Empirical Orthogonal Functions and Statistical Weather Prediction

by

EDWARD N. LORENZ

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
DEPARTMENT OF METEOROLOGY
Cambridge, Massachusetts

DECEMBER 1956

Scientific Report No. 1
STATISTICAL FORECASTING PROJECT

EDWARD N. LORENZ
Director

THE RESEARCH REPORTED IN THIS DOCUMENT HAS BEEN SPONSORED BY THE GEOPHYSICS RESEARCH DIRECTORATE OF THE AIR FORCE CAMBRIDGE RESEARCH CENTER, AIR RESEARCH AND DEVELOPMENT COMMAND, UNDER CONTRACT NO. AF19(604)1566.

ABSTRACT

A "statistical forecasting" formula may be established by determining, from a given sample of data, the linear combination of a set of predictors which forms the best approximation to a given predictand. The dynamical basis for prediction by such formulas is discussed.

Statistical formulas have a greater probability of verifying well, when applied to new data, if the number of predictors is small, relative to the number of independent observations of each predictor.

When the desired predictors consist of a continuous field of some physical quantity, the field may be analyzed into a sum of orthogonal functions of space (Y's), whose coefficients (Q's) are orthogonal functions of time. A small number of Q's with large variances may then be used as predictors. Empirical orthogonal functions (Y's and Q's) may also be determined when the data are heterogeneous. The procedure for determining Y's and Q's is routine, and has been programmed for automatic computation.

The sea-level pressure field over the United States and southern Canada, as represented by observations at 64 stations, has been analyzed into Y's and Q's . Eight Y's and Q's specify 91 per cent of the variance of the pressure field. In predicting the pressure field from the pressure field on the previous day, nothing appears to be gained by using more than a small number of Q's as predictors or predictands.

The possible use of empirical orthogonal functions in nonlinear statistical forecasting, and in dynamic forecasting, is discussed.

TABLE OF CONTENTS

EMPIRICAL ORTHOGONAL FUNCTIONS AND STATISTICAL WEATHER PREDICTION

1. The dynamic basis for statistical weather prediction . . 1

2. Shortcomings of the statistical forecasting method . . 4

3. Representation of a set of predictors 14

4. Representation of weather patterns and situations . . . 19

5. Prediction of the sea-level pressure field 26

6. Further uses of empirical orthogonal functions 30

APPENDIX

1. Expected mean-square errors 33

2. The optimum representation of a set of functions . . . 36

3. Matrix diagonalization by Jacobi's method 39

4. Numerical example 41

ACKNOWLEDGMENT 47

REFERENCES 48

EMPIRICAL ORTHOGONAL FUNCTIONS AND STATISTICAL WEATHER PREDICTION

1. The dynamical basis for statistical weather prediction

Let a set of observed values of a predictand and of one or more predictors be given. One can then usually determine, from these data, the linear combination of the predictors which forms the best approximation to the predictand, in the least-squares sense. During the past few years, meteorologists have become increasingly interested in the problem of predicting the weather by means of such approximations. This method of prediction is popularly called statistical forecasting, notwithstanding the many other statistical methods which have long been used. Among the first systematic studies in statistical forecasting of this sort were those of Wadsworth and Bryan [11; 10]. Subsequently, results obtained by several groups have been presented; these groups have included the Geophysics Research Directorate of the Air Force Cambridge Research Center [12; 13; 14], the Travelers Weather Research Center [7], and the Synoptic Climatology Project at M.I.T. [5; 6; 8; 9].

For the most part these studies have followed by several years the first successful experiments in dynamical forecasting. Studies of the latter sort are now abundant in the literature (see [4]).

To some meteorologists the statistical method has appeared to be the antithesis of the dynamical method, since, although both

methods are objective, the former seems to disregard the dynamics of the atmosphere, while the latter seems to disregard the statistics. It should be mentioned, then, that the dynamic method as practiced is not entirely free of empirical relations, for the dynamic equations are not ordinarily integrated in their primitive form, and the modifications made, such as the geostrophic approximation and the neglect of the vertically integrated divergence, are suggested by the observed behavior of the atmosphere rather than by pure dynamic theory. Likewise, in statistical studies the choice of predictors is often based upon dynamic considerations.

Of more interest at present, however, is the dependence of the success of empirical methods upon dynamics. If atmospheric variations were mere chaotic fluctuations, they could hardly be predicted by empirical formulas. Their predictability results because they are governed by physical laws which presumably do not change with time. These laws tell us that the past and future weather are in some way related.

Recently Wiener [15] has shown that if a statistically stationary system is deterministic, in the sense that its future state is exactly determined from its present by a governing dynamics, and if in addition the entire past history of the system is known, the future of the system may be predicted exactly by linear regression equations, even if the nature of the dynamics is not known. It is of interest to examine the predictability of the weather in the light of this result.

First, the atmosphere by itself is not deterministic, since future influences of the earth upon the atmosphere depend not merely upon the future state of the atmosphere but also upon that of the earth. Even the earth-atmosphere system is not exactly deterministic if the energy received by the sun may undergo significant unpredictable variations. If these variations are in turn governed by dynamic processes within the sun, the sun-earth-atmosphere system may be nearly deterministic. In this case atmospheric observations may be treated as observations of a portion of the sun-earth-atmosphere system.

Secondly, in our problem the infinite past history required for perfect prediction is replaced by a history of only some fifty years for most sea-level observations, and considerably less for most upper-air observations. Moreover these observations cannot be regarded as completely accurate. It is important to know whether, in the absence of infinite and exact past history, good (but not perfect) linear prediction is better aided by considerable knowledge of the far past, or by intensive knowledge of the near past. In the former case Wiener's result would be of doubtful practical value to the forecaster; in the latter case it would imply that useful empirical relations are to be found.

Regardless of whether far-past or near-past history is more desirable, the dynamic equations suggest that near-past history alone may be of considerable use. Although we cannot integrate the general dynamic equations in finite form, we know that the integral of these equations expresses the future state of the atmosphere as a compli-

cated nonlinear function of the present state and of the intervening external influences. Some of the simplified dynamic equations express the pressure and temperature fields in the very near future as fairly simple quadratic functions of the present state. It seems reasonable that these same nonlinear functions are at least partly reflected in the behavior of the atmosphere in the very near past. In addition, some of the information which is lost by observing only certain features or certain portions of the atmosphere, e.g., observing sea-level pressure only, may be reflected in the immediate past behavior of the observed features within the observed portions. Standard methods of linear prediction assure us that this information, to the extent that it is contained implicitly in the near-past behavior of the atmosphere, may be used to good advantage.

It would therefore appear that the more predictable the atmosphere may be by dynamical methods, the more predictable it is likely to be by purely empirical methods. The dynamic equations themselves justify attempts to predict by empirical methods, even when the data are restricted in extent and in kind.

2. Shortcomings of the statistical forecasting method

It is a straightforward process to determine the "best" representation of a predictand as a linear combination of a set of predictors, by the method of least squares. Let the predictand be $x(t)$, and let the M predictors be $p_1(t),---,p_M(t)$, where t

may be time, or any other parameter distinguishing one observation from another. For any choice of $M + 1$ constants $c_0, ---, c_M$, we may then let

$$x(t) = c_0 + \sum_{m=1}^{M} c_m p_m(t) + r(t) \quad . \tag{1}$$

Equation (1) is a prediction formula for $x(t)$; the first two terms on the right form the predicted value of $x(t)$, while the final term $r(t)$, which depends upon the choice of constants, is the residual or error in predicting $x(t)$. The problem at hand is that of determining the set of $M + 1$ constants c_m which minimizes the mean value of r^2.

To accomplish our task we choose a sample consisting of N observations of each quantity. We then minimize $\overline{r^2}$. Here and subsequently, a bar ($\overline{\ }$) denotes the mean value of a quantity over the N observations comprising the sample.

By letting $p_0(t) = 1$, we may write (1) in the more concise form

$$x(t) = \sum_{m=0}^{M} c_m p_m(t) + r(t) \quad , \tag{2}$$

so that

$$\overline{r^2} = \overline{x^2} - 2\sum_{m=0}^{M} c_m \overline{p_m x} + \sum_{m,n=0}^{M} c_m c_n \overline{p_m p_n} \quad . \tag{3}$$

In order that $\overline{r^2}$ be minimized, the derivative $\partial \overline{r^2} / \partial c_m$ must vanish for all values of m, so that

$$\sum_{n=0}^{M} \overline{p_m p_n}\, c_n = \overline{p_m x} \qquad \text{for } m = 0,---,M \ . \tag{4}$$

Equations (4) are a set of $M + 1$ equations in the $M + 1$ unknowns c_m. They may therefore be solved by numerous standard methods; among these, the method of Crout [1] is particularly suitable for both manual and automatic computation.

From (4) and (2) follow the important relations

$$\overline{p_m r} = 0 \qquad \text{for } m = 0,---,M \ . \tag{5}$$

Conversely, from (5) and (2), equations (4) follow. Hence, either equations (5) alone or equations (4) alone form a necessary and sufficient condition that $\overline{r^2}$ be minimized, and that (1) or (2) be the "best" prediction formula for $x(t)$.

