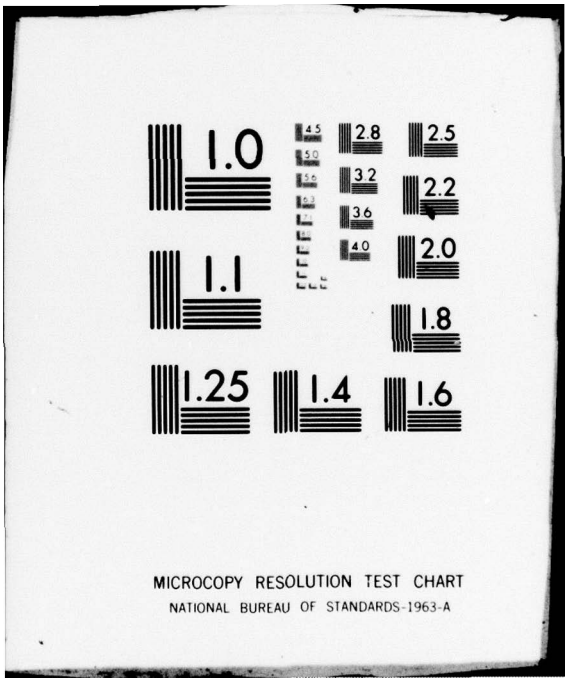
AD-A080 056 NAVAL ENVIRONMENTAL PREDICTION RESEARCH FACILITY MON--ETC F/G 4/2
A USE OF TIME SERIES IN IMPROVING WEATHER FORECASTING.(U)
NOV 79 F D FAULKNER, C COMSTOCK, R R FOSSUM
UNCLASSIFIED NEPRF-CR-79-07 NL

END
DATE FILMED
2-80
DDC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

NAVENVPREDRSCHFAC
CONTRACTOR REPORT
CR 79-07

LEVEL II

A USE OF TIME SERIES IN IMPROVING WEATHER FORECASTING

Prepared By:

Frank D. Faulkner, Craig Comstock, Robert R. Fossum

U.S. Naval Postgraduate School
Monterey, California 93940

N6658679 WR79064

NOVEMBER 1979

Prepared For:
NAVAL ENVIRONMENTAL PREDICTION RESEARCH FACILITY
MONTEREY, CALIFORNIA 93940

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NAVENVPREDRSCHFAC Contractor Report CR 79-07	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) A Use of Time Series in Improving Weather Forecasting.		5. TYPE OF REPORT & PERIOD COVERED (9) Final rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Frank D. Faulkner, Craig Comstock, Robert R. Fossum		8. CONTRACT OR GRANT NUMBER(s) N6685679WR79064 (17)
9. PERFORMING ORGANIZATION NAME AND ADDRESS U. S. Naval Postgraduate School Monterey, CA 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE 62759N, PN WF52551792 NEPRF WU 6.2-6
11. CONTROLLING OFFICE NAME AND ADDRESS Commander, Naval Air Systems Command Department of the Navy Washington, DC 20361		12. REPORT DATE (11) November 1979 13. NUMBER OF PAGES 26
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Naval Environmental Prediction Research Facility Monterey, CA 93940		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report) (16) F52551 (12) 27

Approved for public release; distribution unlimited

1

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(14) NEPRF-CR-79-07

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Time series
Regression analysis
Weather prediction improvement

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Given the assumption that weather can be predicted more effectively by integrating the dynamic equation(s) than by examining trends, the question investigated here is whether time series can be used in the secondary routine to effect improvement. The method is described, the relevant equations are derived, a program using real meteorological data is made and run, and a measure of effectiveness is given. Limited study shows some definite promise.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

407 279

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

CONTENTS

1. INTRODUCTION

The purpose of this study is to investigate a use of time series in improving weather forecasting.

Let us consider a simple model. Let us assume that the relevant variables, or functions, such as, say, the 500 mb height field, can be expressed in terms of a set of orthogonal basis functions, whose coefficients define the field. There is also an associated numerical integration program which yields predictions of this variable in the form of predicted coefficients. This method of forecasting is assumed to be more reliable than any regression scheme. At regular intervals some observations of the variable are made and the predicted coefficients are then updated, or corrected.

