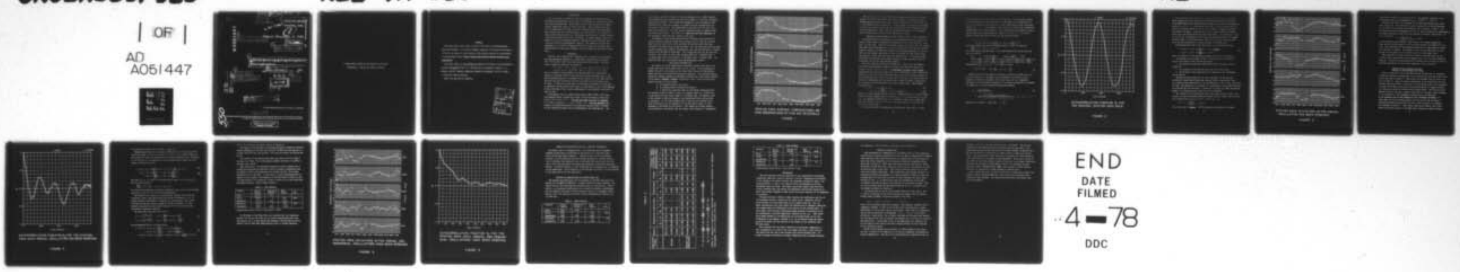
AD-A051 447 NAVY ELECTRONICS LAB SAN DIEGO CALIF F/G 8/10
A PRELIMINARY ANALYSIS OF SEA-SURFACE TEMPERATURE FORECASTING: --ETC(U)
JUN 62 C J VAN VLIET
UNCLASSIFIED NEL-TM-550 NL

| OF |

AD
A051447



END
DATE
FILMED
4-78
DDC

SR 011 01 01, Task 0401

PROBLEM/NEL Z1-24

MOST Project-2

NEL Technical Memorandum 550

ADA051447

Technical Memorandum No. TM-550

25 June 1962

date

NEL-TM-550

A PRELIMINARY ANALYSIS OF SEA-SURFACE TEMPERATURE FORECASTING: REMOVAL OF ANNUAL VARIATION.

C. J. Van Vliet 3050

author *code*

SR01101

SR0110101

22p.

DDC RECEIVED MAR 16 1978 F

AD No.

DDC FILE COPY

COPY AVAILABLE TO DDC DOES NOT PERMIT FULLY LEGIBLE PRODUCTION

U. S. Navy Electronics Laboratory, San Diego 52, California

253 550

The opinions expressed herein are those of the author(s) and are not necessarily the official views of the Navy Electronics Laboratory. If cited in the scientific literature, this document should be described as an unpublished memorandum.

DISTRIBUTION STATEMENT A

Approved for public release; Distribution Unlimited

A PRELIMINARY ANALYSIS OF SEA-SURFACE TEMPERATURE FORECASTING: REMOVAL OF ANNUAL VARIATION

FOREWORD

This memorandum covers only a portion of the work on the Environmental Statistics Studies. It has been prepared primarily for internal distribution at NEL to aid others who are working on time series analyses of oceanographic or meteorological data. ~~Only a limited distribution outside of NEL is contemplated.~~

The author wishes to acknowledge the valuable assistance in the preparation of this memorandum of Dr. E. R. Anderson for oceanographic aspects, J. S. Buehler and Mrs. Jeanne M. Baker for computer programming, and Mrs. Gladys L. Jones for data processing.

Work to 31 May 1962 is reported.

ACCESSION for
NTIS White Section ☑
DDC Buff Section ☐
UNANNOUNCED ☐
JUSTIFICATION Per ltr. on file
BY
DISTRIBUTION/AVAILABILITY CODES
Dist. AVAIL. and/or SPECIAL
A 23

INTRODUCTION

It is desirable to predict sea-surface temperatures into the future. Prediction intervals may be a few days, a few months, or even several years for long term prediction. There exist several long term records of daily surface temperatures taken at fixed locations. These records vary from five to eight years in length with only a moderate amount of missing data. The ultimate goal of a current study is the determination of a method of using certain characteristics of these long term records to forecast surface temperatures several months into the future. The same methods could be extended to forecast tactically important underwater sound velocities as a function of depth and geographical location. The immediate goal of this memorandum is to record a background for such a study, to do certain preliminary work in time series analysis, and to outline the next phase of the work.

