and* $a = 3.76$.

n	X_n (3.74)	X_n (3.75)	X_n (3.76)
0	1.870	1.875	1.880
1	3.497	3.516	3.534
2	0.850	0.824	0.797
3	2.457	2.411	2.362
4	3.153	3.228	3.302
5	1.852	1.684	1.513
6	3.497	3.479	3.400
7	0.851	0.942	1.225
8	2.459	2.646	3.106
9	3.150	2.922	2.033
10	1.858	2.420	3.511
11	3.497	3.218	0.874
12	0.850	1.711	2.522
13	2.457	3.489	3.122
14	3.152	0.912	1.991
15	1.854	2.588	3.522

was rounded off at each step to 28 bits in the memory of the computer.

In each case the initial value $X_0 = \frac{1}{2}a$ was chosen. It can be shown that if a stable periodic solution $\{Y\} = \{Y_0, Y_1, \ldots\}$ of period K exists, the particular solution $\{W\} = \{W_0, W_1, \ldots\}$ with $W_0 = \frac{1}{2}a$, must approach the solution $\{Y\}$ asymptotically.[1]

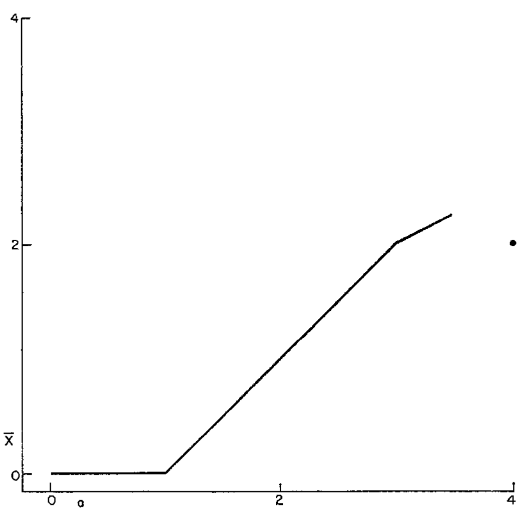

FIG. 2. Graph of $\bar{X}$ as a function of a, for the interval $0 \leqslant a \leqslant 1 + \sqrt{6}$, and for the single point $a = 4$.

For the case $a = 3.74$, the solution is evidently asymptotic to a stable periodic solution of period 5. In general, when a stable periodic solution is discovered, the appropriate value of $\bar{X}$ may be obtained by averaging the values of X_n for a single period.

For the cases $a = 3.75$ and $a = 3.76$, no periodicity is evident. Here, and in general when the solution is nonperiodic, a value of $\bar{X}$ may be estimated by extending the solution to a high value of n, and then averaging all the values of X_n. This value of $\bar{X}$ is indeed an estimate and not a precise value, since it is based upon a finite sample of values of X_n, which may not be representative of a complete solution. This procedure also affords an estimate of $\bar{X}$ when a stable periodic solution does exist, provided that the chosen solution approaches the periodic solution asymptotically, which will be the case if $X_0 = \frac{1}{2}a$.

For any particular value of a, then, the problem of determining the corresponding value of $\bar{X}$ would appear to be solved. Figure 3

[1] To prove this theorem, let U_0 and V_0 be respectively the greatest lower bound of those $X_0 < Y_0$, and the least upper bound of those $X_0 > Y_0$, for which $\{X\}$ does not approach $\{Y\}$ asymptotically. If W_0 does not lie within one of the intervals (U_0, V_0), $\ldots, (U_{K-1}, V_{K-1})$, the mappings $(U_0, V_0) \Leftrightarrow (U_1, V_1) \Leftrightarrow \ldots \Leftrightarrow (U_K, V_K)$ are all one-to-one, and, for $n = 1, \ldots, K$, U_n and V_n, or V_n and U_n, are respectively the greatest lower bound of those $X_n < Y_n$, and the least upper bound of those $X_n > Y_n$, for which $\{X\}$ does not approach $\{Y\}$. But $Y_K = Y_0$, so $U_K = U_0$ and $V_K = V_0$, or $U_K = V_0$ and $V_K = U_0$, and $\{U\}$ and $\{V\}$ are periodic of period K, or $2K$. Moreover $\{U\}$ and $\{V\}$ are unstable solutions, since neighboring solutions approach $\{Y\}$, not $\{U\}$ or $\{V\}$. Thus $\Pi_{n=0}^{k-1} \mu_n \nu_n \geqslant 1$, where $\mu_n = a - 2U_n$ and $\nu_n = a - 2V_n$ are slopes of the parabola (2). Moreover μ_n and ν_n have the same sign, so $\mu_n \nu_n > 0$.