These corrections to the coefficients in the forecast are analyzed for trends. The corrections for each (complex) coefficient of a (complex) eigenfunction or basis function are analyzed separately. The purpose is to predict the corrections.

The basic approach is to define a best-fitting difference equation and a best set of initial conditions, by minimizing weighted sums of squares of the residuals, and then use the difference equation and the initial values to predict the correction. These are used in a secondary routine to improve the forecast.

Testing of the method in the 500 mb field with a simple non-divergent model due to Bourke (1972) shows skill with some coefficients.

2. OUTLINE OF THE PROBLEM

Let us assume that we have some function, or functions, which we are representing in terms of an orthonormal set of basis functions. In particular, let us consider the 500 mb height function, or perhaps, the stream function, expressed on the surface of the earth in terms of spherical harmonics. These are expressed as complex trigonometric functions of longitude and Legendre functions of the sine of the latitude. At time t_N the coefficients of the spherical harmonics form a vector or column matrix A_N; its components are complex numbers $a_{k,N}$, k=1, ..., K, if the series is truncated to K terms. The vectors A_N, and each of the components $a_{k,N}$, form a time series, or sequence.

The sequence is generated conceptually as follows. Let us assume that at time t_{N-1} we have the matrix A_{N-1} of coefficients for the analyzed field. This is considered to represent the true field at that time. We have also an integration procedure by which we can generate predictions $A_{N-1,N}$, $A_{N-1,N+1}$, ..., for the future values A_N, A_{N+1}, ..., as many as we feel are useful or reliable. Then at time t_N, 12 hours later, observations are made. These are combined with the predictions $A_{N-1,N}$ to yield a new analyzed field, and a new set of coefficients A_N. The corrections (changes, errors, or discrepancies),

$$Z_n = A_n - A_{n-1,n}, \quad n = 2, 3, \ldots, N, \tag{2.1}$$

also define a time series or a sequence. If we could in some way analyze this sequence, Z_2, Z_3, .., Z_N, and from it predict Z_{N+1}, Z_{N+2}, ..., Z_{N+m} reliably then we could predict ahead m steps

reliably. Our purpose is to predict more accurately, or more reliably, and perhaps to extend the range of prediction. We shall see that there are two ways, at least, that this improvement might be effected. Both depend on the use of time series as developed in the next sections.

The Z's and/or their components form the data for the analysis. As such they are called observations (not to be confused with meteorological observations). The terms component and coefficient are used interchangeably. Vectors have components and basis functions have coefficients when they are used in the representation of another function. The components $a_{k,n}$ of A_n, for example, are the coefficients in the representation of some meteorological variable. Basis functions are often called modes, particularly in mechanics.

3. INTRODUCTORY TIME SERIES, SIMPLIFYING ASSUMPTIONS

Let us consider the sequence of corrections, Z_n, $n = 2, 3, \ldots, N$. The vector Z_n has K complex components. There is one complex number $z_{k,n}$ for each of the K complex-valued spherical harmonics that is used. To save writing, let z_n, rather than $z_{k,n}$, denote a typical coefficient; we will analyze each of these sequences individually. Part of the rationale is that in a linear uncoupled system the various modes move independently; in a loosely coupled system we would expect similar behavior. There are other important reasons which we will take up later. We shall usually consider z_n to be a pair of real numbers, a two-by-one matrix.

Let us now assume that z_n can be expressed as a sum of two terms, described below,

$$z_n = x_n + y_n. \tag{3.1}$$

Here y_n is a random variable. Its components each are assumed to have expected value zero and expected square σ^2, independent of n; they are also assumed to be uncorrelated. The vector x_n is assumed to be a solution to a difference equation of the form

$$x_n = C_1 x_{n-1} + \ldots + C_p x_{n-p}, \quad n = p + 2, \ldots, N. \tag{3.2}$$

The coefficients $C_1, \ldots, C_p$ are unknown two-by-two matrices, and the order p is to be chosen some way. For the present let us take p = 2, which seems at this time to be a likely value. The C's are not constants necessarily, but may vary adaptively with N.