It is often more convenient to use a prediction formula which refers to departures of x and p_m from their mean values. From (5) it follows that $\bar{r} = 0$. Applying a bar to each term in (1) and subtracting the result from (1), we find that

$$x^*(t) = \sum_{m=1}^{M} c_m p_m^*(t) + r(t) \ , \tag{6}$$

where a star (*) denotes the departure of a quantity from its arithmetic mean value over the sample.

Because of the formal analogy between (6) and (2), relations analogous to (4) and (5) also hold. Thus the sets of equations

$$\sum_{n=1}^{M} \overline{p_m^* p_n^*}\, c_n = \overline{p_m^* x^*} \qquad \text{for } m = 1,---,M \ , \tag{7}$$

and

$$\overline{p_m^* r} = 0 \qquad \text{for } m = 1,---,M \ , \tag{8}$$

are each necessary and sufficient conditions that $\overline{r^2}$ be minimized. In practice the M equations (7) are often used instead of the $M + 1$ equations (4) to determine the constants c_m.

From (6) and (8) it follows that

$$\overline{r^2} = \overline{x^{*2}} - \overline{\left(\sum_{m=1}^{M} c_m p_m^* \right)^2} \tag{9}$$

The first term on the right is simply the variance of the predictand, while the term on the left is the variance of the error, or the unexplained variance. The final term, in addition to being the variance of the predicted value, is therefore the amount of the variance of x "explained" by the predictors, and the ratio of this term to $\overline{x^{*2}}$ is the reduction of variance, a fraction often used as a measure of the goodness of the prediction.

The quantities $\overline{p_m^* p_n^*}$ and $\overline{p_m^* x^*}$ in (7) are covariances with respect to time; moreover, they are sample covariances rather than population covariances. It is a familiar observation that covariances tend to differ considerably from one sample to another, and hence from sample to population. It follows that the coefficients c_m also depend upon the particular sample. Hence the best prediction formula, as

determined by one sample, is in general not the best formula for the whole population.

Indeed, it appears that when meteorological quantities are involved, there is no such thing as a best prediction formula. The best formula for the entire population might be defined as the limit of the best formula for a sample, as the size of the sample approaches infinity. Such a limit, however, does not necessarily exist, and even if it does, it cannot be found because of lack of data. On the other hand, what we usually wish is the best formula not for the entire population, but for some unspecified but not too extensive portion of the future. Certainly we do not wish a formula which will yield good day-to-day predictions during a future ice age, if the formula, in order to do so, must sacrifice some of its effectiveness during the next few decades. If two formulas give nearly equal reductions of variance within a sample consisting of recent past data, we cannot say, with a very high probability of being right, which of these will give the better prediction during the near future.

This state of affairs should not discourage us, because even if we cannot identify a best prediction formula, we can, with a high probability of being right, distinguish between a good formula and a poor formula. We should strive to obtain a good formula, and, having done so, be temporarily satisfied with it, since we cannot pick out the best formula from among the many good ones. Further improvements may be possible when more data are accumulated.

As a corollary, it would appear that any efforts to insure

that a formula is the best, rather than merely good, for a particular sample, are probably wasted. Such procedures as rounding off the data to two significant figures before undertaking manual computations are therefore frequently in order; in the case of automatic computation, this procedure may not result in any saving of effort.

The most obvious danger in sampling is that a formula which appears good for one sample may be poor for the population. A less obvious but equally serious danger is that a formula which is good for the population may be overlooked not because it is poor for the sample, but merely because it is not the best for the sample. We cannot completely escape these dangers, but we should try to minimize the probability encountering them.

At first glance it might seem that the greater the number of predictors, the greater the probability of obtaining a good prediction formula, and, indeed, this would be so if the sample used in establishing the formula could consist of the entire population. When the size of the sample is limited, however, the use of too many predictors can lead to trouble. The difficulty is that the greater the number of predictors, the greater the probability that some linear combination of these predictors will be highly correlated with the predictand within the sample, even though it may be uncorrelated with the predictand within the population. Corresponding to this combination there will be terms in the prediction formula; these terms will increase the reduction of variance within the original sample, but will probably increase the error when the formula is applied to new data.

When formula (6) is applied to a new sample, and the symbol $\sim$ is used to denote a mean value over the new sample, the relation

$$\tilde{r}^2 = \tilde{x}^{*2} - 2\sum_{m=1}^{M} c_m \tilde{p_m^* x^*} + \sum_{m,n=1}^{N} c_m c_n \tilde{p_m^* p_n^*} \qquad (10)$$

does not reduce to a form analogous to (9). The term $\tilde{x}^{*2}$ is no longer the variance of x (unless $\tilde{x} = \bar{x}$), but is the mean-square error in x before prediction, i.e., the mean-square error which would result if x were always predicted to equal $\bar{x}$ (a prediction sometimes called "climatology"). The term $\tilde{r}^2$ is the mean-square error after prediction, and the ratio of the remaining terms on the right of (10) to $\tilde{x}^{*2}$, which replaces the reduction of variance as a measure of the goodness of the prediction, may be called the reduction of error. This ratio is not the only possible measure; sometimes $\tilde{r}^2$ is compared with $\overline{x^{*2}}$ instead of $\tilde{x}^{*2}$.

If we assume that population means do exist, we may let S_0 and R_0 be the reduction of variance and the ratio of the unexplained to the total variance, within the population, so that $S_0 + R_0 = 1$. We may also let S' be the expected reduction of variance within a sample by the best formula for the sample, and let S'' be the expected reduction of error when the best formula for one sample is applied to another sample. Under the assumption that the samples are randomly chosen, we can show that approximately

$$S' \sim S_0 + \frac{M}{N-1} R_0 , \qquad (11)$$

and

$$S'' \sim S_0 - \frac{M}{N+1} R_0 \quad . \tag{12}$$

Thus a considerable discrepancy is to be expected between the reduction of variance and the reduction of error. In many studies the samples are not randomly chosen, since serial correlations are present, and the expected discrepancy is even greater. The derivation of relations (11) and (12) is presented in the appendix, since it does not seem to be readily available in a single paper, although relations (11) and (12) themselves appear to be well known among statisticians. In particular, relation (11) is essentially equivalent to the statement that one should divide the sum of squares of the error by the number of degrees of freedom rather than the number of observations, to obtain an unbiased mean-square error.

It is nearly certain, then, that if many studies are performed, in which the number of predictors is a considerable fraction of the number of observations, some of these studies will yield high sample reductions of variance, especially if low or moderate reductions are present in the population. One is therefore justified in regarding with some skepticism any such formula which has not been tested with an independent sample. There is a high probability that the formula will fail in this test.

The most obvious way to lessen the dangers inherent in sampling is to increase the size of the sample. When this is impossible, it is imperative that the number of predictors be restricted. The question of a suitable method for accomplishing this end then arises.

Perhaps the simplest way to restrict the number of predictors is to choose only a small number in the first place. This was done by Wadsworth and Bryan [11] in some of the first systematic studies in statistical weather prediction. The studies gave reductions of variance of about 50 percent in the 24 hour prediction of daily mean pressure, and over 70 percent in the 24 hour prediction of daily mean temperature, at individual stations. In each case the predictors consisted of daily mean pressures or temperatures one, two, or three days previously, at a small number of nearby stations. Similar reductions of error were found when the formulas were applied to independent samples. Increasing the number of predictors did not appear to increase the reduction of error, possibly for the reasons discussed in this section.

Sometimes the number of predictors believed to be important is not sufficiently small. During the course of the Synoptic Climatology Project at M.I.T. (see Sellers [9]), Malone and Miller [5; 8] attempted to predict the instantaneous sea-level pressure field over the United States and adjacent regions, using as predictors the same field one day and two days previously. This field was represented by the pressures at a set of 91 grid points, and the sample consisted of these pressures on each of 155 January days. If all available pressures on each of two days had been used as predictors, there would have been 182 predictors, and 155 observations of each, and perfect prediction would have occurred within the sample (the least squares method would not have yielded a unique formula), while the probability of success with an independent sample would have been small.

To avoid this situation, Malone and Miller used a method introduced by Wadsworth [11] (see also [9]). The 91 pressures for each day were first represented by their mean and standard deviation, and the normalized coefficients (Z's) of 14 Tschebyscheff orthogonal polynomials. The number of predictors was thus reduced to 32, so that the expected sample reduction of variance would have been 0.22 if all the data had been completely independent.

Actually the sample reduction of the total variance at the 66 grid points used for verification was 0.69. From this one might have estimated a population reduction of variance of about 0.60, and a reduction of error of about 0.50 upon applying the formula to an independent sample. Actual application of the formula to one month of independent data yielded a reduction of error of 0.45. Although this reduction may not be great enough for practical forecasting, it at least indicates that the statistical method may become practical when further refinements are made.

