UNCLASSIFIED

AD 432475

DEFENSE DOCUMENTATION CENTER

FOR

SCIENTIFIC AND TECHNICAL INFORMATION

CAMERON STATION, ALEXANDRIA, VIRGINIA

UNCLASSIFIED

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

64-10

CATALOGED BY DDC
AS AD No. 432475

[redacted]

432475

MEASUREMENT ANALYSIS CORPORATION

DDC
MAR 1964

10962 SANTA MONICA BLVD • LOS ANGELES, CALIFORNIA 90025 • TELEPHONE 477-4555

PREDICTION OF TIME SERIES USING MULTIPLE REGRESSION TECHNIQUES AND SEAKEEPING APPLICATIONS (U)

Loren D. Enochson

December 1963
MAC 307-01

Sponsored by
Bureau of Ships Fundamental Hydromechanics
Research Program, S-R009 01 01
administered by DAVID TAYLOR MODEL BASIN
Contract Nonr-4305(00)

REPRODUCTION IN WHOLE OR IN PART
IS PERMITTED FOR ANY PURPOSE OF
THE UNITED STATES GOVERNMENT

MEASUREMENT ANALYSIS CORPORATION
10962 Santa Monica Boulevard
Los Angeles, California 90025

PREDICTION OF TIME SERIES USING
MULTIPLE REGRESSION TECHNIQUES AND SEAKEEPING APPLICATIONS

CONTENTS

		Page
1.	Introduction	1
2.	Regression for a Single Predictor Variable	2
3.	Regression for Two and More Predictor Variables	4
4.	Application to Time Series	6
	4.1 Computational Details	8
	4.2 Comparison with Other Procedures	11
5.	Extension to Cross-Correlated Variables	12
6.	Seakeeping Applications	15
	6.1 Long Range Ocean Activity Prediction	15
	6.2 Short Term Roll Prediction	17
	6.3 Ocean Wave Amplitude Prediction	20
	6.4 Vibration Data Application	22
7.	Statistical Tests and Errors	25
8.	Digital Computing Considerations	28
References		34

LIST OF FIGURES

		Page
Figure 1 -	Hypothetical Correlation Functions for Sea Activity Prediction Sample	16
Figure 2 -	Hypothetical Autocorrelation Function for Ship's Roll	17
Figure 3 -	Possible Digital Control System for Ship's Roll	19
Figure 4 -	Hypothetical Autocorrelation Function for Ocean Wave Data	21
Figure 5 -	Hypothetical Cross-correlation Functions for Vibration Response Example	23

ABSTRACT

This report discusses methods for predicting future values of discrete time series from past observed values of the time series. The points at which the auto-correlation function is computed are the past points used in the analysis. Classical ideas are reviewed, and then extended to handle more advanced time series problems. Applications of these results are explained for seakeeping applications of: (a) long range ocean activity prediction, (b) short term roll prediction, and (c) vibration response prediction. Further material appears on statistical tests for coefficient determination, and on digital computer requirements.

1. INTRODUCTION

In the subsequent discussion it will be assumed that a discrete random process, namely, a digitized process is being analyzed. The process will be assumed to be a time series $x(t)$, which is defined as a set of observations taken in time sequence. Much physical data is obtained in this way either directly or from digitizing the original continuous data. The procedure to be considered will be that of using statistical multiple regression techniques in order to perform a linear least squares extrapolation for the future employing the previously observed values of the time series.

The past observed values of the variable at certain special points will be thought of as separate variables. More particularly, the points at which the autocorrelation function is computed will be the special points in the regression equation. For example, if the autocorrelation function is computed at the time delay $\tau = 1$ second, 2 seconds, etc., up to n seconds, then the variables will be the value of the time series $x(t)$ at $x(-1)$ seconds, $x(-2)$ seconds, etc., up to $x(-n)$ seconds. In general, there will be n variables present in the regression equation. The final objective will be to obtain a regression equation as a function of previous values of the time series so as to predict an extrapolated value of the time series for some time in the future. This general type of analysis is sometimes termed "autoregressive" analysis.

These matters are discussed in detail in the following sections, and are illustrated on appropriate physical applications. Sections 2 and 3 review classical ideas on multiple regression techniques, while Sections 4 and 5 are devoted to more advanced time series concepts. Application of these results, particularly to seakeeping problems, are developed in Section 6. Further material on statistical tests and on digital computer requirements appears in Sections 7 and 8.

2. REGRESSION FOR A SINGLE PREDICTOR VARIABLE

As an introduction to the basic concept of regression, the case for a single predictor value x will be reviewed, Ref. [1]. Assume a random variable y exists, with mean zero, and also assume that y may be theoretically expressed as a linear function of a variable x plus a random error ϵ. That is

$$y = \beta x + \epsilon \tag{1}$$

where β is an unknown constant to be determined. The usual estimation procedure is to calculate an estimate of β by the method of least squares. Assume one has collected N observations each of x and y denoted by (x_i, y_i), $i = 1, 2, \ldots, N$. One then wants to minimize the quantity

$$\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} (y_i - bx_i)^2 \tag{2}$$

where the quantity β is replaced by b in Eq. (2) and $\hat{y} = bx$ to indicate that a sample estimate b of β will be obtained rather than the true population value β. The hat ($\wedge$) is used over the $\hat{y}$ to indicate $\hat{y}$ is a predicted or estimated value of y.

Differentiating with respect to b and equating to zero, one obtains

$$b = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2} = \frac{\Gamma_{xy} s_x s_y}{s_x^2} = \Gamma_{xy} \frac{s_y}{s_x} \tag{3}$$

where Γ_{xy} is the sample correlation coefficient given by

$$\Gamma_{xy} = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}} = \frac{\sum x_i y_i / N}{s_x s_y} \tag{4}$$

The value of b given by Eq. (3) is the optimum least square estimate of β. The sample variances s_x^2 and s_y^2 in Eqs. (3) and (4) are defined by

$$s_x^2 = \frac{\sum_{i=1}^{N} x_i^2}{N} \tag{5}$$

Then s_x, the sample standard deviation, is the positive square root of s_x^2. It should be noted that s_x^2 is a biased estimator of the true variance σ^2. For large $N > 30$, however, the bias is insignificant.

The variance of b may be shown to be

$$\mathrm{Var}(b) = \frac{\sigma_y^2}{\sum x_i^2} \tag{6}$$

Under the assumptions of an underlying normal distribution the quantity

$$u = \frac{b - \beta}{\sqrt{\sigma_y^2 / \sum x_i^2}}$$

has a normal distribution with zero mean and unit variance. Also, it may be shown, see Ref. [3], that the quantity

$$v^2 = \frac{\sum (y_i - \hat{y}_i)^2}{\sigma_y^2}$$

has a χ^2 distribution with (N - 2) d.f. Therefore, this implies that the statistic

$$t = \frac{u}{v / \sqrt{N-2}} = (b - \beta) \sqrt{\frac{(N-2) \sum x_i^2}{\sum (y_i - \hat{y})^2}} \tag{7}$$

has a t distribution with (N - 2) d.f. Hence, the hypothesis that $b = \beta$ may be tested at a desired level of significance α for a specified value of β.

3. REGRESSION FOR TWO AND MORE PREDICTOR VARIABLES

The case for one predictor variable is easily generalized. This will be illustrated first for the two variable case and then for n variables.

For two variables one desires to compute the coefficients b_1 and b_2 in the regression equation

$$\hat{y} = b_1 x_1 + b_2 x_2 \tag{8}$$

The sum of squares to be minimized for this case is

$$\sum (y_i - \hat{y}_i)^2 = \sum (y_i - b_1 x_{1_i} - b_2 x_{2_i})^2 \tag{9}$$

Differentiating partially with respect to b_1 and b_2, one obtains two simultaneous linear equations which may be solved for b_1 and b_2. These are

$$\begin{aligned} b_1 \sum_{i=1}^{N} x_{1_i}^2 + b_2 \sum_{i=1}^{N} x_{1_i} x_{2_i} &= \sum_{i=1}^{N} y_i x_{1_i} \\ b_1 \sum_{i=1}^{N} x_{1_i} x_{2_i} + b_2 \sum_{i=1}^{N} x_{2_i}^2 &= \sum_{i=1}^{N} y_i x_{2_i} \end{aligned} \tag{10}$$

The coefficients b_1 and b_2 are usually referred to as the <u>sample partial regression coefficients</u>.