BACKGROUND

Previous work at NEL by E.R. Anderson and G. W. Snedecor has included multiple regression analyses of surface sound velocities in the region $30^\circ - 32^\circ$ N, $120^\circ - 123^\circ$ W for the fiscal years 1951 to 1955. In particular, regression equations as complex as

$$V = \beta_0 + \beta_1 T + \beta_2 G + \beta_3 TG + \beta_4 D + \beta_5 D^2 + \beta_6 D^3 + \beta_7 D^4 + \beta_8 D^5 + \beta_9 TD + \beta_{10} TD^2 + \beta_{11} TD^3 + \beta_{12} TD^4 + \beta_{13} TD^5,$$

where D is time, G is longitude, T is latitude and V is surface sound velocity, have been fitted to observations for the individual years. The above regression equation considers the main effects of time, longitude and latitude, and the interactions of time with latitude and longitude with latitude. It should be noted that a fifth degree polynomial has been used to represent the variations with respect to time during a year. This approach differs from that of the present memorandum in which periodic functions of time are used.

Oceanographic data for Ocean Weather Station PAPA (50°N, 145°W) have been analyzed rather extensively as part of the Environmental Statistics Studies (NEL Problem Z1-24). The data are being collected by Canadians of the Pacific Oceanographic Group, Nanaimo, B.C., and are available in a sequence of reports published by the Fisheries Research Board of Canada in their Manuscript Report Series (Oceanographic and Limnological). The data include periodic readings of bathythermograph casts, surface temperature measurements, and surface and depth measurements of salinity, dissolved oxygen and silicates. Twice daily bathythermograph casts are made at 0200 and 1700 G.M.T., and daily samples for surface salinity determination are collected at 0200 G.M.T. This memorandum is concerned with the long term time series analysis of the daily surface temperature data as obtained from baththermograph records. The 0200 G.M.T. data are used when available; if not available, other data for a given day are used.

