1973 AMS Proceedings

3rd Conference on Prob & Statistics in Atmos Sci

PREDICTABILITY AND PERIODICITY:

A REVIEW AND EXTENSION

XEROX

Edward N. Lorenz

Massachusetts Institute of Technology
Cambridge, Massachusetts

## 1. INTRODUCTION AND DEFINITIONS

Among the recent studies which have furthered our understanding of atmospheric predictability, some, such as those based upon numerical integrations of fairly realistic atmospheric models, have been aimed specifically at the atmosphere. Others, including some purely theoretical treatments, have dealt with more general physical or mathematical systems. Regarding the latter it is often loosely stated that periodically varying systems are perfectly predictable, while nonperiodic systems are imperfectly predictable at any range and completely unpredictable at sufficiently long range. The purpose of this review is to examine the basis for such a statement, and to introduce a comprehensive proof.

Before deriving any mathematical results we must have unambiguous definitions of predictability and periodicity. Our definitions will agree with the usual concepts, but in some respects may differ from other definitions, and, to this extent, our conclusions may not be applicable to predictability and periodicity differently defined.

For convenience we shall consider only time series $x(t)$, $y(t)$, etc. defined for equally spaced discrete values of $t$, although many of our remarks will apply equally well to series with continuously varying arguments. Whenever necessary we shall assume that long-term means and other statistics exist.

In the absence of information regarding antecedent conditions, a predictand $y(t_1)$ will possess an <u>a priori</u> (or climatological) probability distribution. Following the specification of antecedent conditions it will also possess an <u>a posteriori</u> distribution. We shall call $y(t_1)$ (completely) <u>unpredictable</u> if and only if these distributions are identical, and (at least) <u>partially predictable</u> if they differ; $y(t_1)$ will be <u>perfectly predictable</u> if and only if the <u>a posteriori</u> probability degenerates to a certainty. The time series $y(t)$ will be considered partially predictable if $y(t_1)$ is partially predictable for at least one value of $t_1$.

Often we wish to use an <u>a posteriori</u> mean as a predicted value, and, in this event, we generally require the predictand to satisfy stricter requirements before being considered predictable. We shall call $y(t_1)$ <u>mean-unpredictable</u> if the <u>a posteriori</u> and <u>a priori</u> probability distributions have equal mean values, and (at least) <u>partially mean-predictable</u> if the mean values differ. Again $y(t)$ will be considered partially mean-predictable if $y(t_1)$ is partially mean-predictable for at least one value of $t_1$.

When a partially predictable series is partially mean-predictable, but not when it is mean-unpredictable, there is a positive <u>reduction of variance</u>; i.e., for the series as a whole (but not for all individual times) the <u>a posteriori</u> variance, which also equals the mean square prediction error, is less than the <u>a priori</u> variance. It is evident that if $y(t)$ is partially predictable, some function of $y(t)$ is partially mean-predictable. A simple example of a series $y(t)$ which is partially predictable but mean-unpredictable from its own past is a series consisting of +1's, 0's, and -1's, where a +1 or -1 is followed with equal probability by a +1, 0, or -1, but a 0 is followed with equal probability by only a +1 or -1; here the time series $y^2(t)$ is partially mean-predictable.

Frequently we wish to predict the occurrence or non-occurrence of some condition or event, e.g., rain at Boston. It is then convenient to introduce, as a predictand $y(t)$, the <u>characteristic function</u> for the event, which assumes the value 1 when the event is occurring and 0 when it is not. In this case the mean value of $y$ is simply the probability of occurrence of the event. Also, since the mean determines the probability distribution, the distinction between predictability and mean-predictability disappears.

Periodicity could be defined in terms of recurrence properties, but we shall define it in terms of the spectrum. The (cumulative) spectral function (the integral of the spectral density function when the latter exists) of a time series is a non-decreasing function. It may therefore be resolved into the sum of three non-decreasing functions with distinct properties, thus resolving the series into the sum of three series. One function is a step function; this corresponds to a line spectrum with a finite or denumerable number of lines, and the corresponding series is by definition periodic. At the other extreme is an absolutely continuous function, whose derivative is the spectral density function. By definition the corresponding series is nonperiodic. In between is the less familiar non-absolutely continuous function, whose increases occur at a nondenumerable but nowhere dense set of points. For our purposes the corresponding series will be considered periodic, although it lacks some features often associated with periodicity. In any particular case one or two of the component series may be absent.