The generalization to k variables x_i, $i = 1, \ldots, k$ follows directly. In this case one wants to determine k coefficients b_i, $i = 1, \ldots, k$ to obtain a regression equation

$$\hat{y} = b_1 x_1 + b_2 x_2 + \ldots + b_k x_k \tag{11}$$

The sum of squares to be minimized becomes

$$\sum (y_i - \hat{y}_i)^2 = \sum (y_i - b_1 x_{1_i} - b_2 x_{2_i} - \ldots - b_k x_{k_i})^2 \tag{12}$$

The result after minimizing the sum of squares is k simultaneous equations which may be solved for the b_i, namely,

$$
\begin{aligned}
b_1\sum x_{1_i}^2 + b_2\sum x_{1_i}x_{2_i} + \ldots + b_k\sum x_{1_i}x_{k_i} &= \sum y_i x_{1_i} \\
b_1\sum x_{1_i}x_{2_i} + b_2\sum x_{2_i}^2 + \ldots + b_k\sum x_{2_i}x_{k_i} &= \sum y_i x_{2_i} \\
\vdots \qquad\qquad \vdots \qquad\qquad\qquad \vdots \qquad & \qquad (13) \\
b_1\sum x_{1_i}x_{k_i} + b_2\sum x_{2_i}x_{k_i} + \ldots + b_k\sum x_{k_i}^2 &= \sum y_i x_{k_i}
\end{aligned}
$$

One should be careful to contrast this model with a correlation analysis. Here it is assumed that y is a random variable of which readings are taken as the x_i are varied through predetermined values. In this case the x_i are not assumed to be arbitrary random variables. In contrast, for a classical statistical correlation analysis, one usually assumes a joint $(k+1)$ dimensional normal distribution where the y variable and the x_i variables, $i = 1, \ldots, k$, are all assumed to be random. That is, one has not controlled the variables x_i but has performed some experiment where all the variables have random outcomes. In this case one wants to estimate all the correlations between the variables. In other words, one wants to estimate the correlation or covariance matrix of the $(k+1)$ dimensional normal distribution.

To explain further, in a true correlation analysis one collects observations on all the variables of interest where none of the variables are controlled but only observed. However, in a true regression analysis one would control one variable, say for example pressure, and then observe another, say temperature, as pressure was stepped through a predetermined range of values. In this case, only the temperature variable would exhibit random fluctuations. In practice, however, one is not usually able to control variables through predetermined ranges as is theoretically required. This gives rise to no real practical problems, though, since if underlying normal distributions

are assumed, the computational procedures for obtaining correlation coefficients are the same as the initial computations required for obtaining the coefficients of the regression equation.

4. APPLICATION TO TIME SERIES

Consider now the situation where one has collected $N+k$ observations from a stationary random process as a function of time which has a zero mean value. Suppose that these are equally spaced observations and are denoted by $x_1, x_2, \ldots, x_{N+k}$ where x_1 is the first observed value and x_{N+k} is the last. These observations could be obtained by reading N values from an analog record or they might naturally arise at discrete points as in certain digital processes. In order to perform an autocorrelation analysis on the data, compute the sample autocorrelation function at k points as defined by the formula

$$R_{xx}(i) = \frac{1}{N} \sum_{j=1}^{N} x_j x_{j+i} \quad , \qquad i = 0, 1, \ldots, k \tag{14}$$

It will be assumed that k is much less than N, that is, there are many more observations available than points at which the autocorrelation function is computed.

A more convenient quantity to work with is the normalized autocorrelation function. This is defined by dividing $R_{xx}(i)$ by $R_{xx}(0)$. In equation form one has

$$\Gamma_{xx}(i) = \frac{R_{xx}(i)}{R_{xx}(0)} \quad , \qquad i = 0, 1, \ldots, k \tag{15}$$

These quantities, $\Gamma_{xx}(i)$, are called correlation coefficients and it may be easily proved that they are bounded in absolute value by unity. That is

$$-1 \leq \Gamma_{xx}(i) \leq 1 \quad , \qquad i = 0, 1, \ldots, k \tag{16}$$

After the correlation function has been computed, a regression analysis may follow. The object of the regression analysis would be to obtain a linear equation which would be used to predict ahead (extrapolate) in the time series. This is a reasonable objective since if there are high correlations indicated by very high peaks in the correlation function at certain time delays, then this implies some prediction can be made this far ahead in the time series. For example, if there is a peak in the correlation function at $i = 10$ seconds, then as the series is being observed and data is being collected in real time, say, one would expect to be able to predict ahead approximately 10 seconds with much greater accuracy than at other times.

Now it is desired to calculate coefficients b_i, $i = 1, 2, \ldots, k$ for a regression equation of the form of Eq. (11), namely,

$$\hat{x}_0 = b_1 x_1 + b_2 x_2 + \ldots + b_k x_k \tag{17}$$

In this case, the variables $x_1, x_2, \ldots, x_k$ all represent observations from the given time series as opposed to being different variables. To be specific, x_1 = the present observed value of the time series, x_2 = the value observed one time unit in the past, and so on up to x_k which represents the observation $(k - 1)$ time units in the past. The variable x_0 is displaced one time unit from x_1 and therefore represents a value of the time series one time unit into the future. This is, of course, the prediction to be made.

To be more precise in notation, Eq. (17) should be written with subscripts as follows.

$$\hat{x}_{t+1} = b_1 x_t + b_2 x_{t-1} + \ldots + b_k x_{t-k+1} \tag{17a}$$

In words, all variables are translated with respect to t. However, avoiding the more complicated subscripts simplifies notation and should cause no confusion.

4.1 COMPUTATIONAL DETAILS

The coefficients, b_i, will be found from solving a set of simultaneous linear equations similar to those indicated by Eq. (13) using the values of the autocorrelation function which have already been computed. The set of simultaneous linear equations may be re-written in a slightly different form as follows.

$$
\begin{aligned}
b_1R(0) + b_2R(1) + \ldots + b_kR(k-1) &= R(1) \\
b_1R(1) + b_2R(0) + \ldots + b_kR(k-2) &= R(2) \\
\vdots \qquad\qquad\qquad\qquad & \quad \vdots \\
b_1R(k-1) + b_2R(k-2) + \ldots + b_kR(0) &= R(k)
\end{aligned}
\tag{18}
$$

In the above equation $R(i) \equiv R_{xx}(i)$ to simplify notation.

The solution of this set of k linear equations requires essentially the inversion of a k by k matrix. If k is large, say on the order of 30 or 40, which would not be at all unreasonable, the necessary matrix inversion would be a considerable computational task, even on a digital computer. Therefore, it would seem to be advisable to restrict one's attention to only those points in the correlation function which exhibit a fairly high peak as determined by some method. By confining one's attention to only those points in the correlation function which are significantly different from zero, one can reduce the order of the matrix to be inverted considerably. It would be most desirable if all non-significant points in the correlation function could be eliminated from consideration entirely. Unfortunately, this is not the case as will be illustrated in the derivation of the necessary least squares equations below.

As an example, assume that significant peaks in the correlation function occur at points 3, 4, 8, and 11. The variables which appear in the regression equation are therefore x_3, x_4, x_8, and x_{11}. The coefficients

to be estimated from the data are denoted by b_3, b_4, b_8, and b_{11} in Eq. (18) below.

$$\hat{x}_0 = b_3x_3 + b_4x_4 + b_8x_8 + b_{11}x_{11} \tag{19}$$

The sum of squares to be minimized is

$$\sum_{i=1}^{N}(x_0 - \hat{x}_0)^2 = \sum_{i=1}^{N}(x_0 - b_3x_3 - b_4x_4 - b_8x_8 - b_{11}x_{11})^2 \tag{20}$$

Differentiating with respect to the b_i coefficient, and equating to zero, one obtains the following set of simultaneous linear equations.

$$\begin{aligned}
b_3\sum x_3^2 &+ b_4\sum x_3x_4 + b_8\sum x_3x_8 + b_{11}\sum x_3x_{11} &= \sum x_3x_0 \\
b_3\sum x_3x_4 &+ b_4\sum x_4^2 + b_8\sum x_4x_8 + b_{11}\sum x_4x_{11} &= \sum x_4x_0 \\
b_3\sum x_3x_8 &+ b_4\sum x_4x_8 + b_8\sum x_8^2 + b_{11}\sum x_8x_{11} &= \sum x_8x_0 \\
b_3\sum x_3x_{11} &+ b_4\sum x_4x_{11} + b_8\sum x_8x_{11} + b_{11}\sum x_{11}^2 &= \sum x_{11}x_0
\end{aligned} \tag{21}$$

In the above equations all summations are assumed to run from i = 1 to N. Note that it is assumed that N + 11 observations are available so that all points of the correlation function are based on N observations.

Rewriting the above equations in terms of the correlation function values, one obtains the following set of simultaneous linear equations

$$\begin{aligned}
b_3R(0) + b_4R(1) + b_8R(5) + b_{11}R(8) &= R(3) \\
b_3R(1) + b_4R(0) + b_8R(4) + b_{11}R(7) &= R(4) \\
b_3R(5) + b_4R(4) + b_8R(0) + b_{11}R(3) &= R(8) \\
b_3R(8) + b_4R(7) + b_8R(3) + b_{11}R(0) &= R(11)
\end{aligned} \tag{22}$$

In the above equations only R(3), R(4), R(8), and R(11) are considered to be significant points on the correlation function. However, in deriving the least squares equation, the points R(0), which is of course the variance, and R(1), R(5), and R(7) also enter into the equations. However, the necessary matrix to be inverted is now only of order 4, as opposed to order k if all k points of the correlation function were employed for the prediction equation.