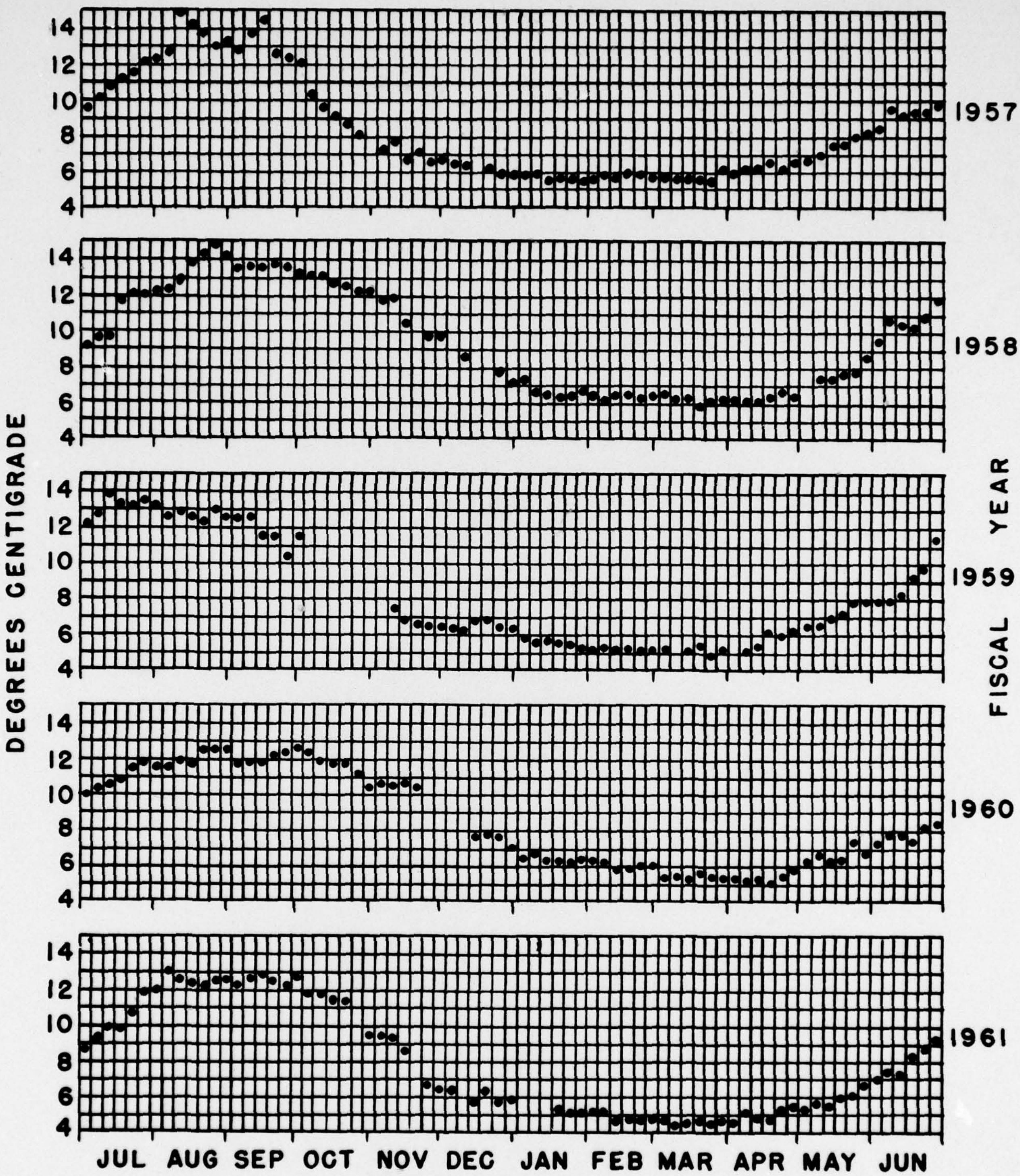
In addition to the Pacific Ocean Station PAPA data, which received the initial intensive analysis, there exist several years of data for Atlantic Ocean Stations DELTA (44°N, 41°W) and ECHO (35°N, 48°W) which have been analyzed in the same manner as that for PAPA. The PAPA data have been collected by oceanographers for specific oceanographic uses, and are definitely of higher quality than the DELTA or ECHO data. The Station PAPA data are for the time interval 1 July 1956 - 30 June 1961, Station DELTA for 1 September 1949 - 31 August 1957, and Station ECHO for 1 September 1949 - 31 August 1956. The DELTA and ECHO data were obtained from the Hydrographic Office.

TIME SERIES DECOMPOSITION

Any record of surface temperature necessarily reflects some sort of an annual oscillation and possibly some sort of a trend. It is convenient to remove the annual oscillation and the trend from a series of observations before applying any statistical prediction technique. The resulting surface temperature anomaly becomes the dependent variable for a time series analysis. If the anomaly exhibits a useable persistence, then a composite prediction formula consists of three terms,

1. an estimate based on the trend,
2. an estimate based on annual oscillation,
3. a statistical predictor based on the anomaly characteristics.

The general concept of trend is that of a gradual change in a system over a long period of time, where "long" is a relative term. An essential part of the concept is that the change in the system over fairly long periods is monotonic. It is always possible that a trend can be confused with a slow oscillation; this is certainly true if the record length is short compared to the period of oscillation. With respect to surface temperature prediction a few months into the future, such confusion is unimportant.



STATION PAPA SURFACE TEMPERATURES. MEDIAN OBSERVATIONS OF FIVE DAY INTERVALS.