The representation of a map by 16 parameters is fairly satisfactory when only two maps are to be used as predictors, but not so suitable for a larger number, for example, two maps at each of three levels, unless the sample size can be increased. Moreover in the study of Malone and Miller it appeared that certain Z's were highly correlated with each other, so that redundancy was present in the representation. An additional method of representing a map, or a set of maps, by a small number of quantities is therefore to be desired. A method of accomplishing this task, and the results which it has yielded, form

the remainder of this study.

3. Representation of a set of predictors

One systematic method of reducing the number of predictors at hand, before using them to predict, consists of determining a smaller number of quantities, in such a way that the original set of predictors may be approximated by a set of linear combinations of the new quantities, and then using the new quantities as predictors. In this section our problem is that of determining suitable new quantities.

Consider a set of M predictors $p_1(t),---,p_m(t)$, each observed at the N times $t_1,---,t_n$. Let

$$V = \sum_{m=1}^{M} \overline{p_m^2} \tag{13}$$

be the sum of the variances, or total variance, of the predictors. Let $q_1(t)---,q_k(t)$ be any K quantities, where $K < M$, and let

$$p_m^*(t_i) = \sum_{k=1}^{K} y_{km} q_k(t_i) + r_m(t_i) , \tag{14}$$

where, as in (6), the coefficients y_{km} are to be chosen to minimize the value of $\overline{r_m^2}$ for each m, and hence to minimize the quantity

$$R = \sum_{m=1}^{M} \overline{r_m^2} , \tag{15}$$

which then becomes the total "unexplained" variance of the predictors. Although R is a minimum with respect to choices of y_{km}, it still depends upon the choice of the quantities q_k. The problem at hand is that of choosing the quantities q_k to minimize the minimum value of R. The quantity $(V - R)/V$ then becomes the fraction of the total variance which may be represented by K quantities.

For this purpose, let

$$p_m(t) = \sum_{k=1}^{M} Y_{km} Q_k(t) \quad , \tag{16}$$

where Y_{km} and $Q_k(t)$ satisfy the relations

$$\sum_{m=1}^{M} Y_{km} Y_{jm} = \delta_{kj} \equiv \begin{cases} 1 & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases} \tag{17}$$

and

$$N \overline{Q_k^* Q_j^*} = a_k \delta_{kj} \quad , \tag{18}$$

and where $a_k \geqq a_{k+1} \geqq 0$.

It will then follow that the quantities $Q_1^*, \text{---}, Q_k^*$, if they exist, satisfy the requirements for the desired functions $q_1, \text{---}, q_k$. Moreover

$$V = \frac{1}{N} \sum_{k=1}^{M} a_k \quad , \tag{19}$$

and

$$V - R = \frac{1}{N} \sum_{k=1}^{K} a_k \quad . \tag{20}$$

A proof of these results is presented in the appendix.

To show that quantities Y_{km} and $Q_k(t_i)$ satisfying (16), (17), and (18) always exist, and to describe a method for determining them, it is convenient to use matrix notation. If P, P^*, Q, and Q^* are matrices of N rows and M columns whose elements are $p_m(t_i)$, $p_m^*(t_i)$, $Q_k(t_i)$, and $Q_k^*(t_i)$, and if Y is a square matrix of order M whose elements are Y_{kj}, the problem consists of expressing P in the form

$$P = QY \quad , \tag{21}$$

where

$$YY' = I \quad , \tag{22}$$

and

$$Q^{*'}Q^* = D \quad , \tag{23}$$

where I is the identity matrix, D is a matrix whose non-diagonal elements vanish and whose diagonal elements are arranged in decreasing order, and a prime (') indicates that the rows and columns of a matrix have been interchanged. Equations (21), (22), and (23) are identical in meaning with (16), (17), and (18).

To determine Y and Q, let

$$A = P^{*'}P^* \tag{24}$$

be the matrix whose elements $N\overline{p_j^* p_k^*}$ are proportional to the covariances of the predictors. If (21), (22), and (23) hold, it follows that

$$Q = PY' \quad , \tag{25}$$

and

$$YAY' = D \quad . \tag{26}$$

Conversely, if Y satisfies (22) and (26), then Q as defined by (25) satisfies (21) and (23).

The problem of solving (22) and (26) for Y and D, given A, is a classical one. It is referred to variously as matrix diagonalization, or determination of characteristic roots and vectors (or latent roots and vectors, or eigenvalues and eigenvectors) of a matrix, and may be solved by numerous methods. A convenient method for both manual and automatic computation, originally used by Jacobi [3], is described in the appendix.

Once having found Y, we can determine Q from (25), or equivalently, determine the functions Q_k from the equations

$$Q_k(t) = \sum_{m=1}^{M} Y_{km} p_m(t) \quad . \tag{27}$$

We may then use K functions Q_k (which we shall call "Q's") as predictors, in place of M functions p_m (which we shall call "p's"). A numerical example, illustrating the determination of A, D, Y, and Q from P, and the representation of M p's by K Q's, is presented in the appendix.

The idea of representing a large number of quantities by a smaller number has been pursued in statistical studies in fields other than meteorology. Among psychologists, in particular, it is known as factor analysis (see [27]). The scheme presented here is equivalent to one of the possible schemas for factor analysis.

Let us see whether this scheme answers the objections presented in the previous section. Certainly the number of predictors has been decreased, so that the discrepancy between the expected reduction of

variance and the expected reduction of error has been reduced. But let us compare a prediction formula based upon K Q's with one based upon M p's . Since M Q's and M p's completely determine one another, the prediction formula based upon M p's is identical with the one based upon M Q's , and we may compare the formula based upon K Q's with the one based upon M Q's .

Each of the p's is likely to contain some noise. Here we use "noise" in a broad sense, and include not only errors in observation, but also any feature of a predictor which shows little or no relation to other features of the predictors, or to different observations of the same feature. For example, in certain studies the anomalous pressure at a station accompanying an isolated thunderstorm would be regarded as noise.

According to (27) the Q's with small variances may be regarded as the small residuals in approximate linear relations connecting the p's . They are therefore likely to consist almost entirely of noise, like many other quantities which are small differences between larger quantities. There is a certain probability, however, that some of these Q's will be highly correlated with the predictand, within the sample, and will therefore enter the prediction formula with large coefficients. They will then, if they are mostly noise, probably lead to large errors when the formula is applied to a new sample. These errors could be eliminated simply by not using these Q's as predictors.

These same remarks do not apply to the Q's with large variances, since, although they may contain as much noise as the other

Q's , they should contain less noise relative to their total variance. They are therefore more likely to represent real physical features of the set of predictors, and any real physical relations involving them should appear in both the sample and the population. They should therefore lead to reductions of error when the prediction formula is applied to a new sample.

Thus it may be expected that if the sample is of only moderate size, a prediction formula based upon the first K Q's will yield a higher reduction of error than one based upon all M Q's , within the population, even though it cannot yield so high a reduction of variance within the original sample.

4. Representation of weather patterns and situations

It sometimes happens that the desired predictor is a continuous field of some physical quantity, for example, sea-level pressure. Of necessity the field must be described numerically by a finite number of quantities, such as pressures at a chosen set of points, or perhaps parameters in an analytic expression. The more precisely we attempt to describe the field, the more we encounter the difficulties which arise when there are too many predictors. To avoid these difficulties, we may first describe the field as precisely as the data permit, or as is convenient, by the pressures at a large number of points. We then represent these pressures by a small number of Q's , in accordance with (16).

Let s_m denote the m^{th} point, and let $Y_k(s_m) = Y_{km}$. The

functions $Q_k(t_i)$ and $Y_k(s_m)$ will be called empirical orthogonal functions, of time and space, respectively. Briefly we shall call them "Q's" and "Y's". The orthogonality is expressed by equations (17) and (18).

The procedure for determining Q and Y, given P, described in the previous section and illustrated in the appendix, is routine. It may therefore be programmed for automatic computation. In the work of the Statistical Forecasting Project at M.I.T. this procedure has been programmed for computation by Whirlwind I, the electronic digital computer at M.I.T. The matrix multiplication program necessary to obtain A from P, and Q from P and Y, has been written by Miss Kelley, Mr. Huschke, and Mr. Brun. The matrix diagonalization program, which will handle any symmetric matrix up to the 64^{th} order, has been written by Mr. Shorr. These programs have also been combined by Mr. Shorr into a single continuous routine, which will yield A, D, Y, and Q as output, if P is the input.

In a preliminary study, the data used were the reported values of sea-level pressure, at a set of 64 stations in the United States and southern Canada, at 1230 Z, during February, in the years 1949-1953. From these 64 p's (one for each s_m) the variances of the first 16 Q's, and the values of the first 8 Y's and Q's, were determined, according to the procedure described in the previous section.

