ON THE USE OF TIME SERIES CONCEPTS AND SPECTRAL AND CROSS-SPECTRAL ANALYSES IN THE STUDY OF LONG-RANGE FORECASTING PROBLEMS

BY

WILLARD J. PIERSON, JR.

New York University

ABSTRACT

Monthly values of sea surface temperature anomalies and wind anomalies for an area of the North Atlantic Ocean are studied as a vector Gaussian process. The covariances and the spectra and cross-spectra provide information on the nature of the interaction of ocean and atmosphere. The long period fluctuations are found to be fairly coherent. Indications are that the sea surface temperature can be predicted for one month into the future on the basis of past values with a 50 % reduction of variance. Prediction problems are discussed.

1. *Introduction.* The concepts of stationary time series methods, of power spectral analysis of time series, and of techniques for prediction of time series are playing an increasingly important role in numerous fields. One field of interest to the writer is that of waves and ship motions (Pierson, 1955, 1957; Putz, 1955).

In meteorological studies, spectra have been studied in connection with numerous problems (see, for example, Cramer and Record, 1955; Griffith, *et al.*, 1956), and regression methods have been used in attempting to derive prediction schemes (see Lorenz, 1956; Malone, 1956; Miller, 1956; and Wadsworth, 1951). It is the purpose of this paper to suggest briefly, by means of studying an interesting stationary Gaussian vector process, a way in which spectral and cross-spectral analyses and regression schemes might be combined to develop prediction procedures and to obtain some fairly accurate long-range prediction methods. The study of spectra and cross-spectra provides information on how the prediction works and on how to improve the planning of further investigations.

The theory used in this paper is that of stationary random vector processes, already well described in the literature. Useful references are Bartlett (1949), Lévy (1948), Grenander and Rosenblatt (1957), Press and Tukey (1956), Blackman and Tukey (1958), and Goodman (1957).

Time series methods require lengthy sample records for analysis. Thus the obvious instability of attempts to extrapolate four or eight year apparent cycles far into the future on the basis of only about 60 years of data is emphasized by such methods.

2. *The Problem for Analysis.* The system considered here consists of three series of data, two of which consist of the mean monthly sea surface temperature anomalies for individual months from 1880 to 1940 for area E and area N of the North Atlantic as published by Smed (1946, 1949). The location of these fields is given in Fig. 1. The third series is the mean monthly pressure gradient anomaly in the form of a Δp measured over a fixed horizontal distance over areas E and N at sea level as defined by Neumann and Pandolfo in Neumann, *et al.* (1958). The average pressure gradient for the month was computed from the gradient of the average pressure for that month. The wind anomaly near the surface over these two ocean areas is of course directly proportional to the pressure gradient anomaly.

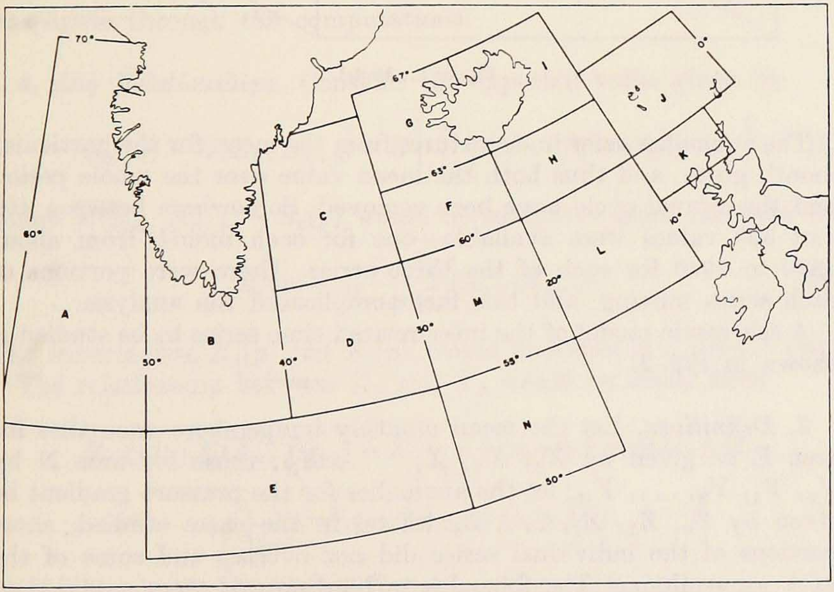

Figure 1. Location of fields for which temperature data are given by Smed.

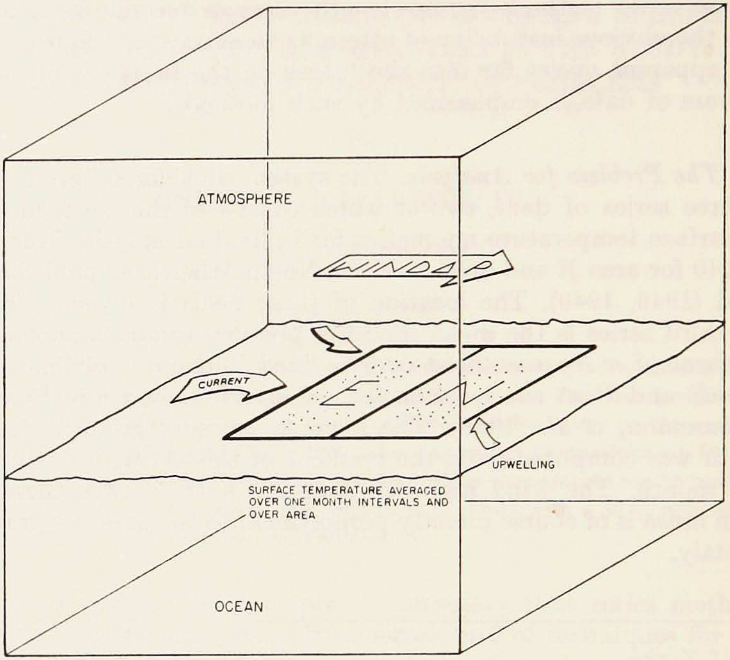

Figure 2. Model.