To determine the coefficients in (3.2), let us choose the C's to minimize a residual of the form

$$R = \frac{1}{2} \sum_{4}^{N} w_n \Big(z_n - \sum_{1}^{p} C_j z_{n-j}\Big)^2, \tag{3.3}$$

where the w_n are positive weight factors to be chosen or determined; we will discuss this later. To minimize R let us differentiate with respect to the elements of the C's; we get a system of equations of the form

$$\sum_{4}^{N} w_n \begin{pmatrix} z_{n-1}z'_{n-1}, & \ldots, & z_{n-1}z'_{n-p} \\ & \ldots & \\ z_{n-p}z'_{n-1}, & \ldots, & z_{n-p}z'_{n-p} \end{pmatrix} \begin{pmatrix} C'_1 \\ \ldots \\ C'_p \end{pmatrix} = \sum_{4}^{N} w_n \begin{pmatrix} z_{n-1} \\ \ldots \\ z_{n-p} \end{pmatrix} z'_n \tag{3.4}$$

the prime (') indicates the transpose of a matrix. The terms like $z_{n-1}z'_{n-p}$ represent two-by-two matrices so that the elements of the matrices are themselves matrices of order two.

Another problem is the choice of the weight factors w_n. For a time-independent process we should use equal weightings, which simplifies several relations. However the fundamental relations may vary rather rapidly with time sometimes, and these are the very times that are most critical. One way to take account of this is to use exponential weightings in (3.3) and (3.4). After some starting routine we weight each new observation with a uniform value w and decrease all earlier weightings then by a factor 1-w. We do this as follows. Let Q_N be the matrix of coefficients in (3.4); it has two-by-two block elements of the form

$$Q_{i,j,N} = \sum_{p+2}^{N} w_n z_{n-i} z'_{n-j}, \quad i,j=1, \ldots, p; \tag{3.5}$$

the right-hand side of (3.4) has elements $Q_{i,0,N}$. When we get the observation z_N, we update $Q_{i,j,N-1}$ by the relation

$$Q_{i,j,N} = Q_{i,j,N-1} + w(z_{N-i}z'_{N-j} - Q_{i,j,N-1}). \tag{3.6}$$

We may actually update Q by first shifting each block element down and to the right one place, dropping out the last row and the last column. Then we update the first (block) row, using (3.6), and then the first column by symmetry. The weight of an observation from time t_n for the coefficients calculated at time t_N is $w(1-w)^{N-n}$; we see that weightings decrease exponentially with time.

The best choice of w is another problem. If w is too small the system is slow to respond to changes, and if w is too large, random errors cause excessive errors in the predicted values.

The above procedure for updating the coefficients in (3.2) is rather efficient, both in terms of storage and computation. The matrix Q is always singular until at least 2p observations have been made. It is usually positive definite after that, but it was sometimes found to be near singular, especially for the smaller values of N.

In order to predict the z's ahead, using (3.2), we also need p starting or initial values; we can take these conceptionally as estimates of x_N, x_{n-1}, ..., x_{N-p+1}. The determination of these is taken up in the next section; it requires more storage and more computation than the above.

In the following section we will generally consider the case p = 2.

There may be some minor discrepancies in the first index; for example, when the a's are involved we tend to think of z_2 as the first of the z's, and later, when treating the z's we think of the first as z_1.

4. ROUTINE TO DETERMINE THE BEST FILTERED OR INITIAL VALUES FOR PREDICTION

Let us assume that we have found the coefficients C_j in the difference Eq. (3.2). We also need estimates of x_N and x_{N-1}, say $\hat{x}_N$, $\hat{x}_{N-1}$, in order to predict ahead using the difference equation. We might just use the last two observations, z_N, z_{N-1}, but these involve individual random errors, and we expect smoothed or filtered values to be more reliable. By using a large value for the weighting parameter w we can drive the smoothed values close to the last two observations, if we wish.