Fig. 1 shows the 64 stations used in the study. Table 1 presents the variance of each of the first 16 Q's, and the cumulative variance, relative to the sum of the variances of the 64 p's. Apparently

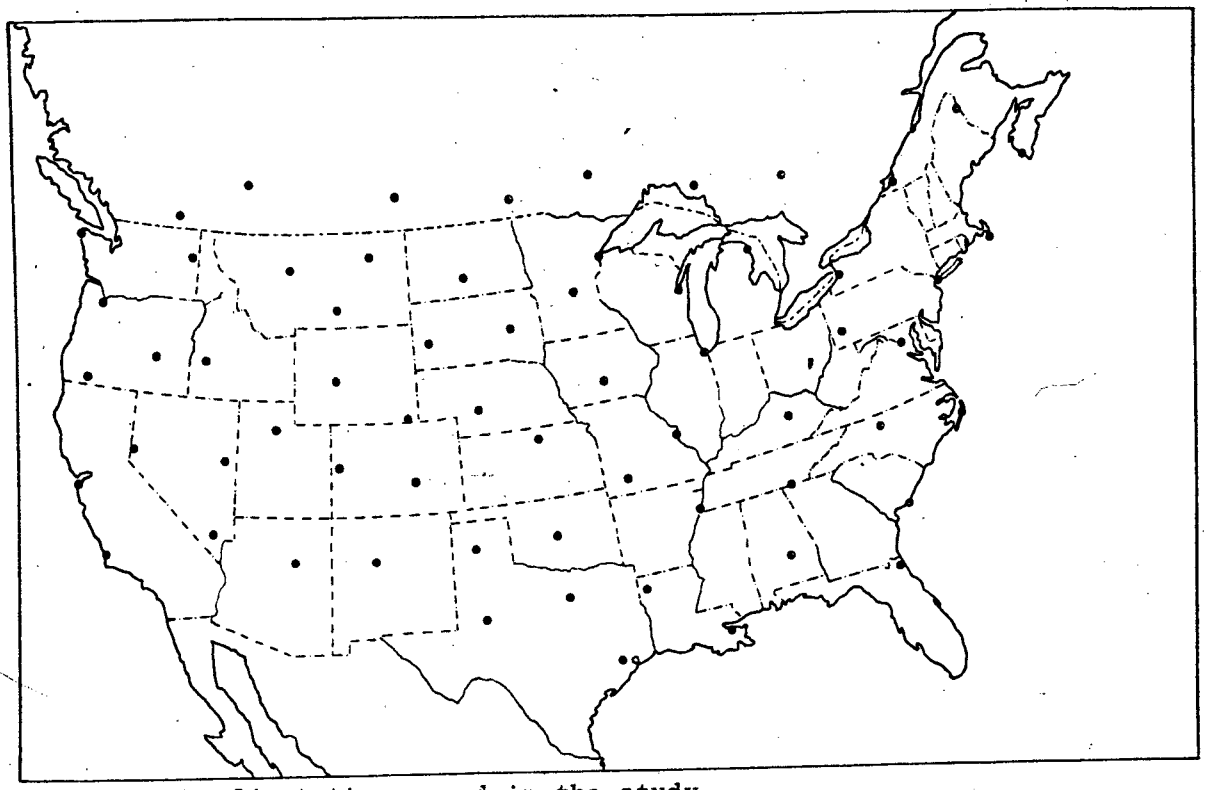


Fig. 1. The 64 stations used in the study.

the pressure field may be fairly accurately described by a small number of Q's ; eight Q's specify 91 percent of the variance, while sixteen specify 97 percent.

The first eight Q's , as defined by means of the data for February 1949-1953, were also computed for a new sample, consisting of similar data for February 1947-1948. Table 1 also includes the mean-square of each function Q_k^* , and the cumulative mean-square, relative to the sum of the mean-squares of the 64 functions p_m^* . (In the new sample the mean-square does not equal the variance because the mean values of p_m^* and Q_k^* are not zero.) In terms of mean-squares, eight Q's again specify 91 percent of the map; evidently their ability to

Table 1 . Fraction of the total variance of the sea-level pressure field (dependent data, Feb., 1949-53), and fraction of the total mean-square of the sea-level pressure field (independent data, Feb., 1947-48), represented by the empirical orthogonal function Q_k , and by k empirical orthogonal functions $Q_1,---,Q_k$, for various values of k .

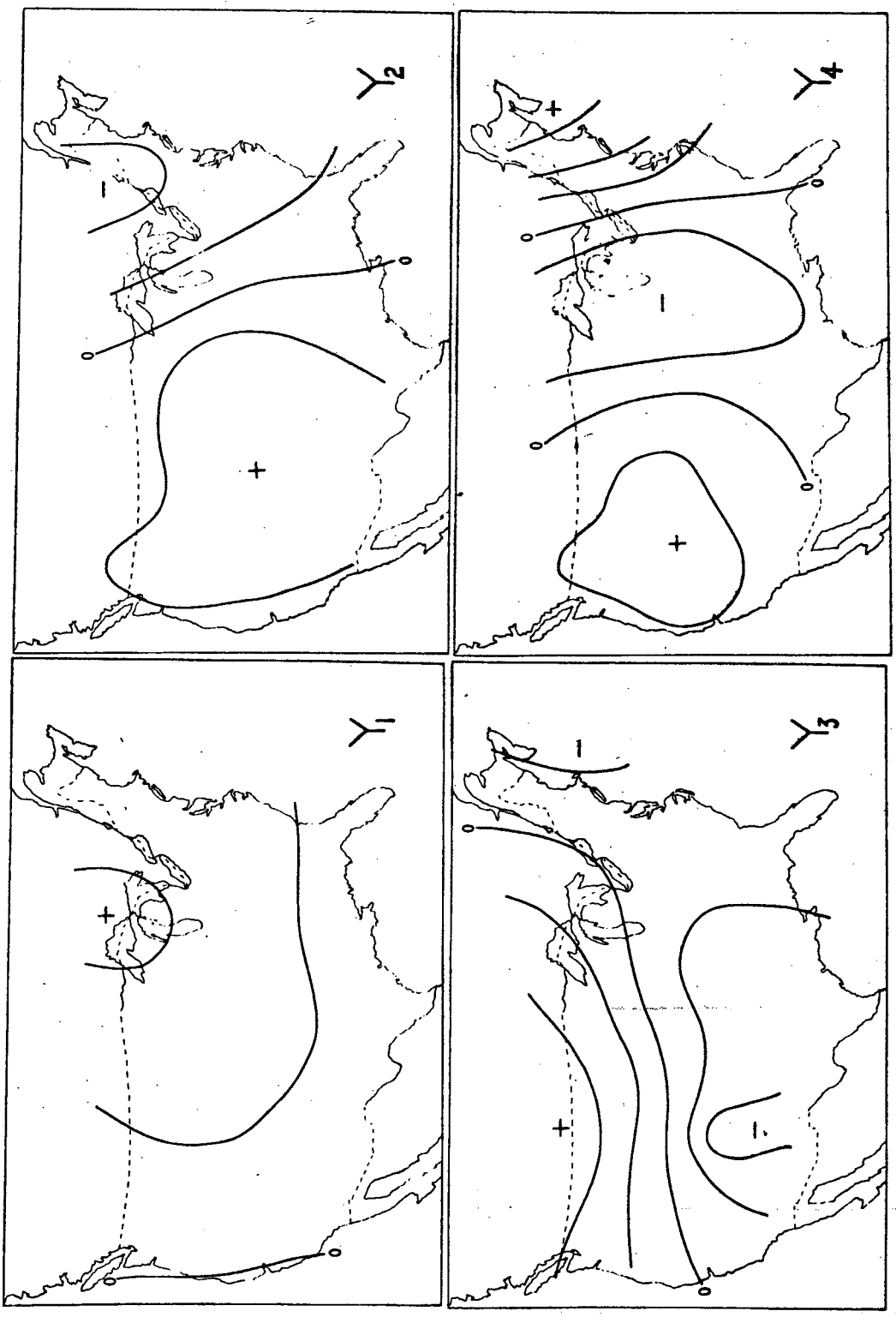
k	1949-53		1947-48	
	variance of Q_k (relative to total variance of map)	total variance of $Q_1,---,Q_k$	mean-square of Q_k^* (relative to total mean-square of map)	total mean-square of $Q_1^*,---,Q_k^*$
1	.302	.302	.239	.239
2	.170	.472	.274	.513
3	.127	.599	.109	.622
4	.117	.716	.125	.747
5	.076	.792	.038	.785
6	.060	.852	.071	.856
7	.036	.888	.033	.889
8	.022	.910	.023	.912
9	.014	.924		
10	.013	.937		
11	.008	.945		
12	.007	.952		
13	.005	.957		
14	.005	.962		
15	.004	.966		
16	.004	.970		

represent the map is not limited to the original sample.

In fig. 2, which was prepared by Mr. Shorr, the first eight Y's are shown as functions of space. A noteworthy feature is that the wave-lengths in the patterns tend to become smaller as the index of Y increases. Thus the information which is included in the first few Q's pertains primarily to larger-scale features, and these Q's may be regarded as circulation indices. The discarded information, contained in the later Q's , pertains primarily to smaller-scale features; it may perhaps be regarded as a form of turbulence, of a scale slightly larger than that of the features which are lost between observing stations. The emergence of the Q's with large variance as circulation indices is additional reason for regarding them as suitable choices for predictors.