The general set of equations, when one chooses some subset of the points of the correlation function, is as follows. Suppose one decides upon r separate points of the correlation function as being significant peaks. Suppose further these are labeled $a_1, a_2, \ldots, a_r$. The points of interest in the time series are, therefore, $x_{a_1}, x_{a_2}, \ldots, x_{a_r}$. The set of simultaneous linear equations to be solved for the b coefficients now becomes

$$\begin{aligned}
b_{a_1} R(0) &+ b_{a_2} R(a_2-a_1) + \ldots + b_{a_r} R(a_r-a_1) = R(a_1) \\
b_{a_1} R(a_2-a_1) &+ b_{a_2} R(0) + \ldots + b_{a_r} R(a_r-a_2) = R(a_2) \\
&\vdots \\
b_{a_1} R(a_r-a_1) &+ b_{a_2} R(a_r-a_2) + \ldots + b_{ar} R(0) = R(a_r)
\end{aligned} \tag{23}$$

No particular systematic method will be considered in this report for choosing the values of the correlation function which are significant and which, therefore, should be included in the regression equation. In many cases of interest, a significant peak or peaks will be obvious in the computed correlation function. In many other typical cases the correlation function will exhibit a damped oscillatory behavior as in the widely observed exponential-cosine autocorrelation function. In this case one should use the first several peaks (positive and negative) as variables for the regression equation.

4.2 COMPARISON WITH OTHER PROCEDURES

One may compare the preceding results to that of the "optimum Wiener linear predictor." The approach here as a regression problem gives the same results since both procedures are based on a least squares error criterion. The optimum Wiener linear predictor is essentially given by Eq. (17a) except that it is usually presented in continuous integral form:

$$x(t) = \int_0^T x(t - \tau)\, b(\tau)\, d\tau \tag{24}$$

where the coefficients b_i or "weighting function" $b(\tau)$ are obtained from solving Eq. (18) or the equivalent.

Also, classical data smoothing procedures extended to an extrapolation procedure would lead to the same results if an underlying linear trend is assumed. In this case there is a difference in concept involved since one never considers "noise" extrapolation or prediction when thinking of smoothing data. One invariably assumes observations are composed of an underlying signal, e.g., an nth degree polynomial, and additive independent noise. One then performs a "curve fitting" procedure with a least squares error criterion to obtain an equation of the form of Eq. (17a) in the discrete case or Eq. (24) in the continuous case.

5. EXTENSION TO CROSS-CORRELATED VARIABLES

The preceding regression techniques may be extended to predict one variable as a function both of past observations of itself and as a function of past observations of another variable. The procedure for obtaining the necessary coefficients of the regression equation will be directly analogous to the preceding method, although now the cross-correlation between the first variable x and the second variable y will be brought into play. The necessary equations will be obtained below for the case where peaks in the autocorrelation function and the cross-correlation function are taken into account to reduce the amount of computation.

Suppose as before that the points of significance for the autocorrelation function are at $a_1, a_2, \ldots, a_r$ for the variable x. For the cross-correlation of x with y, assume the significant points of this function are designated by $d_1, d_2, \ldots, d_s$. The points of the autocorrelation function of the variable x will be denoted by $R_x(a_1), R_x(a_2), \ldots, R_x(a_r)$. The values of cross-correlation function between the variables x and y will be denoted by $R_{xy}(d_1), R_{xy}(d_2), \ldots, R_{xy}(d_s)$. The least squares equations are obtained in exactly the same way as was done previously. After minimizing the appropriate sum of the squares, the simultaneous linear equations are as shown in Eqs.(25). In these equations (25) the relation $R_{xy}(i) = R_{yx}(-i)$ has been employed. From this set of equations the coefficients of the following regression equation are obtained:

$$\hat{x}_0 = b_{a_1} x(a_1) + \ldots + b_{a_r} x(a_r) + c_{d_1} y(d_1) + \ldots + c_{d_s} y(d_s) \qquad (26)$$

To illustrate these equations more concretely, assume that peaks occur at $R_x(1)$ and $R_x(3)$ in the autocorrelation function for x and at $R_{xy}(2)$ and $R_{xy}(4)$ in the cross-correlation function between x and y.

$$
\begin{array}{lllllll}
b_{a_1} R_x(0) & + b_{a_2} R_x(a_2 - a_1) & + \ldots + b_{a_r} R_x(a_r - a_1) & + c_{d_1} R_{xy}(d_1 - a_1) & + \ldots + c_{d_s} R_{xy}(d_s - a_1) & = R_x(a_2) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
b_{a_1} R_r(a_r - a_1) & + b_{a_2} R_x(a_r - a_2) & + \ldots + b_{a_r} R_x(0) & + c_{d_1} R_{xy}(d_1 - a_r) & + \ldots + c_{d_s} R_{xy}(d_s - a_r) & = R_x(a_r) \\
b_{a_1} R_{xy}(d_1 - a_1) & + b_{a_2} R_{xy}(d_1 - a_2) & + \ldots + b_{a_r} R_{xy}(d_1 - a_r) & + c_{d_1} R_y(0) & + \ldots + c_{d_s} R_y(d_s - d_1) & = R_{xy}(d_1) \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
b_{a_1} R_{xy}(d_s - a_1) & + b_{a_2} R_{xy}(d_s - a_2) & + \ldots + b_{a_r} R_{xy}(d_s - a_r) & + c_{d_1} R_y(d_1 - d_s) & + \ldots + c_{d_s} R_y(0) & = R_{xy}(d_s)
\end{array}
\tag{25}
$$

For this special case four simultaneous linear equations are obtained, namely,

$$
\begin{aligned}
b_1R_x(0) &+ b_3R_x(2) + c_2R_{xy}(1) + c_4R_{xy}(3) = R_x(1) \\
b_1R_x(2) &+ b_3R_x(0) + c_2R_{xy}(-1) + c_4R_{xy}(1) = R_x(3) \\
b_1R_{xy}(1) &+ b_3R_{xy}(-1) + c_2R_y(0) + c_4R_y(2) = R_{xy}(2) \\
b_1R_{xy}(3) &+ b_3R_{xy}(1) + c_2R_y(2) + c_4R_y(0) = R_{xy}(4)
\end{aligned}
\tag{27}
$$

From these equations the coefficients for the following prediction are obtained:

$$\hat{x}_0 = b_1x_1 + b_3x_3 + c_2y_2 + c_4y_4 \tag{28}$$

Note that values of the autocorrelation function for y are required and that also $R_{xy}(-1)$ is required or equivalently $R_{yx}(1)$.

It is clear that the above equations can easily be extended to account for cross-correlations between more than two variables. However, the case of two variables shown above adequately illustrates the general form of the equations to the general case will not be developed here.

6. SEAKEEPING APPLICATIONS

Applications of the autoregressive techniques described on the preceding pages are performed most directly for relatively low frequency and quasi-periodic data. Four specific seakeeping applications will be discussed here. The first three concern oceanographic data analysis, and the fourth describes an application to the area of vibration and acoustical data analysis. This application to vibration and acoustics is not involved with direct analysis of an individual vibration record, but from a broader standpoint of predicting vibration properties from other data.

6.1 LONG RANGE OCEAN ACTIVITY PREDICTION

Many applications exist in this general area which can be associated with the weather. When viewed as a time series, the weather is somewhat periodic and of a relatively low frequency nature. In Reference [6, page 129], Pierson suggests the possibility of applying the techniques described in this report for long range weather forecasting. Similarly, since ocean wave activity is influenced by the weather, these techniques could be valuable in obtaining a long or short range forecaster of gross ocean wave activity.

For example, the wind has considerable influence in the generation of a confused sea while air pressure and temperature might exhibit some more indirect effects. Also, ocean activity in one geographical area might lead to activity at a later time in a different geographical area.

To illustrate this matter, assume one has collected time histories $x(t)$ and $y(t)$ of some parameter of ocean wave activity at two different geographical points, and a time history of the wind, $z(t)$, at one of these points. The normalized autocorrelation function of $x(t)$ and the normalized cross-correlation functions between $x(t)$, $y(t)$, and $x(t)$, $z(t)$ might appear as pictured in Figure 1.