Since $\quad \mu_{n+1} - \nu_{n+1} = \frac{1}{2}(\mu_n + \nu_n)(\mu_n - \nu_n)$,

$$(\mu_K - \nu_K)/(\mu_0 - \nu_0) = \prod_{n=0}^{K-1} (\mu_n + \nu_n)/2 = \pm 1,$$

whence
$$1 = \prod_{n=0}^{K-1} \left(\frac{\mu_n + \nu_n}{2}\right)^2$$

$$= \prod_{n=0}^{K-1} \left[\left(\frac{\mu_n - \nu_n}{2}\right)^2 + \mu_n \nu_n\right] > \prod_{n=0}^{K-1} \mu_n \nu_n \geqslant 1,$$

which is a contradiction. Hence W_0 lies within some interval (U_n, V_n) and $\{W\}$ approaches $\{Y\}$ asymptotically.

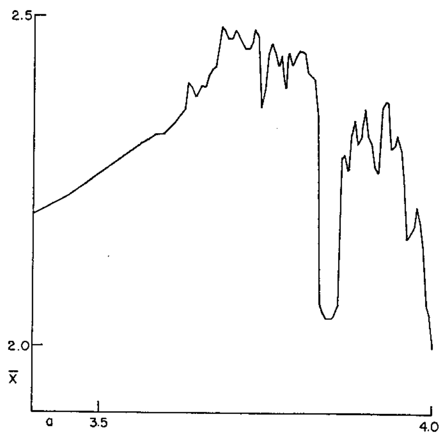

FIG. 3. Graph of $\bar{X}$ as a function of a, as estimated for the interval $3.4 \leq a \leq 4$.

represents a completion of Fig. 2, with estimated values of $\bar{X}$. For values of a from 3.400 to 3.995, at intervals of 0.005, solutions of equation (2) were obtained numerically, in each case with $X_0 = \frac{1}{2}a$. The first 1024 values of X_n were then averaged, yielding the indicated values of $\bar{X}$.

The most striking characteristic of the curve in Fig. 3 is its irregularity. It seems unlikely that the curve can be represented or even closely approximated by any simple combination of familiar analytic functions. Indeed, we are forced to conclude that while the problem of determining $\bar{X}$ for any particular value of a may in essence be solved, the problem of discovering just how $\bar{X}$ varies with a is far from solved.

5. Further considerations

Probably the most unexpected feature of Fig. 3 is the fairly broad band of values of a (from 3.830 to 3.855) in which the estimated values of $\bar{X}$ fall far short of the values of $\bar{X}$ for slightly lower or higher values of a. Although some of the smaller irregularities in Fig. 3 might be attributable to sampling errors, it is unreasonable to expect that sampling would produce a band of this sort.

Inspection of the successive values $X_0, X_1, \ldots$ for $a = 3.83$ reveals the proper explanation for the low values of $\bar{X}$; there is a stable periodic solution of period 3, and $\bar{X}$ is simply the average of the three values. Likewise, the less conspicuous band at $a = 3.74$ corresponds to the stable solution of period 5, which is exhibited in Table 1. If anything requires further explaining, it is not the low values of $\bar{X}$ in the bands, but the higher values outside the bands.

We shall call a value of a for which a stable periodic solution of period K exists a *periodic* value of a, of order K. Values of a which are not periodic of any order will be called *nonperiodic*. Evidently the irregularity of Fig. 3 depends upon the arrangement of periodic and nonperiodic values of a. We shall first consider some properties of periodic values.

For each value of a, let $\{W(a)\} = \{W_0, W_1, \ldots\}$ be the particular solution with $W_0 = \frac{1}{2}a$ (see Fig. 1). Then $W_1 = a^2/4$, and in general W_n is a polynomial of degree 2^n in a.