There is no reason why the original M predictors need be confined to a single level, a single time, or a single weather element. Thus the points s_m may vary in three dimensions rather than two, or in four dimensions if time is regarded as a dimension. The predictors may also be heterogeneous; any weather elements to which numerical values may be assigned may serve as p's in equation (16). In this case, before variances are added, some decision must be made as to how many millibars of pressure, for example, are to be given the same weighting as one degree of temperature. A study of this sort, in which the fields of five-day-mean height and five-day-mean temperature are sumultaneously represented by a set of Q's , is being performed by Elizabeth Kelley.

Thus we may, if we so desire, represent not merely a pressure


Y_2
Y_4
Y_1
Y_3

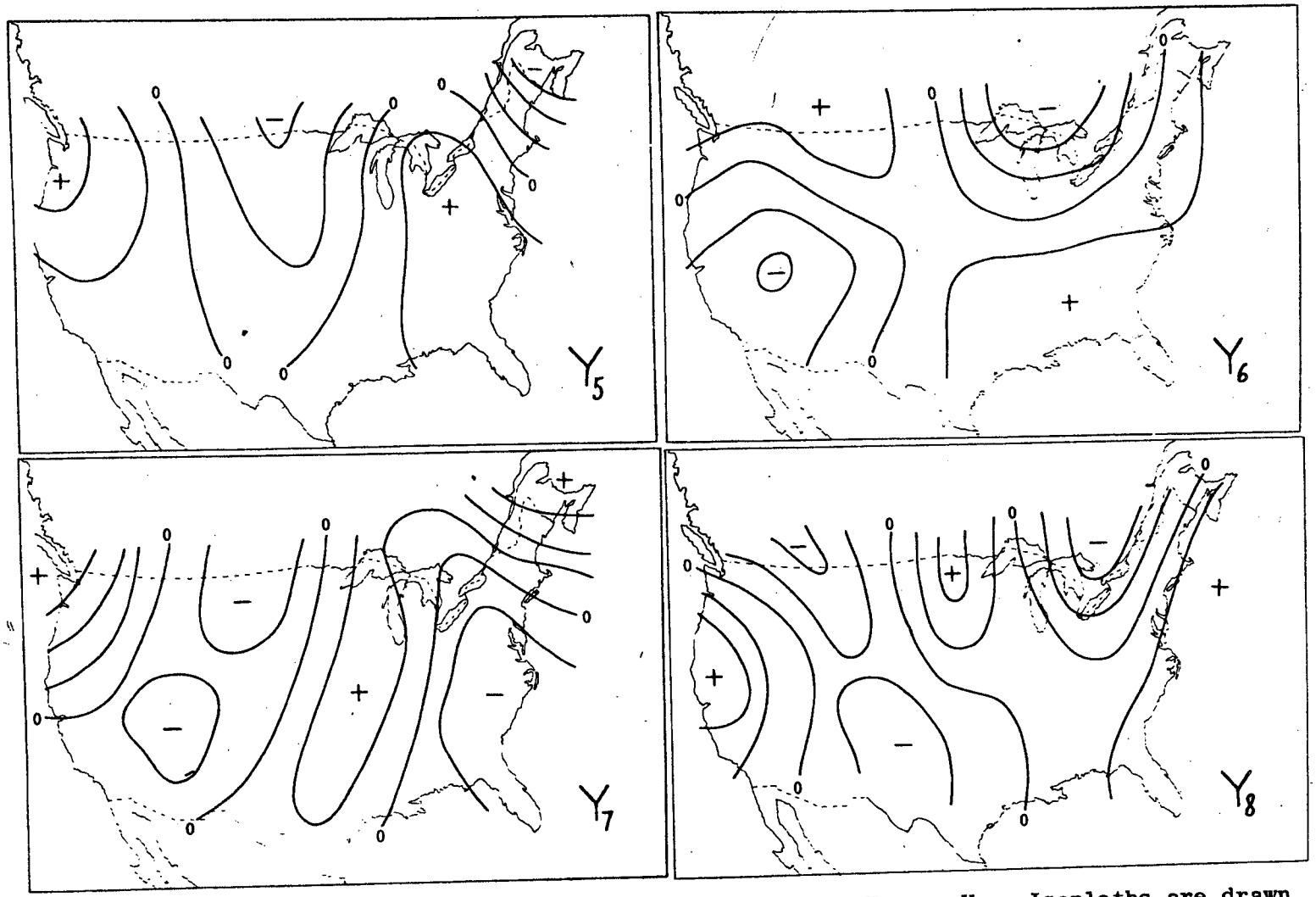


Fig. 2. Maps of the empirical orthogonal functions of space, $Y_1, ---, Y_8$. Isopleths are drawn at intervals of 0.1 unit. The zero isopleths are labeled. The units are chosen so that the sum of the squares of a function, at the 64 stations is 1.

field but an entire weather pattern or weather situation by means of empirical orthogonal functions.

5. Prediction of the sea-level pressure field

The first prediction undertaken was that of the sea-level pressure field, as defined by the pressures at the network of 64 stations, using as a predictor the sea-level pressure field twenty-four hours earlier. In accordance with the previous discussion, the field of predictors was represented by a set of eight Q's . Such a scheme, although yielding smaller reductions of variance within the original sample than would be obtained by using 64 p's , or equivalently 64 Q's , as predictors, may be expected to yield larger reductions of error when the formulas are applied to new samples.

The most direct procedure would be to establish a separate prediction formula for each of the 64 p's , using the eight Q's as predictors. Since the Q's are uncorrelated with each other, the prediction formula of the form of (6) reduces to

$$p_m = \overline{p_m} + \sum_{k=1}^{K} a_k^{-1} N \overline{p_m^* Q_{k(-)}^*} \, Q_{k(-)}^* + r_m \quad . \tag{28}$$

Here the minus signs in parentheses denote that the accompanying values of Q_k occur one day earlier than the corresponding value of p_m .

An alternative procedure is to use the Q's as both predictands and predictors. The prediction formula is then

$$Q_j = \overline{Q_j} + \sum_{k=1}^{K} a_k^{-1} N \overline{Q_j^* Q_{k(-)}^*} \, Q_{k(-)}^* + R_j \quad . \tag{29}$$

The predicted pressure field is then reconstructed from the predicted Q's by the approximation

$$p_m(t) \sim \overline{p_m} + \sum_{k=1}^{K} Y_{km} Q_k^*(t) \quad . \tag{30}$$

The question then arises as to whether there are any advantages, other than convenience, in choosing a small number of Q's as predictands. From (28), (29), and (16) it follows that

$$r_m = \sum_{j=1}^{M} Y_{jm} R_j \quad , \tag{31}$$

so that the error in predicting p_m directly is identical with the error in p_m obtained by predicting all the Q's and reconstructing the pressure field. At first it might appear that we could gain nothing in predicting p_m by discarding some of our information concerning p_m. However, if the portion of p_m represented by the Q's with low variance is largely unpredictable within the population, but at the same time appears to be predictable within the sample, we may actually lessen the errors in prediction, within the population, by not attempting to predict the unpredictable. The Q's may thus play an important part as predictands as well as predictors.

The optimum number of Q's as predictands need not be the same as the optimum number as predictors, and there is no obvious simple

way of determining these optimum numbers. Accordingly, the reduction of variance of the pressure field, as represented by J Q's and predicted by K Q's , has been determined for all values of J and K up to eight. The formulas established for the original sample (1949-1953) have then been applied to the independent sample (1947-1948), and reductions of error have been computed.

The reductions of variance and of error appear in table 2. For the original sample, the reductions of variance necessarily increase (or remain the same) as either J or K increases. A 50 per cent reduction results from the use of eight predictands and eight predictors.

For the new sample, nothing appears to be gained by choosing more than four predictands or more than five predictors. The 31 percent reduction of error when $J = 4$ and $K = 5$ is not surpassed for other values of J and K up to 8. No tests were made for J and K greater than 8, but it is doubtful that additional predictands could greatly improve the prediction, since they account for but 9 per cent of the total variance, and presumably only a small fraction of this 9 per cent is predictable. On the other hand, the errors in predicting the remaining Q's could well be larger than the Q's themselves. Although the new sample used is too small for drawing final conclusions, it is not unlikely that the optimum number of Q's , both as predictors and predictands, is less than 8. The optimum numbers are dependent upon the size of the original sample; if a larger sample had been chosen, there would have been a higher probability of obtaining stable relations among the Q's , and the optimum numbers would presumably have been higher.

Table 2 . Reduction of total variance (Feb., 1949-53) and reduction of total error (Feb., 1947-48), of the sea-level pressure field, as represented by $Q_1,---,Q_J$, and predicted by $Q_1,---,Q_K$, for all values of J and K up to 8. Figures are in hundredths.