The anomalies refer to departures from the mean for the particular month given, and thus both the mean value over the whole period and the annual cycle have been removed. Somewhere between 400 and 500 values were available—one for each month from about 1880 to 1940 for each of the three series. There were portions of each series missing, and this fact complicated the analysis.

A schematic model of the inter-related time series to be studied is shown in Fig. 2.

3. *Definitions*. Let the mean monthly temperature anomalies for area E be given by X_0, X_1, X_2, . . ., X_N, those for area N by Y_0, Y_1, Y_2, . . ., Y_M; let the anomalies for the pressure gradient be given by Z_0, Z_1, Z_2, . . ., Z_S where, in the cases studied, short portions of the individual series did not overlap and some of the data were missing. The formulas to be discussed were modified to take missing data into account.

A mathematical model for the vector time series defined by (X_n, Y_n, Z_n) can be formulated on the assumption, which appears to be justifiable from various tests of the data, that the data are samples from a stationary vector Gaussian process.

$$\left.\begin{aligned} X_n &= \sum_{j=0}^{\infty} a_j \cos(\mu_j n \Delta t) + b_j \sin(\mu_j n \Delta t) \\ Y_n &= \sum_{j=0}^{\infty} c_j \cos(\mu_j n \Delta t) + d_j \sin(\mu_j n \Delta t) \\ Z_n &= \sum_{j=0}^{\infty} e_j \cos(\mu_j n \Delta t) + f_j \sin(\mu_j n \Delta t). \end{aligned}\right\} \tag{1}$$

In (1), $\Delta t = 1$ month and μ_j has the units of radians/month.

Eqs. (1) are to be considered as a very large sum of terms (in the limit infinite) with very small individual amplitudes (in the limit infinitesimal). The numbers, a_j, b_j, c_j, d_j, e_j, and f_j, are considered to be numbers to be drawn at random from a multivariate normal distribution with zero means and with variances and covariances that are a function of μ_j and that as yet have not been determined. The notation has been chosen to show that X_n, Y_n and Z_n are known only at discrete points and to show how the quantities, a_j, b_j, etc. propagate through the computations.

4. *Lag Relationships.* Consider the expected value given by

$$R_x(p) = E(X_n, X_{n+p}) = E\left[\sum_{j=0}^{\infty} \tfrac{1}{2}(a_j^2 + b_j^2) \cos(\mu_j p \Delta t)\right], \tag{2}$$

where $E(X_n, X_{n+p}) = R_x(p)$ would be estimated from the data by

$$R_x(p) = \frac{1}{N-p} \sum_{n=0}^{N-p} (X_n)(X_{n+p}). \tag{3}$$

The covariances, $R_y(p)$ and $R_z(p)$, would be found in a similar way.

The relationship between X_n and Y_n would be found from

$$\begin{aligned} R_{xy}(p) = E(X_p, Y_{n+p}) = E\Bigg[&\tfrac{1}{2}\sum_{j=0}^{\infty} (a_j c_j + b_j d_j) \cos(\mu_j p \Delta t) \\ &+ \tfrac{1}{2}\sum_{j=0}^{\infty} (-b_j c_j + a_j d_j) \sin(\mu_j p \Delta t)\Bigg]. \end{aligned} \tag{4}$$

Similar results for $R_{xz}(p)$ and $R_{yz}(p)$ would come from interchanging appropriate symbols.

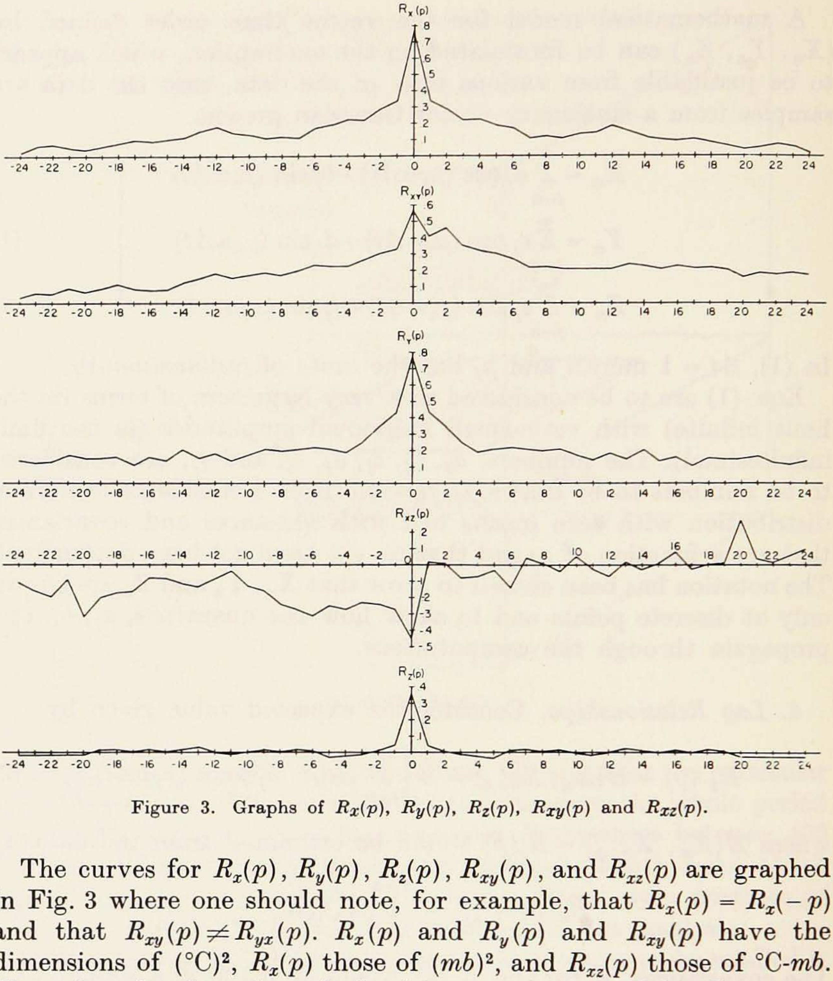

Figure 3. Graphs of $R_x(p)$, $R_y(p)$, $R_z(p)$, $R_{xy}(p)$ and $R_{xz}(p)$.