If the correlation functions were as shown in Figure 1, then the autocorrelation function of $x(t)$ itself would only be of use for very short term predictions. However, since peaks occur in the two cross-correlation

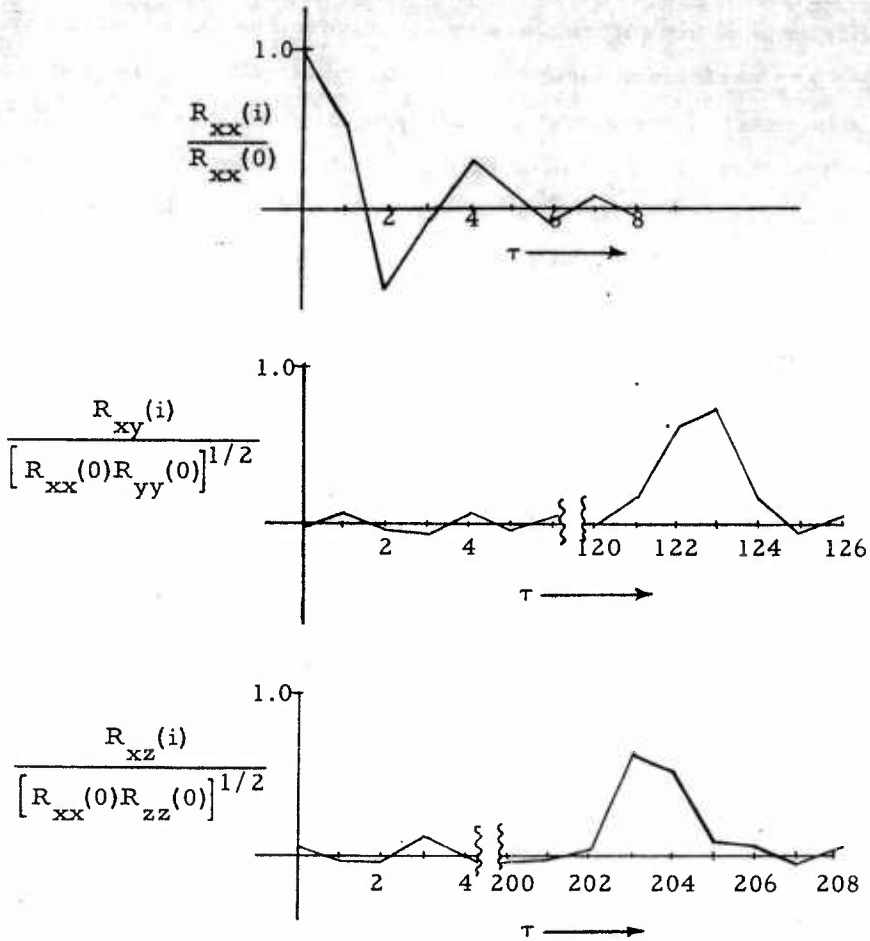

Figure 1

Hypothetical Correlation Functions
for Sea Activity Prediction Example

functions at longer time delays, predictions could be made on the basis of this information for relatively longer prediction times. If the time interval for τ was one minute, then Figure 1 indicates that the prediction could be made for a two-hour period in advance. The regression equation would be

$$\hat{x}_0 = c_{122}y_{122} + c_{123}y_{123} + d_{203}z_{203} + d_{204}z_{204} \tag{29}$$

The linear equations in matrix form to be solved to obtain the coefficients in Eq. (29) are

$$\begin{bmatrix} R_{yy}(0) & R_{yy}(1) & R_{yz}(81) & R_{yz}(82) \\ R_{yy}(1) & R_{yy}(0) & R_{yz}(80) & R_{yz}(81) \\ R_{yz}(81) & R_{yz}(80) & R_{zz}(0) & R_{zz}(1) \\ R_{yz}(82) & R_{yz}(81) & R_{zz}(1) & R_{zz}(0) \end{bmatrix} \begin{bmatrix} c_{122} \\ c_{123} \\ d_{203} \\ d_{204} \end{bmatrix} = \begin{bmatrix} R_{xy}(122) \\ R_{xy}(123) \\ R_{xz}(203) \\ R_{xz}(204) \end{bmatrix} \tag{30}$$

After this set of equations has been solved, then Eq. (29) could be used as a predictor of ocean wave activity.

6.2 SHORT TERM ROLL PREDICTION

The possibility of obtaining short term predictions utilizing information contained in the autocorrelation function of the time series consisting of certain ship motions is suggested by St. Denis and Pierson in Reference [7, page 35]. The practical use of such a prediction would undoubtedly be in a short term control system of some sort in a ship. Assume for example that roll, denoted by $x(t)$, is the motion being considered. As indicated, Ref. [7], these various ship motions look like narrow band noise under an exciting force of a random sea due to the fact that the ship acts as a narrow band filter. The typical autocorrelation function that arises from narrow band noise is a damped exponential-cosine. Therefore, experimental data might give rise to an autocorrelation function as depicted in Figure 2.

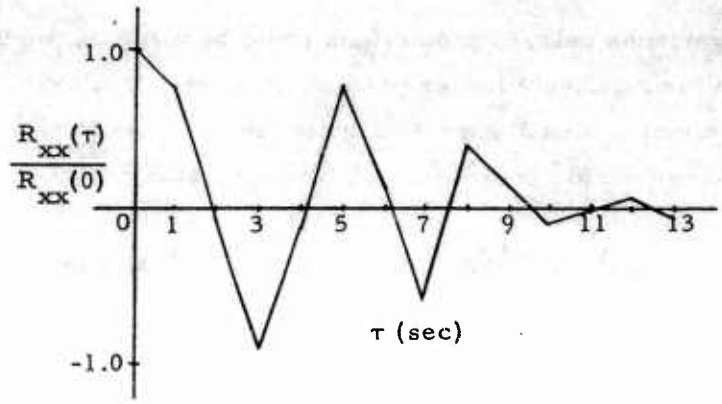

Figure 2

Hypothetical Autocorrelation Function for Ship's Roll

If τ is in seconds, then Figure 2 indicates that a one-second prediction could be made by use of the regression equation

$$\hat{x}_0 = b_1 x_1 + b_3 x_3 + b_5 x_5 + b_7 x_7 + b_8 x_8 \tag{31}$$

In order to obtain the coefficients in Eq. (31), the following set of five linear equations must be solved.

$$\begin{bmatrix} R_{xx}(0) & R_{xx}(2) & R_{xx}(4) & R_{xx}(6) & R_{xx}(7) \\ R_{xx}(2) & R_{xx}(0) & R_{xx}(2) & R_{xx}(4) & R_{xx}(5) \\ R_{xx}(4) & R_{xx}(2) & R_{xx}(0) & R_{xx}(2) & R_{xx}(3) \\ R_{xx}(6) & R_{xx}(4) & R_{xx}(2) & R_{xx}(0) & R_{xx}(1) \\ R_{xx}(7) & R_{xx}(5) & R_{xx}(3) & R_{xx}(1) & R_{xx}(0) \end{bmatrix} \begin{bmatrix} b_1 \\ b_3 \\ b_5 \\ b_7 \\ b_8 \end{bmatrix} = \begin{bmatrix} R_{xx}(1) \\ R_{xx}(3) \\ R_{xx}(5) \\ R_{xx}(7) \\ R_{xx}(8) \end{bmatrix} \tag{32}$$

If a longer range slightly less accurate prediction is desired, x_1 could be discarded or perhaps both x_1 and x_3.

The control procedure could possibly be implemented in one manner by utilizing a modern day shipboard high speed digital computer such as the

AN/UYK-1 (TRW-130). The system would require some sort of device to sample the roll time series $x(t)$, at least once per second and probably more like ten times per second. This information would have to be analog to digital converted and then fed into the computer. The regression equation, such as Eq. (31), would then be evaluated with a predicted value $\hat{x}_0$ obtained. This value could then be processed appropriately, output perhaps to a digital to analog converter, and then used by some anti-roll device. All these procedures would be accomplished in "real time." Perhaps every few seconds a new correlation function estimate could be developed and its values tested for significance by some simple procedure to determine whether or not to use a point in the prediction equation.

A matrix of values such as given in Eq. (32) would have to be inverted in order to obtain the coefficients for the prediction equation which would then replace the previously used prediction equation. This matrix would, of necessity, be restricted to some maximum size so that the inversion could be performed in a reasonable amount of time. Evaluating the regression Eq. (31) would prove no problem timewise since, for example, the add time of the AN/UYK-1 is 12 microseconds (μs) for a 15-bit word and its multiply time is 57 μs maximum. Therefore, the necessary instructions to evaluate Eq. (31) would require about 500 μs, allowing for the necessary load and store operations.

The remainder of the problem, that is, evaluating the correlation function, inverting the correlation matrix, and input/output functions could be performed on a piecemeal basis at a slower rate. Military computers such as the AN/UYK-1 usually have interrupt capabilities such that when an input or output device is ready, it can send a signal to interrupt the computer processing. Therefore, if it was time for the anti-roll device to receive information, it could interrupt, say, the matrix inversion routine. The computer could perform the necessary processing to output information, and then return to the matrix inversion. This type of processing would

possibly allow for an output rate of five or ten control signals per second. A block diagram for this type of system is illustrated in Figure 3.