Equivalent definitions may be given in terms of serial covariances. A periodic component is one whose covariance fails to approach zero as the lag approaches infinity; a nonperiodic component is one whose covariance approaches zero. In the latter case it may be shown that the covariance of the series with any other time series also approaches zero at infinite lag. In all cases the serial covariances of the separate components of a series are additive.

Simple examples of series with discontinuous and absolutely continuous spectral functions are respectively a series of 1's and 0's occurring in alternate succession and a series of 1's and 0's forming the successive digits in the binary expression for $\pi$. A series of 1's and 0's with a non-absolutely continuous spectral function may be generated by the relations $y(0) = 0$, $y(2t) = y(t)$, $y(2t + 1) = 1 - y(t)$.

## 2. BACKGROUND THEORY

In this section we enumerate a few established results which are generally well known, but which will be useful as background material for the following sections. Consider first the general problem of linear prediction. Let $y(t)$ denote any time series to be predicted, and let $x_1(t),\ldots,x_M(t)$ denote separate series to serve as predictors. We seek a formula

$$\hat{y}(t + \tau) = \sum_{i=1}^{M} a_i x_i(t) \quad , \qquad (1)$$

where $\tau$ is the range of prediction, $\hat{y}$ is the predicted value of $y$, and the coefficients $a_i$ are to be chosen to minimize the mean square prediction error. These coefficients then satisfy the linear equations

$$\sum_{j=1}^{M} \overline{x_i(t)x_j(t)}\; a_j = \overline{x_i(t)y(t+\tau)} \quad , \qquad (2)$$
$$i = 1,\ldots M \quad ,$$

where a bar denotes an average with respect to $t$. Often we wish to include a constant term in (1). In this case we may let $x_0(t) \equiv 1$, after which equations (1) and (2) still apply if the lower limits of $i$ and $j$ are changed from 1 to 0.

In practical applications the means must be estimated from samples; if $N$ observations of $y$ form the $N \times 1$ matrix $Y$ and the corresponding observations of $x_j$ (including $x_0$) form the $N \times (M + 1)$ matrix $X$, the $(M + 1) \times 1$ matrix $A$ of coefficients $a_j$ is given by

$$A = (X^TX)^{-1}\, X^TY \quad . \qquad (3)$$

A case of special interest occurs when the predictors $x_i(t)$ consist of all previous values of $y(t)$ back to some remote time. Solution (3) is then still valid, but not particularly convenient, and it fails to capitalize on the fact that the elements of $X^TX$ along any diagonal are identical to each other and to one element of $X^TY$, and except for a factor $N$ and an additive constant $(\bar{y})^2$ are in fact serial covariances of $y(t)$. An alternative solution in this case has been given independently by Kolmogorov (1941) and Wiener (1947). For the nonperiodic case the solution involves the spectral density function $F(\omega)$, which is simply the Fourier transform of the covariance function, and is defined for $-\pi < \omega < \pi$. It is found that for predicting one step in advance the ratio of the a posteriori variance to the a priori variance equals the ratio of the geometric mean of $F(\omega)$ to the arithmetic mean. In particular, there is no prediction error if the geometric mean of $F(\omega)$ vanishes. In this event the series is also perfectly predictable at all ranges. An obvious instance occurs when $F(\omega)$ vanishes throughout some continuum within the interval $-\pi < \omega < \pi$. If, however, a nonperiodic series is not perfectly predictable one step ahead, it becomes completely unpredictable by linear formulas as the range approaches infinity. If a series also possesses periodic components, these are perfectly predictable at any range.

Investigations of nonlinear predictability have followed somewhat different paths. A study by the writer (1963a) invokes the theory of dynamical systems. It is assumed that the system under consideration is describable by a finite number of dependent variables, which are governed by a formally deterministic dynamics. A state of the system may then be identified with a point in M-dimensional phase space, and the evolution of the system may be represented by a moving point.

Assuming that some error in observation is inevitable, the true state of a nonperiodic system, and the predicted state originating from the observed state, will diverge from one another, even if the prediction scheme is optimal. Perfect predictability is therefore absent even at short range, and the predictability diminishes as the range increases. These considerations played a part in some of the early planning for the Global Atmospheric Research Program (Charney et al., 1966).

The preceding conclusions leave a basic matter unsettled. The amplification of small errors does not assure us that the predictability at long range decays all the way to zero. The writer (1963b) has dealt with this matter by noting that if small differences between two states amplify but fail to become as large as differences between randomly chosen states, the two states although not remaining close will remain correlated. Since a moving point in phase space must continue to approach some of its previous positions after arbitrarily long time lapses, at least one variable of the system must possess a covariance which does not approach zero at infinite lag, and the system possesses a periodic component.