The curves for $R_x(p)$, $R_y(p)$, $R_z(p)$, $R_{xy}(p)$, and $R_{xz}(p)$ are graphed in Fig. 3 where one should note, for example, that $R_x(p) = R_x(-p)$ and that $R_{xy}(p) \neq R_{yx}(p)$. $R_x(p)$ and $R_y(p)$ and $R_{xy}(p)$ have the dimensions of $(°\mathrm{C})^2$, $R_x(p)$ those of $(mb)^2$, and $R_{xz}(p)$ those of °C-*mb*. The values for these lag relationships are given in Table I.

5. *Discussion of Lag Relationships.* The graphs in Fig. 3 yield interesting information about the system under study as they stand. Some general statements about the predictability of a given component of the vector process from just its own past or from the past of some other component can be made.

For example, about $25\,{}^0/_0$ of the variance of X_n can be predicted one month into the future on the basis of only its current value.

TABLE I. VALUES OF $R_x(p)$, $R_y(p)$, $R_z(p)$, $R_{xy}(p)$ AND $R_{xz}(p)$.

p	R_x	$R_{xy}(+p)$	$R_{xy}(-p)$	R_y	$R_{xz}(+p)$	$R_{xz}(-p)$	R_z
0	0.775	0.552	0.552	0.778	−0.450	−0.450	0.369
1	.336	.412	.332	.434	.016	− .270	.044
2	.290	.460	.308	.348	.010	− .238	.011
3	.221	.356	.268	.249	− .082	− .230	− .012
4	.225	.322	.226	.193	− .072	− .274	− .020
5	.188	.310	.230	.210	− .030	− .248	− .026
6	.169	.276	.236	.219	− .140	− .308	.011
7	.106	.222	.202	.211	.044	− .274	.014
8	.121	.224	.178	.197	.010	− .238	.005
9	.109	.213	.191	.211	− .042	− .156	.012
10	.134	.210	.174	.171	.048	− .130	− .003
11	.142	.208	.158	.138	− .026	− .038	− .014
12	.179	.208	.186	.185	− .070	− .088	.004
13	.148	.240	.152	.153	.014	− .136	.028
14	.114	.227	.129	.205	− .026	− .116	.008
15	.097	.206	.086	.138	.008	− .084	− .001
16	.091	.227	.083	.158	.076	− .078	.014
17	.098	.217	.093	.108	− .028	− .160	.000
18	.077	.222	.114	.158	− .002	− .168	.014
19	.062	.212	.088	.148	.010	− .200	.012
20	.037	.166	.070	.152	.266	− .318	− .021
21	.050	.187	.093	.173	.040	− .092	− .020
22	.068	.178	.062	.131	.020	− .196	− .021
23	.058	.188	.066	.122	.060	− .160	− .022
24	.019	.181	.035	.124	.098	− .112	− .018

Only 4 % of the variance could perhaps be predicted 12 months into the future on the basis of only its current value. The graphs of both $R_x(p)$ and $R_y(p)$ show that X_n and Y_n depend to some extent on what has happened as much as two years ago and that the prediction of, say, the next monthly temperature anomaly value might use past observations extending back as much as two years.

In contrast, a value of Z_n is virtually independent of all past or future values. Perhaps 1 % of the variance of Z_n could be predicted one month into the future on the basis of its present value. Lack of correlation of the pressure anomaly with its past is probably explained by the fact that frequencies of the pressure gradient variability are associated with the passage of cyclones and anticyclones and hence are associated with much shorter periods—of the order of three to ten days. The averaging process essentially filters out a considerable part of these high frequencies.

The curve for $R_{xy}(p)$ is also interesting. The dependance of the temperature anomaly at area N, Y_n on the anomaly at area E, X_n one or two months ago is greater than the dependence of the temperature anomaly at E on the anomaly at N one or two months ago. Stated another way, prediction of Y_n one or two months into the future on the basis of the present value of X_n would be about twice as good as a prediction of X_n one or two months into the future on the basis of the present value of Y_n. Thus the ocean currents flowing from E to N appear to transport part of the anomaly at E to N with a time delay. The effect is evident even out to a two year lag, since the right side of the curve is consistently higher than the left side.

The curve for $R_{xz}(p)$ is perhaps the most striking of all. Although Z_n is unpredictable from its own past, the future values of X_n are predictable from present and past values of Z_n. Positive values of Z_n tend to be followed by negative values of X_n; negative values of Z_n, by positive values of X_n. (Or, as actually depicted by the graph, negative values of X_n tend to be preceded by positive values of Z_n; positive values of X_n, by negative values of Z_n.) Perhaps 9 $^0/_0$ of the variance of X_n would be predictable six months into the future by the present value of Z_n. The process is not reversible. Knowledge of past and present values of X_n provides no help in predicting Z_n.

The persistent covariances out to two years in all but the winds are subject to sampling variability, as are estimates of any parameters. Unfortunately sampling variability in covariance estimates from time series is not well understood (Blackman and Tukey, 1958; preface), and precise statements about the reliability of these estimates cannot be made on the basis of information given so far. From physical properties of the atmosphere and oceans, results are not contradictory to what one might expect.

The curve for $R_{yz}(p)$ would also be needed to complete an analysis of the problem, but it has not been computed.

6. *Spectra, Cross-spectra and Coherencies*. The lag relationships graphed above can be used to compute the spectra and cross-spectra of the vector Gaussian process under study by using the formulas from Tukey (1949) for spectra and from Goodman (1957) for cross-spectra. This provides important information on the frequencies involved and on the reality of results obtained. The curves defined

by $R_x(p)$, $R_y(p)$, . . ., $R_{xz}(p)$ are made periodic by repetition beyond $p = +\ 24$. The 24 coefficients of the Fourier series that define exactly the even functions and the 48 coefficients that define the even and odd parts of the other functions are then found. If $L_x(h)$, $h = 0$, 1, . . ., 24 are these numbers for the series X_n, then finally they are smoothed by computing $U_x(h) = 0.23\, L_x(h-1) + 0.54\, L_x(h) + 0.23\, L_x(h+1)$ to obtain an estimate of the spectrum of X_n.