Suppose that for some value a_0 of a, $W_K = W_0$ for some K. Then, since $\lambda_0 = a_0 - 2W_0 = 0$, $\Lambda = 0$, and the solution $\{W\}$ is stable. Hence a_0 is a periodic value. We shall call any periodic value of a for which $W_K = W_0$ a *central* value of a, of order K. Since W_K is a polynomial, the set of central values of order K is finite, and the set of all central values is denumerable.

Next, if a_0 is a central value of order K, and if a is sufficiently close to a_0, the equation $X_K(a) = X_0(a)$ will have a root $Y_0(a)$ close to $W_0(a_0)$. Moreover, for the periodic solution $\{Y(a)\}$, Λ will be small, and the solution will be stable. Thus, about each central value of a, of order K, there is a continuum of periodic values of a, of order K. We shall call such a continuum a *periodic band*. The K values $Y_0, \ldots, Y_{K-1}$ are roots of an algebraic equation (equation (18) in the case $K = 2$), and are analytic functions of a within the periodic band.

The converse of this result appears to be true also, i.e., any periodic value of a of order K lies within a periodic band, containing a single central value a_0. We offer no rigorous proof for this assertion; it is merely indicated by the study of many individual numerical solutions, which have failed to reveal any exceptions.

Thus, setting $W_1 = W_0$, we find that $a_1 = 2$ is a central value of a of order 1, while the analytic solution $X_n = a - 1$ is stable within the band $1 \leq a \leq 3$ surrounding a_1. Likewise, setting $W_2 = W_0$, we find that $a_2 = 1 + \sqrt{5} = 3.236$ is a central value of order 2. Surrounding a_2 is the periodic band $3 \leq a \leq 3.449$.

We observe that a_1 and a_2 are the highest central values of orders 1 and 2 respectively. If for some $K \geq 2$ there exists a highest central value a_K of order K, and if $2 < a_K < 4$ then $W_{K+1}(a_K) = W_1(a_K) > W_0(a_K)$ (cf. Fig. 1), while $W_{K+1}(4) = 0$. Therefore, by continuity of W_{K+1}, there exists a greatest value a_{K+1}, with $a_K < a_{K+1} < 4$, for which $W_{K+1}(a_{K+1}) = W_0(a_{K+1})$, i.e., a greatest central value of order $K+1$. The existence of a sequence of central values $a_1 < a_2 < a_3 < \ldots < 4$ is thus established by induction.

Since $X_{n+1} > X_n$ if $X_n < W_0$ (if $a > 2$; cf. Fig. 1), it follows that for $K \geq 3$, $W_2(a_K) < \ldots < W_{K-1}(a_K) < W_0$. The increasing sequence a_1, a_2, ... must approach a limit a'; it follows by continuity that $W_n(a') < W_0(a')$ for all values of $n \geq 2$. This is clearly impossible if $W_2(a') > 0$, i.e., if $a' < 4$; hence $a' = 4$.

For large K, $W_2(a_K)$ and, indeed, all but a small number of the values $W_0(a_K), \ldots, W_{K-1}(a_K)$, will be close to zero. Hence $\bar{X}(a_K)$ will be small; in fact, as $K \to \infty$, $\bar{X}(a_K) \to 0$ (far lower than anything suggested by Fig. 3).

The numerical values of a_K and $\bar{X}(a_K)$ are readily estimated. Letting $a_K = 4 - \varepsilon_K$, where ε_K is small, we find that to first-order terms in ε_K, $W_0 = 2 - \frac{1}{2}\varepsilon_K$, $W_1 = 4 - 2\varepsilon_K$, and $W_2 = 4\varepsilon_K$. Working backward from W_K, and neglecting terms in ε_K, we find in view of equation (21) that $W_K = 2 = 4 \sin^2 \pi/4$, $W_{K-1} = 4 \sin^2 \pi/8$ and finally $W_2 = 4 \sin^2 \pi/2^K$. Equating the values of W_2, and replacing the sine of a small argument by the argument itself, we find that, approximately,

$$a_K = 4 - \pi^2/4^K. \qquad (22)$$

Again neglecting terms in ε_K, we find that

$$\bar{X}(a_K) = \frac{4}{K} \prod_{n=1}^{K} \sin^2 \pi/2^n = 6.828/K. \qquad (23)$$

Thus there is a sequence of central values $a_1, a_2, \ldots$, approaching 4 as a limit, for which the corresponding values $\bar{X}(a_1)$, $\bar{X}(a_2)$, ... approach zero as a limit. In the neighborhood of $a = 4$, then, the behavior of $\bar{X}$ as a function of a is far more complicated than Fig. 3 is able to represent.