Let us resolve the problem as follows. Let us consider

$$\begin{aligned} y_n &= z_n - x_n && (n = 1, \ldots, N) \\ &= z_n - (C_1 x_{n-1} + C_2 x_{n-2}) && (n = 3, \ldots, N) \\ &= z_n - (C_1[C_1 x_{n-2} + C_2 x_{n-3}] + C_2[C_1 x_{n-3} + C_2 x_{n-4}]) \\ &= \ldots, \end{aligned}$$

which we may rewrite in the form

$$y_n = z_n - A_n x_2 - B_n x_1 , \qquad (n = 1, \ldots, N); \qquad (4.1)$$

A_n and B_n are defined by this relation. It is easily verified that they satisfy the relations

$$\begin{cases} A_n = C_1 A_{n-1} + C_2 A_{n-2} & n = 1,2, \ldots, N, \\ A_1 = 0, \; A_2 = I, \\ B_n = C_1 B_{n-1} + C_2 B_{n-2} \\ B_1 = I, \; B_2 = 0. \end{cases} \qquad (4.2)$$

We can simplify the programming somewhat if we define the matrices

$$\begin{cases} D^n = (A_n, B_n) \\ U = (x_2', x_1')' = (x_{12}, x_{22}, x_{11}, x_{21})'; \end{cases} \tag{4.3}$$

x_{ij} is the i'th component of x_j. Then D^n satisfies Eq. (3.2) and has the initial values

$$\begin{cases} D^n = C_1 D^{n-1} + C_2 D^{n-2} \\ D^1 = (0, I),\ D^2 = (I, 0). \end{cases} \tag{4.4}$$

Equation (4.1) now reduces to

$$y_n = z_n - D^n U. \tag{4.5}$$

If we knew the four components of U, we could get starting values $\hat{x}_N$, $\hat{x}_{N-1}$ from Eq. (3.2).

$$x_n = C_1 x_{n-1} + C_2 x_{n-2}, \tag{4.6}$$

and if we select or estimate U in some way we get corresponding estimates for x_N, x_{N-1}. Of course it would be nicer to determine these in a more direct way, but there seems to be no way to do this.

Now, by assumption, the expected value, $E\, y_i = 0$. Hence it seems reasonable that the best choice for U is one which minimizes a weighted sum of squares for the estimates of y_i, say,

$$R = \frac{1}{2} \sum^{N} w_n y_n^2 = \frac{1}{2} \sum^{N} w_n (D^n U - z_n)^2 \tag{4.7}$$

To effect this we differentiate with respect to the components of U; if we call its components u_j, the j'th partial derivative yields

$$\begin{aligned} R_j &= \sum^{N} w_n [d^n_{1,j} (\sum_1^4 d^n_{1,k} u_k - z_{1,n}) + d^n_{2,j} (\sum_1^4 d^n_{2,k} u_k - z_{2,n})] \\ &= 0, \end{aligned} \tag{4.8}$$

where $d^n_{j,k}$ are the components D^n. We solve these equations for U, which defines $\hat{x}_1, \hat{x}_2$, and from these we get our best estimates, $\hat{x}_{N-1}$, $\hat{x}_N$ for starting.

There are several problems. First, the sequence z_n, $n = 1, 2 \ldots, N$, may be very long, so that computing times and errors may be significant. Further the observations z_n for small values of n may not be relevant to today's weather. We also need to recall many observations. These problems suggest limiting the range of summation in (4.7) to, say, the last ten terms, which we did.

Another problem needs some explanation. We really just need $\hat{x}_{N-1}$, $\hat{x}_N$ to predict ahead. However we cannot get these directly. It really does not matter whether we find $\hat{x}_{N-1}$ and $\hat{x}_N$, or $\hat{x}_{N-9}$ and $\hat{x}_{N-8}$; we get the former from the latter by use of difference Eq. (3.2). We might perhaps try to rewrite the difference equation, to solve for x_{n-2} in terms of x_{n-1} and x_n, to avoid this. However we must invert the matrix C_2 to effect this, and there is no reason that it cannot be singular, or nearly so, randomly.

In the program we have 15 $\hat{x}$'s actually indexed from 1 to 15. The first 10 of these are the 10 approximants to the last 10 observations $z_{N-9}, \ldots, z_N$; the last five are predictions for the future values $z_{N+1}, \ldots, z_{N+5}$.

5. TWO SUGGESTED METHODS FOR IMPROVING PREDICTION BY THE USE OF A TIME SERIES

There are two particular methods which seem feasible for improving the forecast. Each of these seems feasible, but each poses some problems.