For the spectra, $U_x(h)$ ($h = 0, 1, \ldots, 24$), is an estimate of that part of the total variance of X_n contributed by those components in (1) that have frequencies between approximately $2\pi(h-\frac{1}{2})/48$ and $2\pi(h+\frac{1}{2})/48$ radians/month (in the absence of aliasing). Or stated another way, $U_x(h)$ estimates approximately

$$\sum_{j=j_a}^{j=j_b} E(a^2+b^2), \tag{5}$$

where

$$\mu_{j_a} = 2\pi(h-\tfrac{1}{2})/48$$

$$\mu_{j_b} = 2\pi(h+\tfrac{1}{2})/48.$$ [1]

The cospectrum, say $C_{xy}(h)$, is the smoothed form of the cosine terms in the Fourier series representation of $R_{xy}(p)$. The cospectrum values are thus approximately an estimate of

$$\sum_{j=j_a}^{j=j_b} E(a_j c_j + b_j d_j), \tag{6}$$

where, since a_j and c_j and b_j and d_j may be uncorrelated, the sum could be zero and where j_a and j_b have the same significance as above.

The quadrature spectrum, say $Q_{xz}(j)$, is an estimate of

$$\sum_{j=j_a}^{j=j_b} E(-b_j e_j + a_j f_j) \tag{7}$$

under the same conditions.

[1] More precisely, the a_j^2 and b_j^2 are weighted by a triangularly shaped filter extending from about $2\pi(h-2)/48$ to $2\pi(h+2)/48$ instead of by a square filter between $2\pi(h-\frac{1}{2})/48$ and $2\pi(h+\frac{1}{2})/48$.

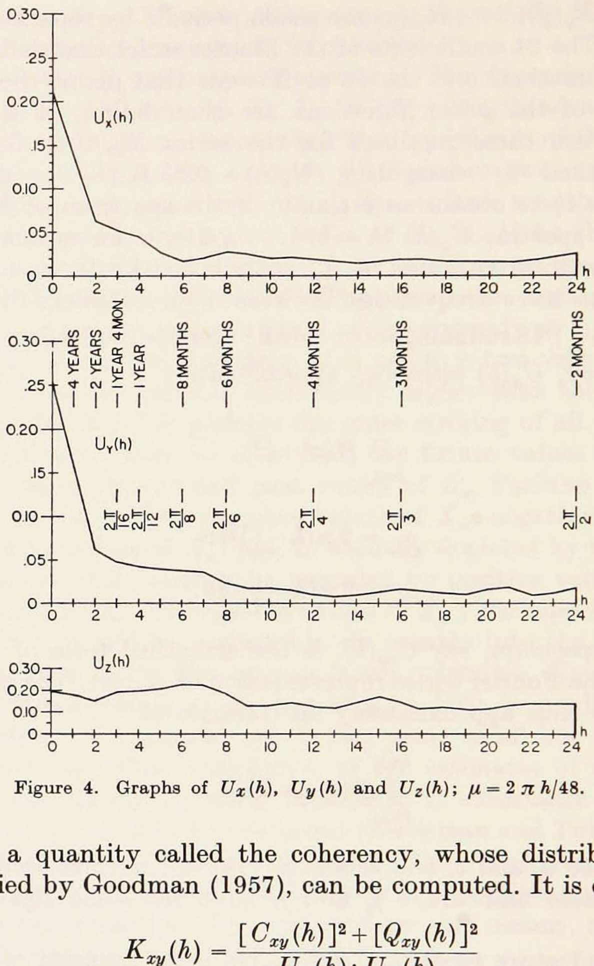

Figure 4. Graphs of $U_x(h)$, $U_y(h)$ and $U_z(h)$; $\mu = 2\pi h/48$.

Finally a quantity called the coherency, whose distribution has been studied by Goodman (1957), can be computed. It is defined by

$$K_{xy}(h) = \frac{[C_{xy}(h)]^2 + [Q_{xy}(h)]^2}{U_x(h) \cdot U_y(h)} \tag{8}$$

for the components X_n and Y_n, and by a similar equation for X_n and Z_n. The coherency is one provided the component of the vector process can be generated by the effect of a linear operator on the other component. In general, in geophysical problems it will be expected to be less than one.

The values of $U_x(h)$, $U_y(h)$, $U_z(h)$, $C_{xy}(h)$, $Q_{xy}(h)$, $K_{xy}(h)$, $C_{xz}(h)$, $Q_{xz}(h)$, and $K_{xz}(h)$ are graphed in Figs. 4 and 5 and tabulated in Table II. $C_{xy}(h) = C_{yx}(h)$ and $Q_{xy}(h) = -Q_{yx}(h)$, for example.

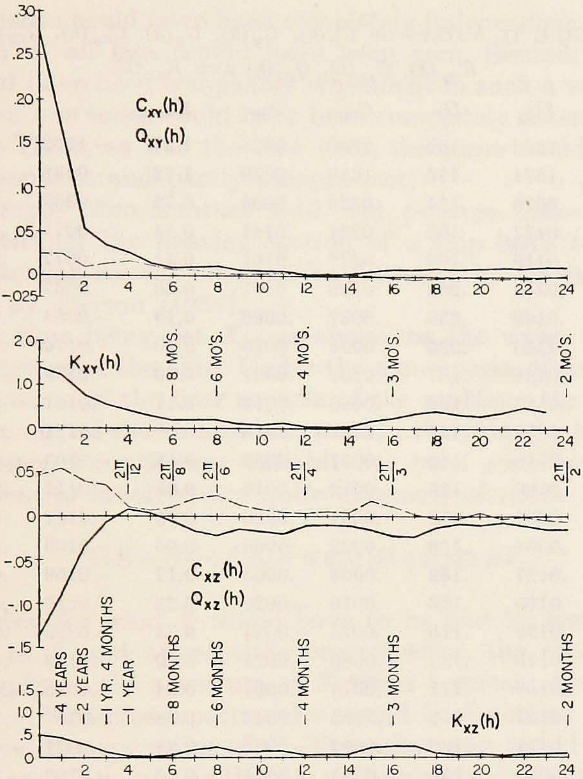

Figure 5. Graphs of $C_{xy}(h)$, $Q_{xy}(h)$, $K_{xy}(h)$, $C_{xz}(h)$, $Q_{xz}(h)$, and $K_{xz}(h)$; $\mu = 2\pi h/48$