Values of a_K and $\bar{X}(a_K)$ may also be determined numerically, using a trial and error procedure. Table 2 shows the values so determined. Values of a_K and $\bar{X}(a_K)$ given by the approximations (22) and (23) are presented for comparison. Also shown are the lower and upper bounds a_K^- and a_K^+ of the periodic band surrounding a_K, determined numerically by trial and error. The extreme narrowness of the bands for large values of K is apparent; indeed, the approximations to a_K, although very close, fall outside the true periodic bands.

TABLE 2. *Numerically determined values of central values a_K, lower and upper bounds a_K^- and a_K^+ of the periodic bands surrounding a_K, and mean values $\bar{X}(a_K)$, and approximations a_K' and $\bar{X}'$ to a_K and $\bar{X}$ given by formulas (22) and (23) respectively.*

K	1	2	3	4	5
a_K^-	1.00	3.000	3.8284	3.96010	3.990258
a_K	2.00	3.236	3.8319	3.96027	3.990267
a_K^+	3.00	3.449	3.8401	3.96047	3.990281
a_K'	1.53	3.383	3.8458	3.96145	3.990462
$\bar{X}$	1.000	2.118	2.059	1.661	1.351
$\bar{X}'$	6.828	3.414	2.276	1.707	1.366

Next, since $W_{K-1}(a_{K-1}) > 1$ and $W_{K-1}(a_K) < 1$ if $K \geq 3$, there exists a b_K, with $a_{K-1} < b_K < a_K$, such that $W_{K-1}(b_K) = 1$. Then $W_K = W_{K+1} = \ldots = b_K - 1$. The steady state solution $X_n(b_K) = b_K - 1$ is unstable. Since a solution $\{W\}$ is always asymptotic to a stable periodic solution, when such a solution exists, it follows that there is no stable periodic solution when $a = b_K$, whence b_K is a nonperiodic value. The long-term statistics are therefore the statistics of a nonperiodic solution, just as in the case when $a = 4$.

Now suppose that $a = b_K'$, where $b_K' - b_K$ is very small (small even compared to $4 - b_K$). Then by continuity, $W_K(b_K')$ is close to $b_K' - 1$. Since the slope λ_K of the parabola at $(b_K' - 1, b_K' - 1)$ is about -2, the successive values W_{K+1}, W_{K+2}, ... lie on alternate sides of $b_K' - 1$, each about twice as far from $b_K' - 1$ as its predecessor, until they are no longer close to $b_K' - 1$. By suitably adjusting the value of b_K', then, we may assure ourselves that for any chosen integer $M > 0$, $W_{K+M}(b_K') = W_0$. Let C_{KM} denote the value of b_K' so determined. Then C_{KM} is a periodic value of a, of order $K+M$.

Thus there exists a sequence of values C_{K1}, C_{K2}, ... of a, approaching b_K as a limit, for which $W_{K+1}(C_{KM})$, $W_{K+2}(C_{KM})$, ... are very close to $b_K - 1$, while $W_{K+M}(C_{KM}) = W_0$. It follows that for fixed K, the sequence $\bar{X}(C_{K1})$,

$\bar{X}(C_{K^2}), \ldots$ approaches $b_K - 1$ as a limit as $M \to \infty$, even though $\bar{X}(b_K)$ is not equal to $b_K - 1$.

The inadequacy of Fig. 3 is thus further revealed. Surrounding each point of an infinite sequence $b_1, b_2, \ldots$, there is an infinite collection of periodic bands of values of a, in which the corresponding values of $\bar{X}$ approximate $a - 1$. In particular, the sequence C_{K,K^2} approaches 4 as $K \to \infty$, while $\bar{X}(C_{K,K^2}) \to 3$ (far higher than anything suggested by Fig. 3).

We have thus "explained" the irregularity of Fig. 3; there are numerous bands of periodic values of a for which $\bar{X}$ is very low, including some where $\bar{X}$ is near zero. Separating these bands are other bands where $\bar{X}$ is very high, including some where $\bar{X}$ is near 3. The curve of $\bar{X}$ against a must therefore undergo wild oscillations.