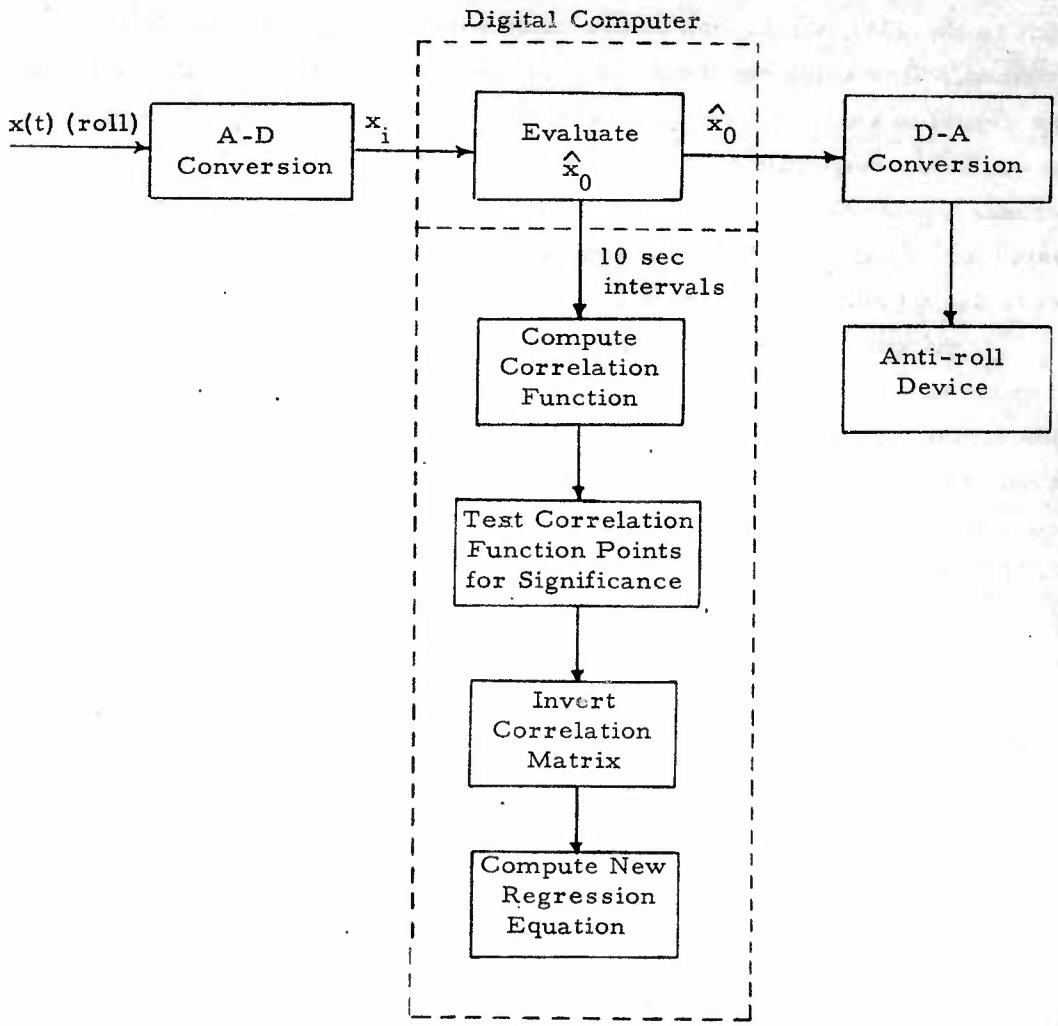

Figure 3

Possible Digital Control System for Ship's Roll

In order to efficiently perform all the required processing, it would probably be necessary to evaluate certain quantities, such as the correlation function in a recursive manner. For example, with the first observation, x_i, one computes x_i^2. With the second observation one computes x_{i+1}^2 and $x_i x_{i+1}$ and accumulates $x_i^2 + x_{i+1}^2$. With the third observation x_{i+2} one computes x_{i+2}^2, $x_{i+2}x_i$, $x_{i+2}x_{i+1}$, and accumulates $x_i^2 + x_{i+1}^2 + x_{i+2}^2$ and $x_i x_{i+1} + x_{i+1}x_{i+2}$. This procedure would then continue until sufficient observations had been obtained to allow reliable correlation function estimates. Perhaps more efficient approximate procedures could be developed.

The prediction procedure might even be improved by including other information, for example, by cross-correlating directly with a record of ocean wave amplitudes. Another possibly would be including a cross-correlation with pitch or heave information. Although, in practice, the six degrees-of-freedom of ship's motion are usually assumed to be independent, in actuality, the motions may be correlated and use of this information might allow for better predictions of motion.

6.3 OCEAN WAVE AMPLITUDE PREDICTION

As a third example for oceanographic problems, suppose that a time series $x(t)$ consists of observations of the height of ocean waves at a given point on the surface of the sea. Assume that observations are taken at one-second intervals. The first step in the analysis would be to compute the points of the autocorrelation function at, say, ten-second intervals as defined by Eq. (14). The correlation function could be expected to fall off fairly rapidly and then a peak should be encountered corresponding to the predominant frequency of the wave process. Assume that this occurs at a delay of 30 seconds. It is not unreasonable that another underlying periodicity might occur of much greater period. This fact would be exhibited by a peak in the correlation function at a greater time delay, say,

for example, at 150 seconds. The over-all normalized correlation function for this data might then appear something like that pictured in Figure 4.

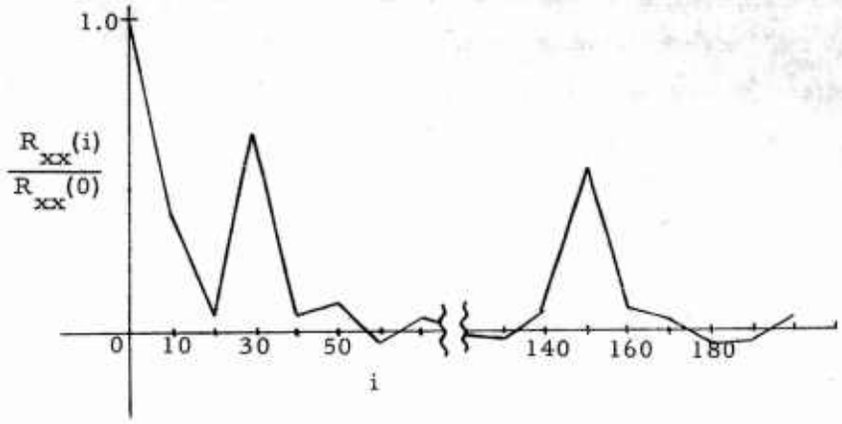

Figure 4

Hypothetical Autocorrelation Function for Ocean Wave Data

For this example one would choose as variables in the regression equation, x_{10}, x_{30}, and finally x_{150}. These variables correspond to the points R(10), R(30), and R(150) of the correlation function. One then wants to estimate the coefficients in the following linear equation.

$$\hat{x}_0 = b_{10}x_{10} + b_{30}x_{30} + b_{150}x_{150} \tag{33}$$

The set of simultaneous linear equations to be solved would be as follows.

$$\begin{aligned} b_{10}R(0) &+ b_{30}R(20) + b_{150}R(140) = R(10) \\ b_{10}R(20) &+ b_{30}R(0) + b_{150}R(120) = R(30) \\ b_{10}R(140) &+ b_{30}R(120) + b_{150}R(0) = R(150) \end{aligned} \tag{34}$$

Once the coefficients b_{10}, b_{30}, and b_{150} have been computed, then Eq. (33) may be used to predict on the basis of presently observed and past observed data for a period of ten seconds into the future. A prediction of ten seconds into the future may or may not be of practical value as far as ocean waves are concerned. However, in this example, due to the strong correlations exhibited at both 30 seconds and 150 seconds, it might be desirable and of interest to compute regression equations based on x_{30} and x_{150}, or possibly even just on x_{150} alone. In the first case, one could then be predicting ahead 30 seconds into the future and in the second case one could be predicting ahead 150 seconds into the future. These predictions might be of more practical value. However, one loses precision in the prediction when pertinent data is neglected such as exists at x_{10} which is indicated by the strong correlation at the point $i = 10$. The over-all correlation-regression analysis described in this example has the value of pointing out the fact that in addition to the basic "periodicity" with a period of 30 seconds, there is an additional underlying "periodicity" with a period of approximately 150 seconds. This information may or may not have been obvious from the original data.