7. *Interpretation of the Spectra.* The spectrum given by $U_x(h)$ can be interpreted as follows, except for the possibility of aliasing which will be discussed in § 9. The contribution to the total variance of the time series for area E made by frequencies between zero and $2\pi/96$ radians/month (half the distance from 0 to $2\pi/48$) is about 0.105 $(°C)^2$; since the frequency interval is half of the range for other h, the plotted value is halved. These frequencies correspond to periods ranging from eight years to infinity. The contribution from frequencies ranging from $2\pi/96$ to $2\pi/36$ is 0.139 $(°C)^2$. These frequencies correspond to periods of from eight to three years, and the point where the estimate is plotted corresponds to a period of four years.

The spectrum is quite flat for frequencies corresponding to periods

TABLE II. Values of $U_x(h)$, $U_y(h)$, $U_z(h)$, $C_{xy}(h)$, $Q_{xy}(h)$, $K_{xy}(h)$, $C_{xz}(h)$, $Q_{xz}(h)$ and $K_{xz}(h)$.

h	U_x	U_y	U_z	C_{xy}	Q_{xy}	K_{xy}	C_{xz}	Q_{xz}	K_{xz}
0	.2112	.2625	.189	.2746	.0228	1.37	–.1362	.0471	.520
1	.1391	.1574	.176	.1619	.0229	1.22	–.0893	.0544	.446
2	.0619	.0626	.154	.0535	.0088	0.76	–.0352	.0392	.291
3	.0517	.0472	.193	.0334	.0141	0.54	–.0071	.0358	.133
4	.0452	.0419	.196	.0227	.0161	0.41	.0077	.0109	.020
5	.0271	.0388	.209	.0139	.0101	0.28	.0037	.0061	.009
6	.0169	.0369	.230	.0087	.0066	0.19	–.0054	.0104	.035
7	.0202	.0355	.225	.0096	.0070	0.20	–.0170	.0202	.153
8	.0272	.0238	.187	.0109	.0037	0.20	–.0224	.0241	.212
9	.0250	.0167	.126	.0065	.0019	0.11	–.0151	.0127	.123
10	.0202	.0172	.128	.0052	.0019	0.09	–.0171	.0080	.137
11	.0219	.0148	.142	.0057	.0030	0.13	–.0203	.0048	.139
12	.0193	.0100	.125	.0012	.0019	0.03	–.0110	.0040	.056
13	.0158	.0082	.118	.0001	.0020	0.03	–.0101	.0054	.070
14	.0138	.0094	.139	.0025	.0008	0.05	–.0106	.0043	.067
15	.0158	.0127	.169	.0058	.0005	0.17	–.0159	.0175	.209
16	.0169	.0160	.146	.0076	–.0020	0.23	–.0216	.0111	.238
17	.0167	.0139	.116	.0076	–.0044	0.33	–.0215	.0011	.238
18	.0158	.0118	.122	.0085	–.0025	0.42	–.0120	–.0057	.091
19	.0163	.0109	.111	.0078	–.0004	0.34	–.0088	–.0028	.046
20	.0160	.0157	.113	.0083	–.0014	0.28	–.0107	.0031	.068
21	.0148	.0171	.130	.0094	–.0026	0.38	–.0077	–.0033	.036
22	.0170	.0090	.135	.0085	–.0021	0.50	–.0109	–.0006	.051
23	.0224	.0108	.136	.0100	–.0003	0.41	–.0156	–.0039	.084
24	.0244	.0164	.136	.0106	.0003	0.28	–.0157	–.0057	.083

of less than four months, and all frequencies above $2\pi/4$ contribute a total variance of about 0.21 $(°C)^2$. A similar interpretation could be given to the power spectrum for area N.

The spectrum for the pressure gradient anomaly is strikingly different from the other two spectra. It shows that contribution to the total variance is nearly the same for all frequencies from zero to $2\pi/2$ radians/month and suggests that aliasing may be a problem. Thus frequencies in the data corresponding to periods of from ten years (or greater) to two months are all of nearly equal importance.

8. *Interpretation of the Cross-spectra.* Several different things could have happened when $R_{xy}(p)$ and $R_{xz}(p)$ were estimated. First,

the three series could have been completely independent, and $R_{xy}(p)$ and $R_{xz}(p)$ at all lags would have been zero. Second, two of the series could have been completely dependent in such a way that the spectrum of one series could have been completely determined from the other. Third, as was the case here, the series could have been partly dependent and partly independent.

An example from another field will perhaps make this point clearer. Consider the heaving motion of a ship hove to in a long crested Gaussian seaway such as those studied by Lewis (1955) and described by Pierson (1957).

For the time being, let X_n, as above, be the waves and Z_n the heaving motion of the ship. Under the assumption of linearity, and neglecting some of the finer aspects of the problem, the response of the ship to one of the sinusoidal waves in the sum can be described by (9), where a continuous variable has been used for simplicity. The frequencies involved are quite different, of course:

$$A\ddot{Z} + B\dot{Z} + CZ = a \cos \mu t + b \sin \mu t. \tag{9}$$

On solving for what Z would have to be and by computing the spectrum of Z and appropriate cross-spectra, the spectrum of Z would simply be the spectrum of X times a certain function af A, B, C and μ. The cross-spectra also would be functions of A, B, C, and μ times the spectrum of X. The coherency would have to be identically equal to one, as there is a completely deterministic linear process connecting X and Z.