The results of linear prediction theory and those derived from the theory of dynamical systmes are alike in indicating the general unpredictability of nonperiodic series, but there are important distinctions. The latter results pertain to a lack of predictability whose presence and extnet depend upon the nature of the errors in observation. The former pertain to an intrinsic lack of linear predictability which is present even if the observations are perfect.

It is noteworthy that, in contrast to anything suggested by the dynamical systems approach, linear prediction theory acknowledges perfect predictability of a nonperiodic series, even at infinite range, if the geometric mean of the spectrum $F(\omega)$ is zero. Let us note what happens, however, when random errors are added to the observations. A white-noise spectrum is added to $F(\omega)$, and the geometric mean no longer vanishes. The predictability at infinite range then disappears completely, in agreement with the dynamical systems approach.

It may seem paradoxical that formulas which yield perfect predictions with perfect observations give useless predictions when the observations are almost but not quite perfect. The paradox may be resolved by noting that the formulas for successively longer lags contain successively larger coefficients. Ultimately the effect of multiplying one of these coefficients by even a very small error renders the prediction worthless.

In general, however, the two approaches appear to be complementary rather than equivalent. They will appear more nearly equivalent in the light of further results which we shall presently discuss.

## 3. A RESULT OF WIENER'S

In this section we comment upon further work by Wiener (1956) dealing with prediction of imperfectly observed systems. Wiener treated a mathematical system, but he had the atmosphere in mind as a model. Indeed, in a thoroughly up-to-date assessment of the situation, he states, "Thus the data on which meteorological prediction is to be done represents a very sketchy sampling of the true data which include every local gust of wind and every cool spot or warm spot in every area. Perhaps it may be possible to maintain that these local fluctuations are unimportant in the development of the weather. It is quite conceivable that the general outlines of the weather give us a good, large picture of its course for hours or possibly even for days. However, I am profoundly skeptical of the unimportance of the unobserved part of the weather for longer periods. To assume that these factors ... will not in the long run play their share in determining major features of the weather, is to ignore the very real possibility of the self-amplification of small details in the weather map."

Wiener's principal result is to the effect that if a predictand $y(t)$ is partially predictable by any scheme, it is equally predictable by a linear scheme. This result is important in its own right, but it is the writer's belief that its principal role in the development of meteorology resulted from its being rather generally misinterpreted. The incorrect interpretation is to the effect that if $y(t)$ is partially predictable as a nonlinear function of a set of predictors, such as the dependent variables in the govening dynamic equations, it is equally predictable as a linear function of the <u>same</u> predictors, in the manner of equation (1). Thanks to the simplicity of the solution (3) of (1), this interpretation had the beneficial effect of stimulating considerable work in statistical weather forecasting. Actually the required predictors in the proper linear scheme are characteristic functions derived from the original predictors.

That the required predictors cannot be the original ones is apparent from an example. Wiener's result is purely mathematical. It is easy to construct a mathematical system consisting of a single dependent variable, governed by a deterministic nonlinear equation, which varies nonperiodically and even possesses a white-noise spectrum. One such system is given by the relations $y(0) = 0.8$, $y(t + 1) = 2y^2(t) - 1$. If small enough "observational" errors are added, nearly perfect prediction at short range is possible through the governing equation, but the variable is mean-unpredictable as a linear function of its past.

For the proper interpretation of Wiener's result we specifically acknowledge that our physical system must be imperfectly observed. Only a finite number of dependent variables is observable, and each of these possesses a finite range and is observed with but finite resolution. It follows that there exist but a finite number of different observable states, albeit a very large finite number. Let us denote these by $S_1, \ldots, S_M$.

Now let $x_1(t), \ldots, x_M(t)$ be the characteristic functions for these states, so that $x_i(t) = 1$ whenever the observed state is $S_i$, and $x_i(t) = 0$ at other times. It is these functions which are to serve as the predictors in equation (1).