	1949-53								1947-48							
J =	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
K=1	14	15	15	15	15	15	16	16	11	13	14	14	14	14	14	14
2	14	18	20	21	22	22	23	23	10	20	21	24	22	21	21	21
3	16	23	27	28	30	31	31	31	10	21	24	27	25	22	22	22
4	17	29	33	35	37	38	39	39	08	24	27	29	26	23	23	23
5	17	29	33	38	40	41	41	42	08	24	28	31	29	25	25	25
6	19	31	36	40	42	44	45	45	06	22	26	29	26	26	26	26
7	19	31	36	41	44	46	46	47	06	22	26	28	26	25	25	25
8	20	32	37	43	46	48	49	50	07	23	28	30	29	30	30	30

The 31 percent reduction of error obtained is hardly sufficient for the formula to be practical However, the purpose of this initial study was to examine the feasibility of using empirical orthogonal functions, and the results clearly justify their use. The reduction of error cannot be directly compared with that of other studies, since it is a reduction of the average error at all 64 stations, including those with unfavorable locations near the boundary of the data, while the physical quantities used as predictors have been restricted to pressure at one level at one time. It is reasonable to believe that when further studies are performed, with larger samples, and with more physical quantities as predictors, formulas of practical value can be established if empirical orthogonal functions are used.

6. Further uses of empirical orthogonal functions

The possible uses of empirical orthogonal functions in the general problem of weather prediction far exceed the particular purpose for which they were introduced, namely, the specification of a set of predictors by a smaller set. Some of these possibilities occur because the functions provide a convenient method of describing a weather map or situation by a small set of numbers. Thus, for example, empirical orthogonal functions might be used for choosing analogues or defining weather types. However, there are two general areas in which empirical orthogonal functions seem to offer special promise.

One of these is nonlinear statistical prediction. There are at least two reasons for believing that good nonlinear prediction formulas are attainable, and that among these may be quadratic formulas. On the one hand, as we have already mentioned, certain simplified forms of the dynamic equations express future states of the atmosphere as quadratic functions of the present state. On the other hand, in attempting to establish linear prediction formulas, one often obtains the impression that one formula holds during one type of "weather regime", and another formula holds during another type. If this impression is justified, so that the prediction coefficients are expressible in terms of a regime, which in turn is expressible in terms of the present weather pattern, the separate linear formulas may be regarded as a single nonlinear formula. In particular, this nonlinear formula is quadratic if the prediction coefficients are linear functions of the weather pattern.

If we attempt to use the $\frac{1}{2}M(M+1)$ products of M original predictors as additional predictors, the total number of predictors is hopelessly large. If the M original predictors are first represented by K Q's, the $\frac{1}{2}K(K+1)$ products of these Q's will represent the additional predictors. This number may still be too large, but in that case, it may be further reduced by letting

$$Q_k^*(t_i)Q_j^*(t_i) = \sum_{h=1}^{\frac{1}{2}K(K+1)} Y_{hkj}^0 Q_h^0(t_i) \quad , \tag{32}$$

where the Y^{0}'s and Q^{0}'s satisfy relations analogous to (17) and (18). We may then choose a small number of Q^{0}'s to supplement the Q's as predictors.

The other area in which empirical orthogonal functions seem to offer much promise is dynamical prediction (which is also nonlinear). The usual method of integrating a system of dynamic equations is a numerical one, in which one or more partial differential equations are replaced by a set of difference equations, whose dependent variables are the values of one or more weather elements at a set of grid points. If these dependent variables are represented by a small number of Q's, as in equation (30), the difference equations may be transformed into a new set, in which the dependent variables are the Q's themselves. This new set of equations may then be integrated numerically. A study of this sort, in which the original dynamic equation is a form of the barotropic vorticity equation, and the original dependent variables are values of 500-millibar height, is currently being performed by William Sellers.

The possible advantages of this method over the usual method of integrating the dynamic equations are several. First of all, in replacing the original dependent variables by a small set of Q's , we are apparently eliminating some of the smaller-scale features of the weather pattern, which seem to be unpredictable by simple statistical methods. These features probably are not unpredictable by dynamic methods, but they are likely to be unpredictable from the available initial data by means of the simplified dynamic equations which have given such gratifying results in predicting the larger-scale features. If this is the case, the elimination of the smaller-scale features should improve the prediction.

A further advantage is simplicity. Not only is the number of difference equations greatly reduced, but the maximum increment of time compatible with computational stability should be increased, since the smaller-scale fluctuations are eliminated. Finally, in approximating the predicted weather pattern by a small number of Q's , we increase the probability that the boundary values, if not predicted by the original dynamic equations, will be compatible with the dynamically predicted values in the interior.

APPENDIX

1. Expected mean-square errors

Given a population E containing M independent predictors $p_1,---,p_M$ and a predictand x. Let $p_0 \equiv 1$. Let F be any particular sample of size N. Let a single bar denote the mean value of a quantity over F, and let a double bar denote the mean value over E.

Let the set of all samples of size N be divided into subsets, such that within any subset the values of $p_1,---,p_M$ are the same for every sample, while the values of x vary from sample to sample. Let square brackets denote the mean value of a statistic over all samples within the subset containing F, and let braces denote the mean value over all samples of size N, i.e., the expected value.

Since predictions are unaltered if the predictors are replaced by independent linear combinations of themselves, we may assume without loss of generality that

$$\overline{\overline{p_m p_n}} = \delta_{mn} \qquad \text{for } m, n = 0,---,M . \tag{33}$$

We may also assume that $\bar{\bar{x}} = 0$. By diagonalizing the matrix of order $M + 1$ whose elements are $\overline{p_m p_n}$, we can obtain linear combinations $Q_0,---,Q_M$ of $p_0,---,p_M$ such that

$$\overline{\overline{Q_m Q_n}} = \delta_{mn} , \tag{34}$$

$$\overline{Q_m Q_n} = \overline{Q_m^2}\,\delta_{mn} \quad , \tag{35}$$

and

$$\sum_{m=0}^{M} \overline{Q_m^2} = \sum_{m=0}^{M} \overline{p_m^2} \quad . \tag{36}$$

We may define prediction coefficients b_m and c_m and residuals ρ and r by the relations

$$x = \sum_{m=0}^{M} b_m Q_m + \rho \quad , \tag{37}$$

where

$$\overline{Q_m \rho} = 0 \quad , \tag{38}$$

and

$$x = \sum_{m=0}^{M} c_m Q_m + r \quad , \tag{39}$$

where

$$\overline{Q_m r} = 0 \quad . \tag{40}$$

We wish to compare $\left\{\overline{r^2}\right\}$ and $\left\{\overline{\overline{r^2}}\right\}$ with $\left\{\overline{\overline{\rho^2}}\right\}$.

It follows from (37) and (39) that

$$\rho = \sum_{m=0}^{M} (c_m - b_m)\, Q_m + r \quad , \tag{41}$$

and hence from (40) and (35) that

$$\rho = \sum_{m=0}^{M} \left(\overline{Q_m^2}\right)^{-1} \overline{Q_m \rho}\; Q_m + r \quad . \tag{42}$$

From (42), (40), and (35),

$$\overline{r^2} = \overline{\rho^2} - \sum_{m=0}^{M} \left(\overline{Q_m^2}\right)^{-1} \left(\overline{Q_m \rho}\right)^2 , \tag{43}$$

while from (42), (38), and (34)

$$\overline{\overline{r^2}} = \overline{\overline{\rho^2}} + \sum_{m=0}^{M} \left(\overline{Q_m^2}\right)^{-2} \left(\overline{Q_m \rho}\right)^2 . \tag{44}$$

Assuming that there are no population relations between ρ and the predictors, and that samples are randomly chosen, we find that

$$\left[\left(\overline{Q_m \rho}\right)^2\right] = \frac{1}{N^2} \sum_{i,j=1}^{N} Q_m(i) Q_m(j) \left[\rho(i)\rho(j)\right] = \frac{1}{N} \overline{Q_m^2}\ \overline{\overline{\rho^2}} . \tag{45}$$

Hence

$$\left[\overline{r^2}\right] = \overline{\overline{\rho^2}} \left(1 - \frac{M+1}{N}\right) \tag{46}$$

and

$$\left[\overline{\overline{r^2}}\right] = \overline{\overline{\rho^2}} \left(1 + \frac{1}{N} \sum_{m=0}^{M} \left(\overline{Q_m^2}\right)^{-1}\right) . \tag{47}$$

The value of $\left[\overline{\overline{r^2}}\right]$ depends upon the particular sample F, while that of $\left[\overline{r^2}\right]$ does not. It follows from (46) that

$$\left\{\overline{r^2}\right\} = \overline{\overline{\rho^2}} \left(1 - \frac{M+1}{N}\right) , \tag{48}$$

and, since the arithmetic mean of the reciprocals of a set of positive numbers exceeds the reciprocal of the arithmetic mean, it follows from (47), (36), and (33) that

$$\left\{\overline{\overline{r^2}}\right\} > \overline{\overline{\rho^2}}\left(1 + \frac{M+1}{N}\right) , \tag{49}$$

Equations (48) and (49) are the desired comparisons.