One of the important results from Goodman (1957) is that, if (9) is replaced by any linear operator on X_n, then the coherency would still be one for sufficiently great resolution in the absence of reading error in the two series.

The situation in the atmosphere-ocean system under study in this paper is much less well defined. The dynamics of the system connecting the two temperature fields is not well understood. It depends on the currents (both transient and quasi-steady) in that section of the ocean and on upwelling on the northern portion of the areas studied, as was shown by Neumann, *et al.* (1958). There is doubt as to whether or not these effects could even be described in part by a linear transformation on X_n to get Y_n, especially since the atmosphere-ocean complex appears to exhibit certain characteristics of a nonlinear rectifying filtering and feedback network.

Also, the terms a_j, b_j, c_j, d_j, e_j, and f_j in (1) should more probably be considered as being of the form

$$a_j = a_{D1j} + a_{D2j} + \ldots a_{R1j} + a_{R2j} + \ldots \text{ (and so on),}$$

where a_{D1j} is an actual effect acting on c_{D1j} and d_{D1j} due to one set of causes such as currents; a_{D2j} is another effect due to, say, upwelling; a_{R1j} is a random component due to, say, errors in the data, and so on. The problem of setting up a physical model for the processes under study is thus quite complex.

However, despite these difficulties, the cross-spectra and coherencies shown in Fig. 5 can be interpreted in a useful way. Some information on the time scale of the effects of anomalous temperature and winds over the area of interest can be deduced.

The graphs of $C_{xy}(h)$ and $Q_{xy}(h)$ show some of the features of the temperature anomalies at areas E and N. The curve for $C_{xy}(h)$ is quite similar in appearance to $U_x(h)$ and $U_y(h)$ except for the fact that it is near zero at $2\pi/4$. The quadrature spectrum is positive, and this shows that a maximum in the temperature anomaly in area E tends to be followed at a later time by a maximum in the temperature anomaly in area N. For example, the value of $C_{xy}(1)$ is about 0.16, the value of $Q_{xy}(1)$ is about 0.02 corresponding to a phase lag of $\tan^{-1}(1/8)$, or about 7°. Since the frequency involved corresponds to a period of four years, this means that a peak in an oscillation with about a four year period for area E tends to be followed about one month later by a peak in the corresponding oscillation at area N. The lag of Y_n behind X_n (N behind E) is between one month and two months for frequencies less than $2\pi/8$.

The graph of $K_{xy}(h)$ shows that the low frequency oscillations are closely related in the two areas. A value greater than one for $h = 0$ and $h = 1$ is possibly explained by sampling variation and by the fact that the missing data in the two series did not occur at the same times. For frequencies corresponding to periods between eight and three months, the coherency is less than 0.25, and this can be interpreted to mean that what occurs in one area for these frequencies has little to do with what occurs in the other area.

The graph of $C_{xz}(h)$ is a modified upside down version of $C_{xy}(h)$. The quadrature spectrum, $Q_{xz}(h)$, is relatively stronger, and at four years the cross-spectra correspond to a phase difference of four months when interpreted in the way that follows. For low frequencies the cospectrum implies that an above-average pressure gradient anomaly produces a below-average temperature anomaly, and conversely. The positive quadrature spectrum at frequencies of $2\pi/48$ to $2\pi/12$ shows that a minimum in the long-period fluctuation of the pressure gradient anomaly tends to be followed several months later (at 4 years, it is 4 months) by a maximum in the temperature anomaly.

The graph of $K_{xz}(h)$ shows that the temperature anomaly is not too strongly related to the pressure gradient anomaly. Only for long period fluctuations of three years or more does the coherency equal 1/2. For the higher frequency fluctuations it is quite small. This does not mean that the study of a more complex system will not lead to significant results inter-relating the wind field and the temperature anomalies at the higher frequencies.

In summary, the spectral analyses yield valuable information in the interpretation of the lag relationships discussed earlier in this paper. The part of the system containing predictable fluctuations is in the longer periods, *i.e.*, greater than about a year; the shorter period fluctuations are only weakly connected from one part of the system to the other. There is some connection between the temperature anomaly fluctuations in the areas at the shorter periods of two to three months.

9. *Aliasing.* Spectra computed from time series that are read at equally spaced points are subject to the phenomenon called aliasing, as explained by Blackman and Tukey (1958). A mean monthly value is essentially an average of the daily values of which it is composed, and effectively the day to day fluctuations have been smoothed by such an operation.

The anomalies could, in principle, have been found for any time by an average of the reports for the 30 days centered about the time of the computation. If, for example, values of the mean anomaly for the average of the last half of one month and the first half of the next month were found, the series could be made to consist of twice as many points. With 48 lags in the covariance estimates, the final spectra would then have twice the frequency range.

The contributions at these higher frequencies, if any, are, in the computations given, folded back on the zero to $2\pi/2$ frequency scale. Due to aliasing, for example, a spectral component with a true frequency of 2π radians/month could actually show up at the origin.

The process of averaging over the time domain effectively imposes a low pass filter in the frequency domain. However, since all of the spectra exhibit a contribution at $2\pi/2$, there is some aliasing in all of the spectra.

If the number of points in the series were doubled, as described above, and if the flat parts of the spectra *were* truly white noise due to a simple random error component at each point, the result on the doubled frequency scale would be to halve the given spectral estimates, report the same values over the new range, and thus preserve the total variance.

Other factors based on experience in meteorology and oceanography must enter to judge these effects in the absence of the ability to extend the frequency range. The aliasing in the spectrum of the pressure gradient anomaly is probably serious, although some of the variance of the pressure gradient anomaly has been assigned to the right frequency since the temperature anomalies have responded to it. This presence of aliasing combined with the effect of the low pass filter does not permit one to learn all that could be learned about the effect of the wind in the system under study.