5.1 Method 1

Let us consider again the matrices for the coefficients in the analyzed fields, A_n, n=1, ..., N; we wish to improve the predictions of these. Let us assume that we have generated and stored the following. First, we have stored the last ten corrections Z_{N-9}, ..., Z_N. We have also stored, for each term in the Z's, the matrix Q which defines the coefficients C_j in the difference equation.

The routine is the following for each component of Z. As soon as the new value of Z, Z_N, is obtained, we store it and discard the old Z_{N-9}. Then we update the matrix Q and solve for the coefficients of the difference equation C_j as described in Section 3. Then we solve for the initial values $\hat{x}_{N-1}$, $\hat{x}_N$, as described in Sect. 4, and finally we predict ahead to get x_{N+1}, x_{N+2}, ..., x_{N+M}, as many as we wish. These define the $\hat{x}$'s, or $\hat{X}$'s.

Next let us use the dynamic equation and integrate ahead one time step, to get an estimate $A_{N,N+1}$, of A_{N+1}. Then let, say,

$$\tilde{A}_{N,N+1} = A_{N,N+1},$$

and

$$\hat{A}_{N,N+1} = \tilde{A}_{N,N+1} + \hat{X}_{N+1}. \qquad (5.1)$$

We get successive values in a similar way. When we have $\hat{A}_{N,N+j}$ we integrate to obtain $\tilde{A}_{N,N+j+1}$, a first estimate of $A_{N,N+j+1}$. Then we add $\hat{X}_{N+j+1}$, to get

$$\hat{A}_{N,N+j+1} = \tilde{A}_{N,N+j+1} + \hat{X}_{N+j+1}. \tag{5.2}$$

With the initial value $\hat{A}_{N,N} = A_N$, the sequence of predictions, or forecasts, is defined.

In the initial part of the study, a simple integration routine was used for predicting, $A_{N,N+M}$, a model using the non-divergent barotropic vorticity equation. When this model is used, the method of correcting above, alternately adjusting and integrating in feasible and practical.

When a more accurate and complicated weather prediction model is used, serious difficulties arise with this method. Restarting a time integration after rather arbitrary adjustments to one or more variables may generate spurious gravity waves which degrade the prediction and offset the adjustments. To eliminate this difficulty a second method may be used, in which a separate series is generated for each interval of prediction. The appropriate term is used for each interval and the adjusted functions are never integrated.

5.2 Method 2

There is another method which gets around the above problem, at the expense of increasing the storage and computation by a factor of roughly M, if we use M different intervals of prediction. That is, if we wish, say to adjust the 12-, 24-, and 36-hour forecasts we must generate three time series.

Now assume that we wish to predict ahead M steps. At the time t_N we will have the predictions based on integration,

$$A_{N,N+1}, \ldots, A_{N,N+M} = \{A_{N,N+m}\}_{m=1,M} \tag{5.3}$$

For the corrections we will save M sequences Z_n^m

$$Z_n^m = A_n - A_{n-m,n}, \quad m = 1, \ldots, M, \quad n = m+1, \ldots, N. \tag{5.4}$$

Let us consider any particular value of m. To generate the data for the time series we must store each time the uncorrected predictions $A_{n,n+m}$. Then at time t_{n+m} we will obtain and store Z_{n+m}^m from (5.4).

We will thus have M time series to be analyzed. Each is analyzed as discussed earlier. In this case however we will use a single prediction from each one: from Z_n^m we generate the single correction for m time steps ahead. If we denote the corrections by $\hat{X}_n^m$ the modified values will be

$$\hat{A}_{n,n+m} = A_{n,n+m} + \hat{X}_n^m \tag{5.5}$$

This method has the advantage that the modified terms are not integrated. It has the disadvantage that it requires one time series for each value of m used, which leads to larger storage and computational requirements.

6. COMMENTS

The problem here differs from the most common applications of time series. We have a rather small amount of information which tends to be masked by a large random element. From this we are trying to predict transients, or trends, for relatively short periods, perhaps one to ten time steps. We are not directly interested in the smoothed values, since we are not concerned with what has happened, except as our ability to fit it reflects on our ability to predict. Our feeling is that the variables we are predicting may change rather rapidly so that the weather two weeks ago is of little interest; even the equation which governs the behavior may well have changed. We cannot really make use of long-term observations, as for a stable system, which allow more accurate predictions.