We shall close this section with some speculations concerning the prevalence of nonperiodic values of a. Again, our conclusions will be based partly on hypotheses suggested by the study of many numerical solutions.

In all cases investigated numerically, if a' and a'' are two distinct nonperiodic values of a, the sequences $\{W(a')\}$ and $\{W(a'')\}$ diverge from one another until, for some K, $W_K(a')/a' < 1/2 < W_K(a'')/a''$. It follows by continuity that for some value of a, between a' and a'', $W_K(a) = a/2 = W_0$, i.e., a is a central value. Surrounding a there is then a periodic band, which must lie entirely between a' and a''.

In Table 1, for example, the solutions $\{W(3.75)\}$ and $\{W(3.76)\}$ are fairly close together for $n \leqslant 6$, but have lost all resemblance when $n = 11$. There must exist a periodic value of a, of order 11, somewhere between 3.75 and 3.76.

In short, every pair of nonperiodic values of a is separated by a continuum of periodic values. The periodic values therefore form an everywhere dense set, and moreover the nonperiodic values form a nowhere dense set; i.e., arbitrarily close to *any* value of a there can be found a periodic value which itself lies within a continuum of periodic values. There are no continua of nonperiodic values.

One might suppose, then, that almost all values of a are periodic, in the sense that there is zero probability that a randomly selected value of a is nonperiodic. Such a conclusion is by no means justified; it is easy to construct a nowhere dense set, analogous in some ways to the set of nonperiodic values of a, which may be shown to have positive measure.

Consider the unit interval $0 < \alpha < 1$. The points of the form $\alpha_{mn} = (2m-1)/2^n$ form a denumerable subset which is everywhere dense. About each point α_{mn} construct an interval of length 2^{-2n}. The sum of the lengths of all these intervals is $\frac{1}{2}$, and, since some of the intervals overlap, the measure of the set of all points contained in these intervals is less than $\frac{1}{2}$. The set of points not contained in these intervals therefore has a measure greater than $\frac{1}{2}$, and this set is nowhere dense.

The crucial point is that the widths of the intervals decrease very rapidly as n increases. This feature seems to have its analogy in the periodic bands. Certainly the widths of the bands about the central points a_K decrease very rapidly as K increases, according to the data in Table 2. Although the range $0 \leqslant a \leqslant 3.449$ consists entirely of periodic values, it seems highly likely that in the range $3.9 < a < 4$, for example, a large majority of the values are nonperiodic.

However, the only individual values of a which we have identified as being nonperiodic are those for which $W_{K+M} = W_K$ for some K and some M, while $W_M \neq W_0$. There are only a denumerable number of values of this sort, since for a particular K and M the equation $W_{K+M} = W_K$ has a finite number of roots.

6. Concluding remarks

We have presented several procedures by which the climate, or the long-term statistical properties of a system, might be deduced from the equations governing the system. We have illustrated these procedures by means of a first-order quadratic difference equation in one variable. By specifically choosing the simplest possible nonlinear governing equation, we have abandoned any direct effort to make the system resemble the atmosphere, or any other real physical system. It is noteworthy then that, as far as its solutions are concerned, our equation resembles certain hydrodynamic systems in spite of itself.

Consider, for example, the laboratory experiments of Hide (1958) and Fultz (1959), in which a circularly symmetric vessel containing water is rotated about its (vertical) axis, while being heated near its rim and cooled near its center. If the apparatus rotates slowly enough, the resulting flow is also symmetric about the axis. At slightly higher rates of rotation, particularly when the vessel is annular in shape, waves develop, and travel around the annulus at a uniform rate, without changing their shape. At still higher rotation rates the waves vacillate, i.e., they alter their shape in a regular periodic fashion. At even higher rotation rates, the waves progress and alter their shape irregularly.

Let us agree to identify the parameter a in our simple difference equation (2) with the rate of rotation in the laboratory experiments, and let us identify the variable X_n with the kinetic energy of the waves. We then find that for the slowest rotation ($0 \leqslant a \leqslant 1$), X_n approaches zero, i.e. there are no waves. For somewhat higher rotation rates ($1 < a \leqslant 3$), X_n approaches a positive constant $a-1$, i.e., the waves exist, and their energy remains constant with time. For still higher rotation rates ($3 < a \leqslant 3.449$), X_n oscillates periodically between two values, i.e., the waves vacillate. Finally, for at least some of the highest rotation rates ($3.449 < a \leqslant 4$), X_n oscillates nonperiodically, i.e., the waves move irregularly.