6.4 VIBRATION DATA APPLICATION

For a vibration data application, the emphasis of the procedure will be shifted. A problem that is of interest is to obtain a vibration data time series as a function of other time series. For example, pressure transducers might be mounted at various external points on a ship's structure and an accelerometer might be located at an internal point of interest on the structure to measure the vibration response at that point. The pressure transducers at various points on the structure would effectively measure sources of vibration excitation. These exciting forces would transmit directly through the structure, or acoustically through the air to produce vibration at the structural point where the accelerometer is located. There

would be time delays between the excitation and response due to the finite amount of time that it takes to transmit the vibration through the structure or the surrounding medium. Therefore, the vibration response at the accelerometer might be obtained as a function of the response measured at the various pressure transducers at some time in the past. The vibration response would then be given as a function of lagged values of the pressure variables. In this example only the cross-correlation analysis would be of interest.

For purpose of this example, let $x(t)$ represent readings taken from an accelerometer located at a point on a ship structure, and let $y(t)$ and $z(t)$ represent the readings of two pressure transducers located at other points on the structure. The normalized cross-correlation functions of $x(t)$ with $y(t)$ and of $x(t)$ with $z(t)$ might then appear as illustrated in Figure 5.

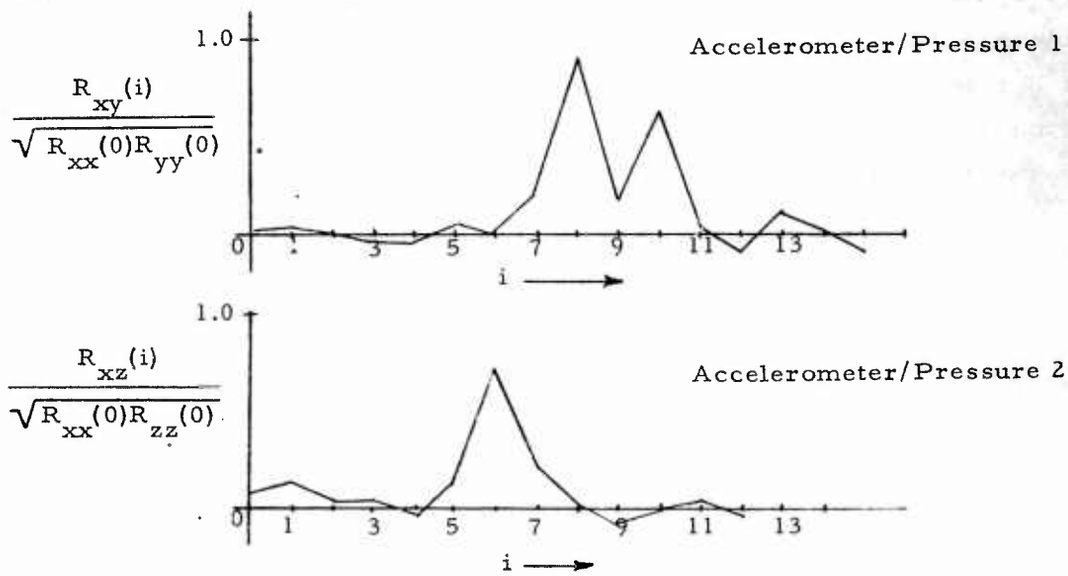

Figure 5

Hypothetical Cross-correlation Functions
for Vibration Response Example

Variables to choose in predicting vibration as indicated by the above accelerometer data, then, would be pressure transducer No. 1 readings at lags of 8 and 10 time units, and pressure transducer No. 2 at a lag of 6 time units. Hence, the variables would be y_8, y_{10}, and z_6. The regression equation to give accelerometer readings as a function of the two pressure transducer readings is then

$$\hat{x}_0 = c_8 y_8 + c_{10} y_{10} + d_6 z_6 \tag{35}$$

The coefficients of this equation may be obtained from the following set of linear equations

$$\begin{bmatrix} R_{yy}(0) & R_{yy}(2) & R_{yz}(-2) \\ R_{yy}(2) & R_{yy}(0) & R_{yz}(-4) \\ R_{yz}(-2) & R_{yz}(-4) & R_{zz}(0) \end{bmatrix} \begin{bmatrix} c_8 \\ c_{10} \\ d_6 \end{bmatrix} = \begin{bmatrix} R_{xy}(8) \\ R_{xy}(10) \\ R_{xz}(6) \end{bmatrix} \tag{36}$$

In this problem one needs the autocorrelation function for $y(t)$ and the autocorrelation function for $z(t)$, the cross-correlation function between $x(t)$ and $y(t)$, and the cross-correlation function between $y(t)$ and $z(t)$ at negative values of the lag, or equivalently values of the cross-correlation function between $z(t)$ and $y(t)$ at positive values of the lag. For this problem the end objective is not an extrapolation of the vibration time series, but rather to predict the vibration as a linear function of the pressure transducer readings.

7. STATISTICAL TESTS AND ERRORS

After the coefficients in the regression equation have been obtained, it is desirable to apply a statistical test to see if the coefficients are significantly different from zero. This is equivalent to testing if the variable associated with the given coefficient contributes a statistically significant amount to the prediction of the time series. For classical regression analysis, the variables are assumed to be normally distributed and the deviations from the predicted values used in the regression equations are assumed to be normally distributed and independent from one prediction to the next. However, in the application of regression techniques to time series, the problem is more difficult. Even when the process may be assumed to be a Gaussian or normal process, it will still have a non-zero autocorrelation function. Hence, the residuals from the prediction will not necessarily be independent from prediction to prediction, but will themselves be correlated. Fortunately, for large sample sizes, it is pointed out, Ref. [2], that the classical formulas hold approximately true. This means that classical formulas for the standard errors and the sampling distributions of the regression coefficients are asymptotically valid even if the residuals are correlated.

An approximate test for significance on the sample regression coefficients may be performed in the following way. See Reference [3] for more details. Let Γ represent the $(k+1)$ by $(k+1)$ matrix of correlation coefficients.

$$\Gamma = \begin{bmatrix} 1 & \Gamma(1) & \dots & \Gamma(k) \\ & 1 & \dots & \Gamma(k-1) \\ & & & \vdots \\ & & & 1 \end{bmatrix} \tag{37}$$

Then, if C_{ij} is the cofactor of i-jth element in the matrix Γ, the sample regression coefficient b_i may be computed from

$$b_i = \frac{-C_{1i}}{C_{11}} \tag{38}$$

The sample variance of residual errors in the regression, i.e., the variance of the distribution of the deviations of the predicted values from the true values, is obtained from the formula

$$s^2_{0.12\ldots k} = R(0)\,\frac{|\Gamma|}{C_{11}} \tag{39}$$

where $|\Gamma|$ is the determinant of the matrix Γ. The standard errors of the regression coefficients are given by the formula

$$s_{b_{i-1}} = \frac{1}{C_{11}}\sqrt{\frac{C_{11}C_{ii} + C^2_{1i}}{N}} \tag{40}$$

This is derived from manipulation of formulas

$$s^2_{b_i} = \frac{1}{N}\,\frac{s^2_{0.23\ldots k}}{s^2_{i.34\ldots k}}\,(1 - \rho^2_{0i.23\ldots k}) \tag{41}$$

and

$$\rho^2_{0i.34\ldots k} = \frac{C^2_{1,\,i+1}}{C_{11}C_{i+1,\,i+1}} \tag{42}$$

which may be found in Chapter 27 of Reference [5].

The meaning of the above terms is as follows. The term $s^2_{0.12\ldots k}$ in Eq. (39) is the sample variance of x_0 when the best linear estimates of $x_1, x_2, \ldots, x_k$ have been subtracted out. This is the reason for the form of the notation. The term $\rho^2_{0i.23\ldots k}$ in Eq. (42) is the <u>partial correlation</u> coefficient between x_0 and x_i. The partial correlation coefficient is the correlation coefficient between the variables x_0 and x_i

after the best linear estimate of the other variables involved have been subtracted from x_0 and x_i.

Since the t distribution approaches the normal distribution for a large number of degrees-of-freedom, the normal distribution may be substituted for t in the above tests if (N - k) is larger than, say, 30. This should be the case in most practical situations. Chapter 3, Reference [4], gives correction factors for the standard errors of the regression coefficients but have the disadvantage of being quite complicated and requiring a large amount of additional computations.

It may then be shown, Ref. [3], that the statistic

$$t = (b_i - \beta_i) \frac{(N - k)^{1/2}}{s_{b_i}} \tag{43}$$

has a t-distribution with (N - k) degrees-of-freedom (d.f.). Thus, one computes t from Eq. (43), and if

$$|t| \geq t_{1-\alpha/2}(N - k) \tag{44}$$

where $t_{1-\alpha/2}(N - k)$ is obtained from tables of the student t distribution, then the hypothesis $b_i = \beta_i$ is rejected at the α level of significance. Note this is a "two-tailed" test. The most usual test would be for $b_i = 0$ and Eq. (43) would be modified accordingly by setting $\beta_i = 0$.

8. DIGITAL COMPUTING CONSIDERATIONS

Some basic computing procedures for regression analysis are indicated in Section 5.6 of Reference [1]. An outline will now be given here of suggested computations to implement the preceding discussions which encompass those techniques discussed in Reference [1]. No distinction need usually be made between the auto or cross-correlation cases since the computational procedures are basically the same.