We are trying to predict the short-range behavior of a part of a non-linear system of high order by solutions to a linear system of low order. Whether we try to do this by a difference equation or by Taylor series depends on the type of data, and the nature of the expected solutions. In our case both the form of the data and the anticipated periodic properties of the expected solutions suggest difference equations.

The papers by Jones (1963), (1964) particularly, and (1965) suggested the general problem here. These treat more the descriptive and the long-range forecasting problem. The series are considered to be stationary in the sense that all observations are weighted equally; several formulas are then simplified.

The references found most useful in this study were Whittle (1963) and Gelb (1974). Recent developments in the use of spherical harmonics in meteorology (see GARP, 1974) and Fast Fourier Transforms make the methods appear feasible.

The method is clearly that of least squares. It is also a linear regression method; all of the equations to be solved are linear. If we were to use, say, our estimates $\hat{x}_{N-1}$, $\hat{x}_N$ in a routine to try to improve the C's, then it would be non linear. The solutions to the difference equations are basically complex exponentials, that is, a combination of real exponentials and trigonometric functions.

There are a number of points that should be discussed or clarified.

6.1 *Consistency*

It should be pointed out that even for a stationary system the solutions are not consistent, as follows. Let us assume that we have a solution to a difference Eq. (3.1) to which random uncorrelated terms y_n, each with expected value 0 and expected square σ^2 is added. Then Eq. (3.4) will not yield the desired coefficients in (3.2). The expected elements of the matrix of coefficients Q in (3.5), from (3.4) will be

$$Q_{i,j,N} = \sum^{N} w_n (x_{n-i}\, x'_{n-j} + \delta_{ij}\, \sigma^2\, I_2) \quad i,j=1, \ldots, p; \qquad (6.1)$$

here I_2 is the second-order identity matrix, and δ is the Kronecker δ, (=1 when i=j and 0 when i≠j). That is, the Q matrices have an extra term σ^2 on the main diagonal. This suggests that if we have a measure of σ^2 we could subtract it from the elements on the main

diagonal of Q, in a sort of negative ridge-regression scheme. This is very risky procedure however, since it drives the matrix of coefficients toward singularity. Difficulty has been encountered several times because this matrix Q was singular, and the danger is aggravated when exponential weighting with a large decay rate is used.

6.2 Best Order for the Difference Equation

We have suggested that we use a second-order difference equation, p = 2. Some of the reasons follow. Let us consider a system which is time independent, and assume that have found the correct order p and the true coefficients C in the difference equations. In this case the residuals yield an estimate for σ^2. Now consider the residual

$$R = \sum_{p+1}^{N} w_n \{x_n + y_n - \sum_{1}^{p} C_j (x_{n-j} + y_{n-j})\}^2$$

$$= \Sigma\, w_n (y_n - \Sigma\, C_j y_{n-j})^2 ,$$

since the x's satisfy (3.2). The expected value of the residual is then

$$E(R) = (\Sigma\, w_n)\ (2 + \Sigma\, c^2_{j,k,i})\ \sigma^2,$$

where the last term in the parenthese denotes the sum of the squares of all of the components of all of the C's.

Now if we try to fit the solutions to a second-order system by a third-order difference equation, in the absence of noise, the matrix of coefficients is singular, since the solutions satisfy many third-order difference equations. If we have a little noise

we may have a near-singular matrix and have large values for the $c_{i,j,k}$; in this case we get a large expected value in (6.1) for the residual. Limited numerical checks make the second-order equation appear satisfactory.

6.3 Determining Initial Values

There was no particular reason for choosing ten values to fit when we determine the starting values for prediction. It was felt that six would be the smallest number to be considered, that fewer would make the methods vulnerable to random errors in the last terms. Many more than ten could cause roundoff errors.

6.4 Criteria for Adjusting Terms

We do not expect to adjust or correct all of the coefficients. For some the errors may be too small to warrant the time and trouble. Other coefficients may contain such random elements and change so unpredictably that trends are effectively obscured. The decision to adjust or correct will probably depend on an analysis of similar sequences in the past. There are two basic ways to do this, one based on a long sequence of past data (that is, on climatology), and the other on recent data (corresponding to recent weather).