Aliasing is probably not too serious in the temperature spectra. The spike is truly at low frequencies and not at some higher frequency that has been aliased, as would be evidenced by the fact that other quantities measured over the ocean exhibit true anomalous behaviour over entire months and years (for example, iceberg counts).

10. *Sampling Variability.* The three spectra and the various cross-spectra of the process are not precisely known. For example, on the basis of 400 points in the series and 24 lags, each spectral estimate has less than 40 degrees of freedom in a χ^2 distribution. This means that the values of $U_x(h)$, $U_y(h)$, and $U_z(h)$ can be multipled by 3/2 and 2/3 and that the true value of the spectrum will be between these bounds with a probability of 0.90. A similar lack of precision also exists for the cross-spectra. In the year 2000, the degrees of freedom will be doubled, but this is not of much help now.

Even with these broad confidence bands, it is evident that the contributions at low frequencies in $U_x(h)$ and $U_y(h)$ are significantly higher than the contributions at middle and higher frequencies. Within sampling variability, $U_z(h)$ might easily be the mean horizontal line through the indicated values. Similarly the cross-spectra and coherencies indicate a structure at low frequencies that cannot be explained on the basis of sampling variability.

It must therefore be concluded at a fairly high confidence level that some part of the structure of the spectra and cross-spectra is real. This in turn implies that the general shapes but not the details of the graphs of the covariance functions given in Fig. 3 are correct.

11. *Prediction.* The above results have shown that temperature anomalies in some sense propagate from E to N (X to Y) and that large long-period pressure gradient anomalies are followed by temperature anomalies of an opposite sign. To complete the study, it would be necessary to study $R_{yz}(p)$ and the appropriate spectra and cross-spectra, but on the basis of the above partial analysis, results suggest that the temperature anomalies at area N could be predicted from their own past and present values and from the past and present values of X_n and Z_n. Although the theory would work just as well in principle, the direction of the phase shifts and the coherencies involved suggest quite strongly that a prediction of Z_n from its own past and present and from the past and present of X_n and Y_n would not be very successful.

A complete study of the prediction problem has not been carried out. A one-month prediction of the temperature anomaly for area E was made on the basis of the least squares regression equation computed from the inverse of the determinant of $R_x(p)$ (see, for example, Cramer, 1946). The prediction is thus a simple weighted linear combination of 24 past values of the X_n series. As pointed out earlier, the one month lag would permit an explanation of 25 $^0/_0$ of the variance of the record. The full utilization of $R_x(p)$ permits an explanation of 50 $^0/_0$ of the variance of the fluctuations. (This suggests that certain of the regression schemes that use just one or two 24 hour lags to predict future temperatures and pressures might be improved by going back several weeks into the past.)

Fig. 6 shows a portion of the graph of the temperature anomalies, the predicted anomalies, and the residual error. The longer period fluctuations are quite successfully predicted. The short period fluctuations are only partially predicted. The error graph, quite distinctly different from the original anomaly graph, appears to emphasize the higher frequencies more than the original series.

The prediction has not been made on independent data. This un-

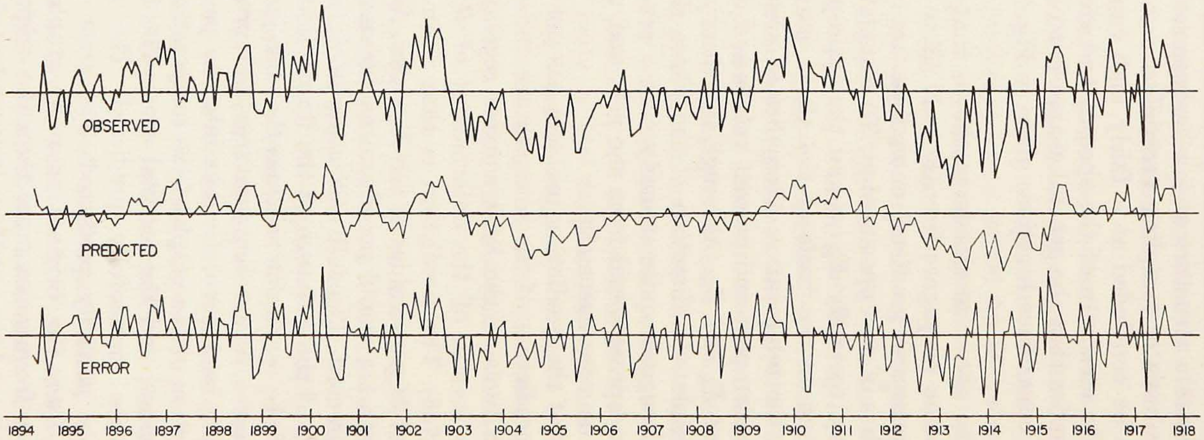

Figure 6. Selected portion of predictions for area E.

fortunate situation was forced on the writer because the data available to him consisted only of the period of years given at the start of this paper. If a big enough portion of it had been set aside as independent data, the sampling variability of the spectral and cross-spectral estimates would have suffered to the detriment of the study of their interesting features. The value of the 50 % reduction in variance might thus be questioned both on the basis of the lack of an independent check and also on the basis of what good it might be, even if real.

It is believed, however, that the interesting features of the graphs shown in Fig. 3, 4, and 5 are an important result of this study. Since the structure shown in these figures is real, results of the prediction are quite likely to be not entirely accidental.

12. *Paths for Further Investigation.* There are many ways in which the investigation described above could be extended. The prediction of the Y_n series could be carried out by a least squares regression and by inverting a 96×96 determinant of the values of $R_x(p)$, $R_y(p)$, $R_z(p)$, $R_{xy}(p)$, $R_{xz}(p)$, and $R_{yz}(p)$ so as to determine the coefficients in the following equation:

$$\begin{aligned} Y_{+1} &= a_0 Y_0 + a_{-1} Y_{-1} + \ldots + a_{-24} Y_{-24} \\ &+ b_0 X_0 + b_{-1} X_{-1} + \ldots + b_{-24} X_{-24} \\ &+ c_0 Z_0 + c_{-1} Z_{-1} + \ldots + c_{-24} Z_{-24} \end{aligned} \tag{10}$$

Or perhaps the spectra and cross-spectra given above could be smoothed, described analytically in a simpler way, reinverted, and used to find a simpler predictor.