For the special case $M = 0$, whence $Q_0 = 1$, $x = \rho$, and $x^* \equiv x - \bar{x} = r$, (48) and (49) reduce to the familiar relations

$$\left\{\overline{x^{*2}}\right\} = \left(1 - \frac{1}{N}\right)\overline{\overline{x^2}} , \tag{50}$$

$$\left\{\overline{\overline{x^{*2}}}\right\} = \left(1 + \frac{1}{N}\right)\overline{\overline{x^2}} . \tag{51}$$

If we approximate the expected value of a quotient by the quotient of the expected values, we obtain from (48), (50), (49), and (51) the approximate relations (11) and (12) in the text of the report.

2. The optimum representation of a set of functions

Given a set of M functions $p_1(t),---,p_M(t)$. Let a bar denote the mean, and let a star denote a departure from the mean, with respect to the independent variable t. For any K functions $q_1(t),---,q_K(t)$, where $K < M$, let

$$p_m^* = \sum_{k=1}^{K} y_{km} q_k + r_m \tag{52}$$

where the coefficients y_{km} are chosen so that

$$\overline{q_k r_m} = 0 \qquad \text{for} \quad k = 1,---,K; \quad m = 1,---,M \quad . \tag{53}$$

We wish to determine the set of K functions $q_1,---,q_K$ which minimizes the total unexplained variance

$$R = \sum_{m=1}^{M} \overline{r_m^2} = \sum_{m=1}^{M} \overline{p_m^{*2}} - \sum_{m=1}^{M} \overline{\left(\sum_{k=1}^{K} y_{km} q_k \right)^2} \quad . \tag{54}$$

Let

$$p_m = \sum_{j=1}^{M} Y_{jm} Q_j \quad , \tag{55}$$

where

$$\sum_{m=1}^{M} Y_{km} Y_{jm} = \delta_{kj} \quad , \tag{56}$$

and

$$\overline{Q_k^* Q_j^*} = a_k \delta_{kj} \quad , \qquad \text{where} \quad a_k \geq a_{k+1} \geq 0 \quad . \tag{57}$$

We shall show that the functions $Q_1^*(t),---,Q_K^*(t)$ satisfy the requirements for $q_1,---,q_K$.

The functions $q_1,---,q_K$ which minimize R are not uniquely determined, since any K independent linear combinations serve the purpose equally well. We may therefore restrict our choice of functions so that

$$\overline{q_k Q_j^*} = 0 \qquad \text{if} \quad k > j \tag{58}$$

and, retaining this restriction, further restrict our choice so that

$$\overline{q_k q_j} = 0 \qquad \text{if} \quad k \neq j \quad . \tag{59}$$

From (52), (53), and (59),

$$y_{km} = \left(\overline{q_k^2}\right)^{-1} \overline{q_k p_m^*} \quad , \tag{60}$$

and equation (54) reduces to

$$R = \sum_{m=1}^{M} \overline{p_m^{*2}} - \sum_{m=1}^{M} \sum_{k=1}^{K} \left(\overline{q_k^2}\right)^{-1} \left(\overline{q_k p_m^*}\right)^2 . \tag{61}$$

Let

$$q_k = \sum_{j=k}^{M} b_{jk} Q_j^* + \rho_k \tag{62}$$

where

$$\overline{Q_j^* \rho_k} = 0 \qquad \text{for} \quad j = 1,---,M; \quad k = 1,---,K \quad . \tag{63}$$

The lower limit $j = k$ in the summation is possible because of restriction (58).

From (55) and (62), with the aid of (56), (57), and (63), equation (61) further reduces to

$$R = \sum_{j=1}^{M} a_j - \sum_{k=1}^{K} \left(\frac{\sum_{j=k}^{M} a_j^2 \, b_{jk}^2}{\sum_{j=k}^{M} a_j \, b_{jk}^2 + \overline{\rho_k^2}} \right) . \tag{64}$$

We must choose the quantities b_{jk} and ρ_k, without violating restriction (59), to minimize R, or equivalently, to maximize the fraction in parentheses in (64) for each value of k.

Temporarily neglecting restriction (59), we can, for any choice of b_{jk}, maximize the fraction by choosing $\rho_k \equiv 0$. The

fraction then becomes the ratio of a weighted average value of a_j^2 (with $j \geq k$) to a similarly weighted average value of a_j; its maximum value, namely a_k, will occur if $b_{kk} = 1$ and $b_{kj} = 0$ when $j > k$. Substituting these choices of ρ_k and b_{kj} into (62), we find that the minimum value of R, namely

$$R = \sum_{j=1}^{M} a_j - \sum_{k=1}^{K} a_k = \sum_{k=K+1}^{M} a_k \tag{65}$$

occurs when

$$q_k = Q_k^* \quad \text{for} \quad k = 1,---,K \ . \tag{66}$$

Since this value of q_k does not violate (59), our conclusion is verified.

3. Matrix diagonalization by Jacobi's method

Given a square symmetric matrix A of order M. We wish to determine a square matrix Y such that

$$YY' = I \ , \tag{67}$$

and

$$YAY' = D \ , \tag{68}$$

where I is the identity matrix and D is a square matrix whose non-diagonal elements vanish.

Let the non-diagonal element of A of largest absolute value be A_{jk} (with $j < k$).

Let θ satisfy the equation

$$\tan 2\theta = 2A_{jk} \Big/ (A_{jj} - A_{kk}) \tag{69}$$

(with $0 \leqq \theta < \pi/2$). Let Y_1 , be a square matrix of order N in which

$$\left.\begin{aligned} (Y_1)_{jj} &= (Y_1)_{kk} = \cos\theta \\ (Y_1)_{jk} &= (-Y_1)_{kj} = \sin\theta \\ (Y_1)_{mm} &= 1 \qquad \text{unless} \quad m = j \quad \text{or} \quad m = k \\ (Y_1)_{mn} &= 0 \qquad \text{otherwise} \ . \end{aligned}\right\} \tag{70}$$

Then

$$Y_1 Y_1' = I \tag{71}$$

while

$$Y_1 A Y_1' = A_1 \tag{72}$$

is a symmetric matrix in which the diagonal sum equals that of A ; the sum of the squares of the non-diagonal elements is $(2A_{jk}^2)$ less than that of A , and $(A_1)_{jk} = 0$.

Now select the non-diagonal element of A_1 of largest absolute value, and repeat the process, determining a matrix Y_2 . After L iterations,

$$(Y_L Y_{L-1} \cdots Y_2 Y_1)(Y_L Y_{L-1} \cdots Y_2 Y_1)' = I \ , \tag{73}$$

while

$$(Y_L Y_{L-1} \cdots Y_2 Y_1) A (Y_L Y_{L-1} \cdots Y_2 Y_1)' = A_L \ , \tag{74}$$

is a symmetric matrix whose diagonal sum equals that of A . The non-diagonal elements of A_L may be made as small as desired by choosing L large enough, so that

$$Y = \lim_{L \to \infty} (Y_L Y_{L-1} \cdots Y_2 Y_1) \tag{75}$$

satisfies (67) and (68).

4. Numerical example

Suppose that the observed sea-level pressures, in millibars, at three stations on five days have been tabulated as follows:

	station 1	station 2	station 3
day 1	1028	1022	1019
day 2	1026	1025	1015
day 3	1020	1020	1010
day 4	1009	1015	1013
day 5	1012	1008	1023

The matrix P is then simply the array of numbers in this table, arranged just as they are in the table. To obtain P^* from P subtract from each number the mean value of the numbers in that column, so that

$$P^* = \begin{pmatrix} 9 & 4 & 3 \\ 7 & 7 & -1 \\ 1 & 2 & -6 \\ -10 & -3 & -3 \\ -7 & -10 & 7 \end{pmatrix}$$

Each element of $A = P^{*\prime}P$ is obtained by multiplying corresponding numbers in two columns of P^*, and summing; thus $A_{11} = (9)^2 + (7)^2 + (1)^2 + (-10)^2 + (-7)^2 = 280$, $A_{12} = (9)(4) + (7)(7) + (1)(2) + (-10)(-3) + (-7)(-10) = 187$, etc, so that

$$A = \begin{pmatrix} 280 & 187 & -5 \\ 187 & 178 & -68 \\ -5 & -68 & 104 \end{pmatrix}$$