We may analyze a long sequence for a coefficient, perhaps over a period of years, and from this make a decision, beforehand always to correct, or never, in the program.

An alternate way is to see how well the smoothed values fit the last ten corrections. We generate these values naturally in determining the starting values for prediction, so this entails little computation. We may consider, for example, the ratio

$$1.25 \sum_{1}^{10} w_n (z_{N-10+n} - \hat{x}_n)^2 / \sum_{1}^{10} w_n z_{N-10+n}^2 .$$

If this is much less than one we have fitted the past ten values well, and if it exceeds one we have no skill in fitting the observations. The factor 1.25 adjusts for the fact that we could fit two terms exactly with the initial conditions. We might correct or not accordingly as this number was below or above some criterion, say, .75. The philosophy is that if we can fit a sequence of observations well, then we can predict well. (More logical is the converse; if we cannot fit the observed values, then we cannot predict well.) It is probably not worthwhile to use this criterion unless we usually correct.

6.5 Criteria for Improvement

In the numerical study we used two principal criteria for improvement. Both of these compared in some way the magnitude of the errors in prediction with adjustment with the errors generated without the adjustment. In the first criterion, the error in the predicted values(s) was compared with the rms value of the last ten corrections (without adjustment). This figure was calculated as the predictions were under way. For the second criterion, a long sequence of adjustments were calculated. Then the rms values of the errors in prediction after adjustment was compared with the rms value of the error in prediction without adjustment:

$$\Sigma |z_n - \hat{x}_n|^2 / \Sigma |z_n|^2;$$

the range of summation of n is over each term for which a correction to z was made. These measures are easily made in the machine, and they define a simple norm for the error relation.

It is not clear these norms are good ones from a meteorological point of view. Very likely the ultimate test will rest on maps and criteria for judging them.

6.6 Real Coefficients

Most of the coefficients are complex. The one associated with the modes that are independent of longitude (meridianal) are real and hence one dimensional. The routines developed then involve scalar equations, rather than two-dimensional vectors. The corresponding spherical harmonics are the Legendre polynomials, the simplest of the Legendre functions.

7. RESULTS

The procedure for analyzing the time series was checked on several sets of made-up data, and on two coefficients a_2 and a_4 (of P_1^1 and P_2^2), using a difference equation of second order, $p = 2$. For data made up of solutions to a difference equation, with random noise added, the solution could be recovered if the noise was not too large, but the routine did not perform very well when the energy in the noise (rms value) was larger than in the solution.

For meteorological data, on a string of 60 pairs of corrections, the predictions x_{11} to x_{55} led to a definite improvement for one coefficient. The best value for the weighting w of the new values seemed rather small: the values .025, .05 and .1 gave similar results and were better than larger values. The numbers below were determined as follows. We considered the coefficient of P_1^1. A set of 10 data points were used at the start to get coefficients for the best difference equation and the initial

values. Five values were predicted, corresponding to the 12-hour correction to be made in 12 hours, 24 hours, ..., 60 hours. The error in predicting these was then calculated and normalized by dividing by the rms of the 10 original data. Then the time index was advanced by one and the operation repeated.

The numbers in row one are the rms values of $a_{n+1} - a_{n,n+1}$, normalized by the rms values of $z_{n-9}, z_{n-8}, \ldots, z_n$. The numbers in row m denote the rms value of $a_{n+m} - a_{n+m-1,n+m}$, normalized by the rms value of $z_{n-9}, \ldots, z_n$. Each is then a correction to be applied after integrating one time step. Each entry comes from the sum of 46 terms.

No. of Time Steps	Weight w	.025	.05	.1	.2	.4
1		.6239	.6257	.6131	.6404	.8138
2		.6370	.6401	.6312	.6629	.8640
3		.6670	.6751	.6834	.7599	1.3095
4		.6573	.6698	.6960	.8314	1.5523
5		.6513	.6636	.6880	.8496	1.9109

There are several interesting points about the data. First, the results from the small values of w, from .025 to .1 are all comparable. This indicates a stable trend: when w = .025 the first of the 10 data points had a weighting almost .8 of the last; they are weighted nearly equally. For w = .10 the relative weighting is about .39. A surprising feature is that all five of the predictions are so similar. This is further indication of a stable trend, a large "random" element, and a satisfactory value of p, and suggests longer predictions are feasible.