Other analogies could also be drawn; for example, the appearance of successively more complicated forms of convection as a Rayleigh number increases, or the progression from laminar motion to complicated turbulence as a Reynolds number increases.

The writer feels that this resemblance is no mere accident, but that the difference equation captures much of the mathematics, even if not the physics, of the transitions from one regime of flow to another, and, indeed, of the whole phenomenon of instability. In the instances just mentioned, a more complicated type of flow sets in as soon as the simpler flow becomes unstable with respect to perturbations of small amplitude. In other instances, a more complicated flow may set in when a simpler flow goes out of existence altogether. Thus equation (2) possesses a stable solution of period 3 when $3.828 < a < 3.840$. The solution still exists, but is unstable, when $a > 3.840$, but no solution of period 3 exists at all when $a < 3.828$.

Having presented a case for a close mathematical analogy between a simple difference equation in one variable and a complicated system of hydrodynamic equations, let us see what is to be learned from our attempts to deduce the climate governed by the difference equation. Because the real atmosphere varies in a somewhat irregular fashion, we must concede that the difference equation is a better analogue of the atmosphere for the higher values of a (say between 3.449 and 4). Because of the failure of purely analytic methods to yield values of $\bar{X}$ for many of these values of a, and because of the relative ease of determining $\bar{X}$, for any particular value of a, by solving the equation numerically, we might conclude that straightforward numerical integration affords the best method of deducing climate.

However, we have found that for determining the nature of $\bar{X}$ as a function of a, the numerical procedure alone leaves much to be desired. The graph of $\bar{X}$ against a (Fig. 3) reveals the rather broad band near $a = 3.84$ where $\bar{X}$ is relatively small; it does not suggest the existence of considerably narrower bands where $\bar{X}$ is even smaller. It is true that these bands would eventually be detected if $\bar{X}$ were computed for more and more values of a, but the bands are so narrow that it is questionable whether one would continue the search to the point of discovering them. The bands with very high values of $\bar{X}$ are narrower still.

Straightforward analytic reasoning, on the other hand, readily reveals the existence of bands where $\bar{X}$ is very low, and also where $\bar{X}$ is very high. Yet the reasoning necessary to establish the existence of these bands might never have been performed, had not the numerical procedure revealed the first band.

Analytic reasoning also yields an interesting result concerning the probability that a stable periodic solution will exist if a is chosen at random. Here, however, the result is not rigorously proven; it merely follows if certain hypotheses are accepted. These very hypotheses might never have been formed without previous examination of the numerical solutions.

We thus see that a computing machine may play an important role, in addition to simply grinding out numerical answers. The machine cannot prove a theorem, but it can suggest a proposition to be proven. The proposition may then be proven and established as a theorem by analytic means, but the very existence of the theorem might not have been suspected without the aid of the machine.

Ulam (1960, Ch. 8) has discussed the general problem of the computing machine as a heuristic aid to mathematical reasoning, and has presented examples from a number of different branches of mathematics.

As for the problem of deducing the climate, this would appear to be best handled by numerical integration, preferably with the most powerful computing machine available, accompanied by a large amount of careful mathematical reasoning.

Acknowledgement

The writer wishes to express his thanks to Mrs. Ellen Gille for performing a multitude of numerical compuatations, and preparing the diagrams.

REFERENCES

FULTZ, D., LONG, R. R., OWENS, G. V., BOHAN, W., KAYLOR, R., and WEIL, J., 1959, *Studies of thermal convection in a rotating cylinder with some implications for large-scale atmospheric motions.* Meteor. Monographs. Amer. Meteor. Soc., Boston, pp. 104.

HIDE, R., 1958, An experimental study of thermal convection in a rotating liquid. *Phil. Trans. Roy. Soc. London,* (A), 250, pp. 441–478.

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. Marine Res.* V, II, pp. 38–55.

STEIN, P. R., and ULAM, S. M., 1963, *Non-linear Transformation Studies on Electronic Computers.* Los Alamos Sci. Lab., Los Alamos, New Mexico.

ULAM, S. M., 1960, *A Collection of Mathematical Problems.* Interscience Publishers, New York, pp. 150.