A point to note is that the structure of the wind system has been completely suppressed to the level of white noise by using monthly anomaly values. (But the wind still affected the temperature.) However, the wind does have a definite spectral shape at shorter periods. If other areas of the ocean were included and if five day mean anomalies or weekly anomalies were studied instead of monthly values, the role of the atmosphere could be followed better because the effects of aliasing and filtering would be greatly reduced. This might result in improved monthly anomaly predictions, since the shorter period fluctuations of from two to four months are missed in the predictions shown.

13. *Conclusions.* The system studied in this report is an interesting one from both a meteorological and an oceanographic point of view. The analysis of its properties was far from complete in this study. Results show that predictability lies in the continuous spectrum of the long period components, that temperature anomalies downstream follow upstream fluctuations, and that the long period part of the wind fluctuation is followed by a temperature fluctuation but not conversely. The short period wind fluctuations that have been aliased in part and that have been filtered out in part by using monthly values may explain the two to four month variations in the temperature anomalies.

With the advent of larger, faster electronic computers, studies of more complex systems similar to this can be carried out more easily, and they should yield results of considerable importance. The results of this study can serve as a practical guide in planning such a program.

14. *Acknowledgments.* This work was sponsored by the Geophysics Research Directorate of the Air Force Cambridge Research Center, Air Research and Development Command. This paper is a revision under the same title of Chapter 3 of the final report for this project on Contract No. AF 19(604)–1284, prepared by G. Neumann, *et al.* (1958). The help of Raymond Stevens in discussing the problem and in carrying out the computations is gratefully acknowledged. The author is indebted to Professor Gerhard Neumann and Dr. Eberhard Wahl for their helpful suggestions and discussions.

REFERENCES

Bartlett, M. S.
1949. An introduction to stochastic processes. Cambridge University Press, England. 312 pp.

Blackman, R. B. and J. W. Tukey
1958. The measurement of power spectra from the point of view of communications engineering. Dover Publications, Inc., New York. 190 pp.

Cramer, Harald
1946. Mathematical methods of statistics. Princeton University Press. 575 pp.

Cramer, H. E. and F. A. Record
1955. Power spectra of the eddy-velocity components. J. Met., *12* (2): 146–151.

Goodman, N. R.
1957. On the joint estimation of the spectra, co-spectrum and quadrature spec-

trum of a two-dimensional stationary Gaussian process. Sci. Pap. Engng. Statist. Lab., N.Y. Univ. No. 10. 169 pp.

GRENANDER, ULF AND MURRAY ROSENBLATT

1957. Statistical analysis of stationary time series. John Wiley and Sons, Inc., New York. 300 pp.

GRIFFITH, H. L., H. A. PANOFSKY AND I. VAN DER HOVEN

1956. Power spectrum analysis over long ranges of frequency. J. Met., *13* (37): 279–282.

LÉVY, P.

1948. Processes stochastiques et mouvement brownien. Gauthier-Villars, Paris.

LEWIS, E. V.

1955. Irregular waves in model tanks, *in*, Ships and waves. Coun. on Wave Res. and Soc. of Naval Architects and Marine Engineers. J. W. Johnson, Ed. Pp. 324–331.

LORENZ, E. N.

1956. Empirical orthogonal functions and statistical weather prediction. Sci. Rep. Statist. Forecast. Proj., M.I.T., No. 1; 49 pp.

MALONE, T. F.

1956. Application of synoptic climatology to weather prediction, *in*, Petterssen's Weather Analysis and Forecasting. McGraw-Hill Book Co., New York; Chap. 28: 238–255.

MILLER, R. G.

1956. Statistics and the predictability of weather. Bull. Amer. meteor. Soc., *35*: 495–496 (Abstr.).

NEUMANN, GERHARD, EDWIN L. FISCHER, J. P. PANDOLFO AND W. J. PIERSON, JR.

1958. Studies on the interaction between ocean and atmosphere with application to long range weather forecasting. Final Rep. Engng. Res. Div., N.Y. Univ. Contract AF 19(604)–1284. 100 pp.

PIERSON, W. J., JR.

1955. Wind generated gravity waves, *in*, Advances in Geophysics. Vol. 2, Academic Press, Inc., New York. Pp. 92–178.

1957. On the phases of the motions of ships in confused seas. Tech. Coll. Engng. Res. Div., N.Y. Univ., Contract NONR 284(17). 37 pp.

PRESS, HARRY AND J. W. TUKEY

1956. Power spectral methods of analysis and their application to problems in airplane dynamics, *in*, Flight Test Manual, NATO, Advisory Group for Aeronautical Research and Development, IV-C, pp. 1–41. (Reprinted as Monogr. Bell System No. 2606.)

PUTZ, R. R.

1955. Prediction of wave motion, *in*, Ships and waves. Counc. on Wave Res. and Soc. of Naval Architects and Marine Engineers. J. W. Johnson, Ed. Pp. 33–54.

SMED, JENS

1946. Monthly anomalies of the surface temperature in some areas of the north-western North Atlantic in 1876–1939 and 1945–1946. Ann. biol. Copenhague, *3*: 12–13.

1949. Monthly anomalies of the surface temperature in areas of the northern North Atlantic during the years 1876–1939 and 1945–1949. Anal. biol. Copenhague, *6*: 18–27.

TUKEY, J. W.

1949. The sampling theory of power spectrum estimates. Symposium on applications of autocorrelation analysis to physical problems, Woods Hole, June 13, 1949, Navexos–P-735. Office of Naval Research.

WADSWORTH, G. P.

1951. Application of statistical methods to weather forecasting. Compendium of Meteorology. Amer. Met. Soc., Boston. Pp. 849–858.