To determine Y_1 and A_1 from A, let A_{jk} be the non-diagonal element of A of largest absolute value (with $j < k$), so that $j = 1$, $k = 2$, $A_{jj} = 280$, $A_{kk} = 178$, and $A_{jk} = 187$. Determine $\cos 2\theta$ by the relation

$$\cos 2\theta = \pm \frac{A_{jj} - A_{kk}}{\sqrt{(A_{jj} - A_{kk})^2 + (2A_{jk})^2}} = \frac{102}{\sqrt{15028}} = 0.2631 ,$$

the sign being positive if $A_{jj} - A_{kk}$ and A_{jk} have the same signs, and negative otherwise. Then determine $\cos \theta$ and $\sin \theta$ by the relations

$$\cos \theta = \sqrt{\tfrac{1}{2}(1 + \cos 2\theta)} = \sqrt{.6315} = 0.795 ,$$

$$\sin \theta = \sqrt{\tfrac{1}{2}(1 - \cos 2\theta)} = \sqrt{.3685} = 0.607 .$$

Next, write the identity matrix I, and write the matrix A below it. Augment both matrices by two new columns, the first of these being

$$(\cos \theta \times j^{th} \text{ column} + \sin \theta \times k^{th} \text{ column}) ,$$

and the record being

$$(-\sin \theta \times j^{th} \text{ column} + \cos \theta \times k^{th} \text{ column}) .$$

Similarly, augment matrix A, but not I, by two rows similarly determined, so that the augmented matrix A becomes symmetric. Then strike out the j^{th} and k^{th} columns in both matrices, and the j^{th} and k^{th} rows in A, but not in I. The result is

1.000	0	0	.795	-.607
0	1.000	0	.607	.795
0	0	1.000	0	0

280	187	-5	336	-21
187	178	-68	257	28
-5	-68	104	-45	-51
336	257	-45	423	0
-21	28	-51	0	35

The numbers enclosed in rectangles are those which have not been struck out. The upper of these arrays of numbers is the matrix Y_1' , and the lower is A_1 . The order in which the columns and rows occur has been changed, but this is immaterial.

The largest non-diagonal element of A_1 is next selected, and the procedure is repeated, until the non-diagonal elements are as small as desired. The procedure is illustrated below (with the decimal point omitted from I).

iteration	1	2	3	4	5
A_{jj}	280	104	423	8	126
A_{jk}	187	-51	40	-21	3
A_{kk}	178	35	131	429	7
cos 2θ	.2631	-.5603	.9645	.9950	.9987
cos θ	.795	.469	.991	.999	1.000
sin θ	.607	.883	.133	.050	.025

1000	0	0	795	-607	-536	-285	750	-388	-498	776	-400	-488
0	1000	0	607	795	702	373	651	289	734	615	307	727
0	0	1000	0	0	469	-883	-117	-875	463	-140	-863	485

280	187	-5	336	-21								
187	178	-68	257	28								
-5	-68	104	-45	-51	4	-116						
336	257	-45	423	0	-21	40	425	-17				
-21	28	-51	0	35	7	61						
		4	-21	7	8	0	-21	3	7	-21		
		-116	40	61	0	131	57	125				
			425		-21	57	429	0	0	430		
			-17		3	125	0	126	3	0	126	0
					7		0	3	7	0	3	7
					-21		430	0	0	431	0	0
								126	3	0	126	0
								0	7	0	0	7

In this example, within the limits of the round-off error, five iterations are sufficient to reduce the non-diagonal elements of A to zero. The original matrices I and A, and the final matrices Y' (above) and D (below), have been enclosed in rectangles. The increase in the diagonal sum of A by two units is the result of round-off error. The numbers which at the beginning of each step were non-diagonal elements of largest absolute value have been underlined.

The number of iterations necessary to diagonalize A should increase roughly as rapidly as the square of the order of A, while the number of multiplications per iteration increases approximately as the first power. In manual computation the labor per multiplication may also increase with the order of A, since more significant figures must be carried when the number of iterations is large. The total labor therefore probably increases at least as rapidly as the cube of the order.

The matrix Y, the transpose of Y', is

$$Y = \begin{pmatrix} .776 & .615 & -.140 \\ -.400 & .307 & -.863 \\ -.488 & .727 & .485 \end{pmatrix}$$

so that

$$YY' = \begin{pmatrix} 1.000 & .004 & .001 \\ .004 & .999 & .000 \\ .001 & .000 & 1.002 \end{pmatrix}$$

which approximates the identity I, the small discrepancies resulting from round-off error.

Next,

$$Q^* = P^*Y' = \begin{pmatrix} 9.0 & -5.0 & 0 \\ 9.9 & 0.2 & 1.2 \\ 2.8 & 5.4 & -1.9 \\ -9.2 & 5.7 & 1.2 \\ -12.6 & -6.3 & -0.5 \end{pmatrix}$$

As in the original tabulation of observations, each row refers to one day, and each column refers to one empirical orthogonal function. The large variance of the first column and the small variance of the last are immediately apparent. The sums of squares within the columns are 430.2, 126.4, and 6.7, which agree with the diagonal elements of D, except for round-off error.

Since the first two orthogonal functions explain nearly 99 per cent of the variance of the three pressures, it is possible to approximate the p's as linear combinations of two Q's. Thus

$$\hat{P}^* = \begin{pmatrix} 9.0 & -5.0 \\ 9.9 & 0.2 \\ 2.8 & 5.4 \\ -9.2 & 5.7 \\ 12.6 & -6.3 \end{pmatrix} \begin{pmatrix} .776 & .615 & -.140 \\ -.400 & .307 & -.863 \end{pmatrix} = \begin{pmatrix} 9.0 & 4.0 & 3.1 \\ 7.6 & 6.2 & -1.6 \\ 0.0 & 3.4 & -5.0 \\ -9.4 & -3.9 & -3.6 \\ -7.2 & -9.6 & 7.2 \end{pmatrix}$$

The first two matrices in this relation are Q^*, with the final column omitted, and Y, with the final row omitted. Comparing $\hat{P}^*$, the approximation for P^*, with P^*, we see that in only one instance is the discrepancy greater than 1.0 millibar. Thus the approximation should be satisfactory for most purposes.

ACKNOWLEDGMENT

The results presented in this report could not have been accomplished without the teamwork of the entire M.I.T. Statistical Forecasting Project. The data used were gathered and checked for accuracy by Mrs. Heidi Basler and Mr. Matthew Mulkern. The various portions of the program for automatic computation were written by Miss Elizabeth Kelley, Mr. Bernard Shorr, Mr. Ralph Huschke, and Mr. Herbert Brun. The automatic computations were performed on Whirlwind I, operated by the Scientific and Engineering Computation Group of the Digital Computer Laboratory at M.I.T. Many of the manual computations were performed by Miss Kelley and Mrs. Basler. The figures were constructed by Mr. Shorr, and drafted by Miss Isabelle Kole. Finally, the writer has greatly benefited from numerous meetings with Dr. Thomas Malone and Mr. Robert Miller of the Travelers Weather Research Center, and from several discussions with Prof. George Wadsworth and Dr. Joseph Bryan of M.I.T.

REFERENCES

1 Crout, P. D., 1941: A short method for evaluating determinants and solving systems of linear equations with real or complex coefficients. Trans. Amer. Inst. Elec. Eng., 60, 1235-1241.

2 Holzinger, K. J., and Harman, H. H., 1941: Factor Analysis. Chicago, U. of Chicago Press, 417 pp.

3 Jacobi, C. G. J., 1846: Über ein leichtes Verfahren die in der Theorie der Säcularstörungen vorkommenden Gleichungen numerisch aufzulösen. J. reine angewandte Math., 30, 51-95.

4 Kramer, M. P , and Rigby, M , 1954: Annotated bibliography on numerical forecasting. Meteor. Abstr. and Bibliog. 5, 734-758.

5 Malone, T. F., 1955: Application of statistical methods in weather prediction. Proc. Nat. Acad. Sci., 41, 806-815.

6 Malone, T. F., 1956: Application of synoptic climatology to weather prediction. Chapter 28 in Petterssen, Weather analysis and forecasting, New York, Mc Graw-Hill Book Co., 238-255.

7 Miller, R. G., 1956: Statistics and the predictability of weather. Abstract in Bull. Amer. Meteor. Soc., 37, 382.

8 Miller, R. G., and Malone, T. F., 1954: Application of Synoptic Climatology to an objective technique for preparing sea-level prognostic pressure maps. Abstract in Bull. Amer. Meteor. Soc., 35, 495-496.

9 Sellers, W. D., 1956: Studies in synoptic climatology. Final report, contract N5 ori-07883 between O.N.R. and M.I.T., 215 pp.

10 Wadsworth, G. P., 1951: Application of statistical methods to weather forecasting. Compendium of Meteorology, Boston, Amer. Meteor. Soc., 849-855.

11 Wadsworth, G. P., and Bryan, J. G., 1948: Short range and extended forecasting by statistical methods. Air. Wea. Serv. Tech. Rep. no 105-37, 202 pp.

12 White, R. M., 1956: Pressure-height predictability as a function of the type and amount of initial data. J. Meteor., 13, 40-45.

13 White, R. M., and Galligan, A. M., 1956: The comparative accuracy of certain statistical and synoptic forecasting techniques. Bull. Amer. Meteor. Soc., 37, 1-7.

14 White, R. M., and Palson, W. C., 1955: On the forecasting possibilities of empirical influence functions. J. Meteor., 12, 478-485.

15 Wiener, N., 1954: Non-linear prediction and dynamics. Unpublished manuscript.