To solve the corresponding equations (2) we note first that for any t, $x_i(t) = 1$ for exactly one value of i. Hence

$$\sum_{i=1}^{M} x_i(t) \equiv 1 \quad , \tag{4}$$

whereupon a constant term in (1) would be superfluous. Next, we recall that

$$\bar{x}_i = p_i \quad , \tag{5}$$

where $p_i$ is the a priori probability that the state of the system is $S_i$. Since $x_i(t)$ and $x_i^2(t)$ are always equal,

$$\overline{x_i^2} = p_i \quad . \tag{6}$$

Moreover, since two distinct states cannot occur simultaneously

$$\overline{x_i x_j} = 0 \quad \text{if} \quad i \neq j \quad . \tag{7}$$

Finally, given a predictand y(t),

$$\overline{x_i(t)y(t + \tau)} = p_i \bar{y}^i(\tau) \quad , \tag{8}$$

where $\bar{y}^i(\tau)$ is the a posteriori mean of y for those times following occurrences of state $S_i$ by a time lag $\tau$. Introducing (6), (7), and (8) into (2), we find that

$$a_i = \bar{y}^i(\tau) \quad . \tag{9}$$

Now let the state of the system at time t be $S_I$. Substituting $x_I = 1$, and $x_i = 0$ when $i \neq I$, into (1) yields the prediction

$$\hat{y}(t + \tau) = \bar{y}^I(\tau) \quad . \tag{10}$$

This is indeed the best possible prediction. If separate occurrences of $S_I$ are always followed after lag $\tau$ by the same value of y, the prediction is perfect. If different values of y may occur, there is no sure procedure for choosing among them, and (10) minimizes the mean square error. In the event that y(t) is itself a characteristic function for some event, $\bar{y}^I(t)$ is the a posteriori probability of occurrence of the event, and (10) constitutes an optimum probability forecast. This formally linear prediction scheme is in fact seen to be identical to the most nonlinear of schemes, namely, the well known analogue method, with the strictest conditions to be met by states purporting to be analogues.

We have tacitly assumed that all of the potentially useful information for predicting y is known when the observed state $S_i$ is known. It is conceivable that additional predictive information may be available if the present together with several past states are known. If this is the case, we may introduce characteristic functions for distinct sets of successive states instead of distinct states. The subsequent development remains unchanged.

It should be evident that the procedure just described is not operationally practical for weather forecasting, in view of the vast number of distinct states $S_i$. Indeed, experience indicates that we must be very liberal in recognizing states as analogues if we are to discover any analogues at all. The real importance of Wiener's result is that it allows certain established theorems regarding linear prediction to be extended to nonlinear prediction.

## 4. UNPREDICTABILITY OF NONPERIODIC SYSTEMS

We have by now completed most of the work needed to formulate the proper relation between predictability and periodicity. We suppose a physical system is imperfectly observed, and that some variable y(t) representing a property of the system varies nonperiodically. The serial covariance of y(t) then approaches zero at infinite lag. This means that the covariance of y(t) with any other quantity whatever also approaches zero at infinite lag. Hence, if $S_i$ is any state of the system, and $x_i(t)$ is its characteristic function,

$$\lim_{\tau \to \infty} \left( \overline{x_i(t)\ y(t + \tau)} - \bar{x}\ \bar{y} \right) = 0 \quad . \tag{11}$$

It follows from (5) and (8), upon dividing by $p_i$, that

$$\lim_{\tau \to \infty} \bar{y}^i(\tau) = \bar{y} \quad . \tag{12}$$

Thus at infinite range the a posteriori and a priori means are equal, regardless of the state $S_i$ prevailing when the prediction is made, and y(t) is mean-unpredictable.

The possibility remains that y(t) is partially predictable. In this instance, however, some function of y(t), say z(t), which must also represent a property of the system, is partially mean-predictable. Reversing the reasoning of equations (11) and (12) we see that z(t) possesses a periodic component. If, then, an entire physical system varies nonperiodically and is imperfectly observed, it is not perfectly predictable at any range and becomes completely unpredictable as the range becomes infinite.


## ACKNOWLEDGMENT

This work has been supported by the Atmospheric Sciences Section, National Science Foundation, under Grant GA-28203X.


## REFERENCES


Charney, J.G., et al., 1966: The feasibility of a global observation and analysis experiment. Bull. Amer. Meteor. Soc., 47, 200-220.

Kolmogorov, A.N., 1941: Interpolation und Extrapolation von stationären zufälligen Folgen. Bull Acad. Sci. U.S.S.R., Ser. Math. 5, 3-14.

Lorenz, E.N., 1963a: Deterministic nonperiodic flow. J. Atmos. Sci., 20, 130-141.

Lorenz, E.N., 1963b: The predictability of hydrodynamic flow. Trans. N.Y. Acad. Sci., Ser. II, 25, 409-423.

Wiener, N., 1947: Extrapolation, interpolation, and smoothing of stationary time series. Cambridge, Technology Press, and New York, Wiley.

Wiener, N., 1956: Nonlinear prediction and dynamics. Proc. 3rd Berkeley Sympos. Math. Statistics and Prob. Vol. III, 247-252.