A convenient way of handling the set of $(N+k)$ observations of x_i is to arrange them in a $(k+1)$ by N matrix as indicated below.

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \\ x_2 & x_3 & \cdots & x_{N+1} \\ \vdots & & & \\ x_{k+1} & x_{k+2} & \cdots & x_{N+k} \end{bmatrix} \tag{45}$$

The product of this matrix X with its transpose X' gives a $(k+1)$ by $(k+1)$ matrix whose elements are the discrete points of the sample correlation function. Thus

$$R_x = XX' = \begin{bmatrix} \sum x_i^2 & \sum x_i x_{i+1} & \cdots & \sum x_i x_{i+k} \\ \sum x_i x_{i+1} & \sum x_i^2 & \cdots & \sum x_i x_{i+k-1} \\ \vdots & & & \vdots \\ \sum x_i x_{i+k} & \cdots & & \sum x_i^2 \end{bmatrix} \tag{46}$$

The top row of the matrix then gives the $k+1$ discrete points of the autocorrelation function $R(0), R(1), \ldots, R(k)$. In actual computational practice

for the autocorrelation case, one only needs to compute the first row of the matrix (46) and then shift this row to the right to obtain the additional portions of the $(k-1)$ rows necessary to fill in the upper right half of this symmetric matrix. However, in some cases where computing time is not important, one might perform the complete matrix multiplication to simplify programming.

The values of the correlation function $R_{xx}(i)$ would then be inspected to determine what values of x_i to utilize in the regression equation. The rows and columns corresponding to the unwanted points would then be deleted from the general matrix equation.

For example, to obtain Eqs. (32) of Section 6.2, one eliminates as indicated below.

$$\begin{bmatrix} R(0) & R(1) & R(2) & R(3) & R(4) & R(5) & \cdots \\ \cancel{R(1)} & \cancel{R(0)} & \cancel{R(1)} & \cancel{R(2)} & \cancel{R(3)} & \cancel{R(4)} & \cancel{\cdots} \\ R(2) & R(1) & R(0) & R(1) & R(2) & R(3) & \cdots \\ \cancel{R(3)} & \cancel{R(2)} & \cancel{R(1)} & \cancel{R(0)} & \cancel{R(1)} & \cancel{R(2)} & \cancel{\cdots} \\ R(4) & R(3) & R(2) & R(1) & R(0) & R(1) & \cdots \\ \cancel{R(5)} & \cancel{R(4)} & \cancel{R(3)} & \cancel{R(2)} & \cancel{R(1)} & \cancel{R(0)} & \cancel{\cdots} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix} \begin{bmatrix} b_1 \\ \cancel{b_2} \\ b_3 \\ \cancel{b_4} \\ b_5 \\ \cancel{b_6} \\ \vdots \end{bmatrix} = \begin{bmatrix} R(1) \\ \cancel{R(2)} \\ R(3) \\ \cancel{R(4)} \\ R(5) \\ \cancel{R(6)} \\ \vdots \end{bmatrix} \qquad (47)$$

The set of linear equations remaining after the appropriate deletions then would correspond to Eqs. (32) in Section 2. The reduced system of equations indicated by Eq. (47) may now be solved directly for the b_i coefficients or in some cases it is desirable to obtain the explicit form of the inverse of the matrix of correlation function values. To avoid the necessity of introducing new symbols, let R_x represent the matrix of correlation function values whether or not deletions have been made.

For digital computing purposes, it is usually desirable to work with the matrix of correlation coefficients defined by

$$\Gamma(i) = \frac{R(i)}{R(0)} = \frac{R(i)}{\sigma^2} \tag{48}$$

for the autocorrelation case and by

$$\Gamma_{xy}(i) = \frac{R_{xy}(i)}{\sqrt{R_{xx}(0)\,R_{yy}(0)}} = \frac{R_{xy}(i)}{\sigma_x \sigma_y} \tag{49}$$

for the cross-correlation case. This normalization requires divisions to be performed in the autocorrelation case and square roots and divisions in the cross-correlation case. However, the final quantities are then in the range -1 to +1 which provides advantages in the general handling of the numbers in the matrix inversion process. Let Γ_x represent the matrix of correlation coefficients and let A represent the inverse of Γ_x, that is $A = \Gamma_x^{-1}$.

The system of equations to be solved directly for the coefficients is then

$$\begin{bmatrix} 1 & \Gamma(1) & \cdots & \Gamma(k-1) \\ \Gamma(1) & 1 & \cdots & \Gamma(k-2) \\ \vdots & & & \\ \Gamma(k-1) & \Gamma(k-2) & \cdots & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} = \begin{bmatrix} \Gamma(1) \\ \Gamma(2) \\ \vdots \\ \Gamma(k) \end{bmatrix} \tag{50}$$

Note that the matrix on the left is a k by k matrix corresponding to that of Eq. (37) with the first row and column deleted. If the computations are being performed in a real time control system and computational speed is important, one would solve for the b_i directly without obtaining the inverse

matrix explicitly. However, in other situations, it is desirable to obtain the inverse of the $(k+1)$ by $(k+1)$ matrix explicitly since certain quantities of statistical interest may be conveniently obtained from the elements of A.

For the solution of the equations, a reasonable elimination and back substitution method which takes advantage of symmetry is termed the "Banachiewiez-Cholesky-Crout Method." A description of this method, Ref. [8], is presented here. For the computational procedure below, let d_{ij} represent the elements of the k by k matrix of Eq. (50). The computational steps are as follows:

(a) Define

$$\gamma_{i1} = d_{i1} \quad , \quad i = 1, 2, \ldots, k$$

$$\alpha_{1j} = d_{1j} \quad , \quad j = 2, 3, \ldots, k$$

(b) Compute

$$\gamma_{ij} = d_{ij} - \sum_{n=1}^{j-1} \gamma_{in}\alpha_{nj} \quad , \quad i, j = 2, 3, \ldots, k \quad i \geq j \tag{51}$$

(c) Compute

$$\alpha_{ij} = \frac{\gamma_{ji}}{\gamma_{ii}} \quad , \quad i < j \tag{52}$$

(d) Compute

$$\omega_1 = \Gamma(1)$$

$$\omega_i = \frac{1}{\gamma_{ii}} \left[\Gamma(i) - \sum_{n=1}^{i-1} \gamma_{in}\omega_n \right], \quad i = 2, 3, \ldots, k \tag{53}$$

(e) Finally, the coefficients b_i are given by

$$b_k = \omega_k$$

$$b_i = \omega_i - \sum_{n=i+1}^{k} \alpha_{in} b_n \quad , \quad i = k-1, k-2, \ldots, 1 \tag{54}$$

If the elements a_{ij} of the inverse matrix A are desired, the same basic procedure given above applies with minor modifications, except now one starts with the full (k + 1) by (k + 1) matrix of correlation coefficients. In Step (d) above, the $\Gamma(i)$ are replaced with unit row vectors ϵ_i where ϵ_i represents a unit vector with a one in the ith position. For example, $\epsilon_1 = (1, 0, \ldots, 0)$ replaces $\Gamma(1)$, etc. Steps (d) and (e) are modified as follows:

(d') Compute

$$\omega_1 = \epsilon_1$$

$$\omega_i = \frac{1}{\gamma_{ii}} \left[\epsilon_i - \sum_{n=1}^{i-1} \gamma_{in} \omega_n \right], \qquad i = 2, 3, \ldots, k+1 \tag{55}$$

Note that the ω_i are now row vectors instead of scalars.

(e') Compute the rows a_i of the inverse matrix A

$$a_{k+1} = \omega_{k+1}$$

$$a_i = \omega_i - \sum_{n=i+1}^{k} \alpha_{in} \omega_n, \qquad i = k, k-1, \ldots, 1 \tag{56}$$

(f) The regression coefficients are given by

$$b_i = \left(\frac{a_{1i}}{a_{11}} \right) \left(\frac{s_1}{s_i} \right) \quad i = 1, 2, \ldots, k \tag{57}$$

In Eq. (57), s_i is the sample standard deviation of the ith variable under consideration. In the strict autocorrelation case $s_1 = s_2 = \ldots = s_k$, but in the cross-correlation case they are in general different.

(g) The sample variance of the residual errors defined in Section 7 by Eq. (39) is given by

$$s^2_{0.12\ldots k} = \frac{R(0)}{a_{11}} \tag{58}$$

(h) The standard errors of the regression coefficients are computed from

$$s_{b_{i-1}} = \frac{1}{a_{11}} \left(\frac{a_{11}a_{ii} + a_{ij}}{N} \right)^{1/2}, \quad j = 1, 2, \ldots, k \tag{59}$$

(i) The t statistic to test the regression coefficients for significance may then be computed from

$$t = (b_i - \beta_i) \frac{(N-k)^{1/2}}{s_{b_i}} \tag{60}$$

where the β_i may be arbitrary hypothesized values, but usually are chosen as zero.