The corrections to the fourth coefficient were not predicted successfully; the error in the predicted values was consistently a little large than the corrections. However various features of the solutions are quite similar to the other solutions; the smaller values of w give better predictions and all five predicted values are similar in size, with later predictions often anomalously having smaller error than the first.

Acknowledgments

We are particularly indebted to Dr. C. E. Leith of NCAR, Boulder, for making available to us both a large set of data from analyzed fields and an integration program for generating data to try this method. We are also greatly indebted to our colleague, Professor H. J. Larson, for a number of clarifying discussions.

SYMBOLS: DEFINITIONS, AND PLACES WHERE INTRODUCED

Section 2

t_n, t_N	time N is usually associated with the last or next-to-last observation.
A_n	the matrix, or vector, of coefficients for the analyzed field at time t_n.
$a_{k,n}$	the k'th element of A_n, k=1, ..., K. Most of the $a_{k,n}$ are complex numbers and are treated like two-dimensional vectors.
K	the number of spherical harmonics used to represent a function.
$a_n (=a_{k,n})$	a typical element of A_n, written without index k to save writing.
A_{n,n_1}	the value predicted for A_{n_1} at time t_n.
$Z_n=A_n-A_{n-1,n}$	This represents the discrepancy between the analyzed field at t_n and the value predicted for it at time t_{n-1}. Also called a correction.

Section 3

z_n	a typical element of Z_n.
y_n	the component of z_n that is assumed to be random.
x_n	the component of z_n that is assumed to satisfy a difference equation (see below).
C_j	two by two matrices in the difference equation $x_n = C_1 x_{n-1} + \cdots + C_p x_{n-p}$.
R	weighted sum of residual errors.
w	weighting of new terms.
w_n	the weighting of term associated with t_n, at time $N \geq n$; w_n decreases with N for a fixed value of n: $w_n = w(1-w)^{N-n}$.

Section 3 (continued)

$\hat{x}_n$ — a best approximation for x_n, made at some definite time t_N. (It might also have been designated by $\hat{x}_{k,N,n}$.)

$\hat{X}_n$ — the vector, or matrix, of $\hat{x}_n$'s.

Section 4

A_n, B_n — matrices defined in Eq. (4.2), used in finding x_n's.

$D^n = (A_n, B_n)$ — (see Eqs. (4.3), (4.4)).

x_{ij} — the i'th component of x_j, i,j=1,2.

$U = (x_{12}, x_{22}, x_{11}, x_{12})'$

Section 5

$\hat{A}_{N,N+m}$ — an optimal predictor for A_{N+m}, made at t_N.

$\tilde{A}_{N,N+m}$ — a temporary estimate of $A_{N,N+m}$, used in obtaining $\hat{A}_{N,N+m}$.

$Z_n^m = A_n - A_{n-m,n}$

z_n^m — typical elements of Z_n^m.

R — a general symbol for a quadratic residual to be minimized.

REFERENCES

Bourke, William, 1973: An efficient, one-level, primitive-equation, spectral model, Mon. Wea. Rev., 100, 683-689.

GARP/WGNE, 1974: Global Atmospheric Research Program/Working Group on Numerical Experimentation, Report of International Symposium on Spectral Methods in Numerical Weather Prediction, 12-16 August 1974, Copenhagen, Rept. No. 7.

Gelb, Arthur, editor, 1974: Applied Optimal Estimation, Cambridge, MA, MIT Press.

Jones, Richard H., 1962: Stochastic Processes on a Sphere as Applied to Meteorological 500-Millibar Forecasts, Ch. 7, (Proceedings of the Symposium on) Time Series Analysis, Murray Rosenblatt, John Wiley, NY, 1862.

_____, 1964: Prediction of multivariate time series, J. Appl. Meteor., 3, 285-289.

_____, 1965: Optimal estimation of initial condition for numerical prediction, J. Atmos, Sci., 22, 658-663.

Whittle, Peter, 1963: Prediction and Regulation, Princeton. D Van Nostrand.