This concludes the computational procedure except, of course, for the evaluation of the regression equation which is a straightforward calculation.

REFERENCES

1. Bendat, J.S., Enochson, L.D., Klein, G.H., and A.G. Piersol. The Application of Statistics to the Flight Vehicle Vibration Problem, ASD TR 61-123, Aeronautical Systems Division, AFSC, USAF, Wright-Patterson AFB, Ohio. December 1961. ASTIA AD 271 913.

2. Bartlett, M.S., An Introduction to Stochastic Processes, Cambridge Press, London, (1955).

3. Cramer, Harald, Mathematical Methods of Statistics, Princeton Univ. Press, Princeton, New Jersey (1946).

4. Lyttkens, E., "Standard Errors of Regression Coefficients in the Case of Autocorrelated Residuals," Chap. 3, Time Series Analysis (ed. Murray Rosenblatt), John Wiley and Sons, Inc., New York (1963).

5. Kendall, Maurice G., and Alan Stuart, The Advanced Theory of Statistics, Hafner Publishing Co., New York (1961). Vol. 2.

6. Pierson, W.J., Jr, "On the Use of Time Series Concepts and Spectral and Cross-spectral Analyses in the Study of Long-range Forecasting Problems," Sears Foundation: Journal of Marine Research (October 30, 1960). Vol. 18, No. 2, pp. 112-132.

7. St. Denis, Manley and Willard J. Pierson, Jr., "On the Motions of Ships in Confused Seas," Paper presented at annual meeting of the Society of Naval Architects and Marine Engineers in New York, November 1953.

8. Korn, G.A., and T.M. Korn, Mathematical Handbook for Scientists and Engineers, McGraw-Hill Book Co., Inc., New York (1961).

DISTRIBUTION LIST

Commanding Officer and Director 65
David Taylor Model Basin
Washington, D.C. 20007 Code 513

Chief of Naval Research 4
Department of the Navy
Washington 25, D.C. Code 438

Commanding Officer 1
Office of Naval Research
Branch Office
230 North Michigan Avenue
Chicago 1, Illinois

Commanding Officer 1
Office of Naval Research
Branch Office
207 West 24th Street
New York 13, New York

Commanding Officer 1
Office of Naval Research
Branch Office
1030 East Green Street
Pasadena 1, California

Commanding Officer 1
Office of Naval Research
Branch Office
495 Summer Street
Boston 10, Massachusetts

Commanding Officer 1
Office of Naval Research
Branch Office
1000 Geary Street
San Francisco 9, California

Commanding Officer 5
Office of Naval Research
Navy No. 100, Box 39
Fleet Post Office
New York, New York

Director 2
U.S. Naval Research Laboratory
Washington 25, D.C. (Code 2021)

Chief, Bureau of Aeronautics 3
Department of the Navy
Washington 25, D.C.
Attn: Research Division,
Code AD-3 and Code DE-3

Chief, Bureau of Ordnance 1
Department of the Navy
Washington 25, D.C.
Attn: Research and Development Div.

Chief, Bureau of Ships 9
Department of the Navy
Washington 25, D.C.
Attn: Tech Library (Code 210L)
Tech Information (Code 335)
Ship Design (Code 410)
Preliminary Design (Code 420)
Scientific Research (Code 442)
Landing Ships (Code 529)
Minesweeping (Code 631)
Ship Silencing (Code 345)
Ship Research Br. (Code 341B)

Chief, Bureau of Yards and Docks 1
Department of the Navy
Washington 25, D.C.
Attn: Research Division

Hydrographer 1
Department of the Navy
Washington 25, D.C.

Commander 1
Portsmouth Naval Shipyard
Portsmouth, New Hampshire

Commander 1
Naval Ordnance Test Station
3202 East Foothill Blvd.
Pasadena, California
Attn: Head, Underwater Ord. Dept

Commanding Officer 1
U.S. Naval Base
Key West, Florida
Attn: Surface Anti-Submarine
Development Detachment

Commanding Officer and Director 1
U.S. Naval Engineering Experiment Station (Wave Mechanics)
Annapolis, Maryland

Librarian 1
U.S. Naval Postgraduate School
Monterey, California

Dr. J.H. McMillen 1
National Science Foundation
1520 H Street N.W.
Washington, D.C.

Director 1
National Bureau of Standards
Washington 25, D.C.
Attn: Fluid Mechanics Section
Dr. G. B. Schubauer

Director 1
Langley Research Center
NASA
Langley Field, Virginia

Mr. J. B. Parkinson 1
Langley Aeronautical Laboratory
NASA
Langley Field, Virginia

Office of Technical Services 1
Department of Commerce
Washington 25, D.C.

Division, Applied Mathematics 1
Brown University
Providence 12, Rhode Island

California Institute of Technology 2
Hydrodynamics Laboratory
Pasadena 4, California

University of California 4
Berkeley 4, California
Attn: Prof. H.A. Einstein (1)
Dept. of Engineering
Prof. H.A. Schade (3)
Dept. of Naval Architecture

Cornell University 1
Graduate School, Aeronautical Engr.
Ithaca, New York
Attn: Prof. W.R. Sears, Director

University of Illinois 1
Dept., Theoretical and Applied Mechanics
College of Engineering
Urbana, Illinois
Attn: Dr. J. M. Robertson

State University of Iowa 2
Iowa Institute of Hydraulic Research
Iowa City, Iowa
Attn: Dr. Hunter Rouse, Director
Dr. L. Landweber

University of Maryland 1
Institute for Fluid Dynamics and Applied Mathematics
College Park, Maryland
Attn: Prof. H.M. Martin, Director

Massachusetts Inst. of Technology 3
Cambridge 39, Massachusetts
Attn: Prof. A.T. Ippen, Hydrodynamics Lab
Prof. M.A. Abkowitz, and Prof. L. Troost, Dept., Naval Architecture

Michigan State College 1
Hydraulics Laboratory
East Lansing, Michigan
Attn: Prof. H. R. Henry

University of Michigan 1
Applied Mechanics Department
Ann Arbor, Michigan
Attn: Prof. J.S. McKnown

University of Minnesota 1
St. Anthony Falls Hydraulic Lab
Minneapolis 14, Minnesota
Attention: Director

New York University 2
Inst., Mathematical Sciences
25 Waverly Place
New York 3, New York
Attn: Prof. R. Courant, Director
Prof. J. J. Stoker

Director of Research 1
NASA
1512 H Street N.W.
Washington 25, D.C.

New York University 1
Dept. Meteorology and Oceanography
College of Engineering
New York 53, New York
Attn: Dr. W.J. Pierson, Jr.

University of Notre Dame 1
College of Engineering
Notre Dame, Indiana
Attn: Dr. A.B. Strandhagen

Pennsylvania State University 1
Ordnance Research Laboratory
University Park, Pennsylvania
Attn: Prof. G.F. Wislicenus
Director, Water Tunnel

Stanford University 1
Applied Mathematics and
Statistics Laboratory
Stanford, California

Stevens Institute of Technology 1
Davidson Laboratory
711 Hudson Street
Hoboken, New Jersey
Attn: Dr. J.P. Breslin

University of Tennessee 1
Engineering Experiment Station
Knoxville, Tennessee
Attn: Dr. G.H. Hickox, Director

Woods Hole Oceanographic Inst. 1
Woods Hole, Massachusetts
Attn: Dr. C. O'D. Iselin

Director 1
Scripps Institute of Oceanography
La Jolla, California
Attn: Adm. O.D. Wheelock, USN(Ret)

Mathematical Reviews 1
80 Waterman Street
Providence, Rhode Island

Colorado State University 1
Dept. of Civil Engineering
Fort Collins, Colorado
Attn: Dr. A.R. Chamberlain

Harvard University 1
Dept. of Mathematics
Cambridge 39, Massachusetts
Attn: Dr. G. Birkhoff

Commander 1
Air Research and Development Command
Andrews Air Force Base
Washington 25, D.C.
Attn: Fluid Mechanics Division

Commander 10
Defense Documentation Center
Cameron Station
Alexandria, Virginia 22314

Technical Research Group, Inc. 1
Route 110
Melville, New York 11749
Attn: Dr. Jack Kotik

Oceanics, Incorporated 1
Technical Industrial Park
Plainview, Long Island, New York
Attn: Dr. Paul Kaplan, President

Engineering Experiment Station 1
Kansas State University
Manhattan, Kansas 66504

Tuskegee Institute 1
School of Engineering
Tuskegee, Alabama

Itek Corporation, Vidya Division, 1
1450 Page Mill Road
Palo Alto, California

Illinois Institute of Technology 1
Aeronautical Laboratory
Chicago 16, Illinois
Attn: Prof. Irving Michelson

UNCLASSIFIED

UNCLASSIFIED