FIGURE 1

For the time series under consideration in this memorandum, the record lengths of five to eight years are rather marginal in size for any trend determination, particularly in light of the existence, to be seen later, of a very dominant annual oscillation. Without going into detail, any changes in the system seem to be year to year fluctuations not satisfying the long term requirements of trend. For the following discussion, it is assumed that no trend exists in the surface temperature records.

ANALYSIS OF THE STATION PAPA DATA: REMOVAL OF THE ANNUAL VARIATION

The Station PAPA data analyzed in this memorandum consist of 1136 of the possible 1826 surface temperature measurements for the time interval 1 July 1956 - 30 June 1961. Except for two gaps of about one month each, the missing data days are distributed more or less at random within the five year period. Figure 1 displays the PAPA data modified for plotting convenience to the extent that the median observation in each successive five day interval has been plotted, rather than the individual observations. It should be emphasized that all analysis was actually performed on the individual daily observations, and not on these plotted medians.

The first obvious feature of the PAPA data is the annual oscillation with fairly sharp maxima in August to October and rather flat minima in February to May. Since the data are oscillatory in nature with a period necessarily close to one calendar year, it is justifiable, from a statistical point of view, to fit to the data some theoretical function which oscillates with period one year. Such a function would be fitted to the data either to provide an understanding of the oscillatory nature of surface temperatures, or to provide a prediction formula for these temperatures. The simplest such function is

$$X' = \beta_0 + \alpha \sin\frac{2\pi}{365}(D-\theta), \qquad (1A)$$

$$= \beta_0 + \beta_1 \sin\frac{2\pi}{365} D + \beta_2 \cos\frac{2\pi}{365} D, \qquad (1B)$$

where D is time measured in days from some arbitrary origin, X' is the fitted value of surface temperature, and either β_0, α, θ, or equivalently, $\beta_0, \beta_1, \beta_2$ are regression coefficients to be estimated. The year's length, here shown as 365 days, would be written to whatever accuracy is justified in the actual computations.

A conventional way of fitting such a function as Equation (1) to the points of Figure 1 is the Method of Least Squares. It is decidedly easier to use Equation (1B) rather than (1A) in this method of fitting since $\beta_0, \beta_1, \beta_2$ will appear as the solution of three simultaneous linear equations, whereas β_0, α, θ would appear as the solution of three quite complicated simultaneous non-linear equations. More specifically, the method of Least Squares consists of minimizing with respect to $\beta_0, \beta_1, \beta_2$ the sum of squares of the differences of observed and predicted values

$$Q = \Sigma (X - X')^2,$$
$$= \Sigma \left(X - \beta_0 - \beta_1 \sin \frac{2\pi}{365} D - \beta_2 \cos \frac{2\pi}{365} D\right)^2,$$

where the summation is over the N sample points. Thus the equations

$$\frac{\partial Q}{\partial \beta_i} = 0, \text{ for } i = 0, 1, 2,$$

are to be solved simultaneously. After some manipulation these equations can be written as

$$\begin{aligned}
&\beta_0 + \beta_1 \Sigma \sin \tfrac{2\pi}{365} D + \beta_2 \Sigma \cos \tfrac{2\pi}{365} D = \Sigma X \\
&\beta_0 \Sigma \sin \tfrac{2\pi}{365} D + \beta_1 \Sigma \left(\sin \tfrac{2\pi}{365} D\right)^2 + \beta_2 \Sigma \left(\sin \tfrac{2\pi}{365} D\right)\left(\cos \tfrac{2\pi}{365} D\right) = \Sigma X \sin \tfrac{2\pi}{365} D \qquad (2) \\
&\beta_0 \Sigma \cos \tfrac{2\pi}{365} D + \beta_1 \Sigma \left(\sin \tfrac{2\pi}{365} D\right)\left(\cos \tfrac{2\pi}{365} D\right) + \beta_2 \Sigma \left(\cos \tfrac{2\pi}{365} D\right)^2 = \Sigma X \cos \tfrac{2\pi}{365} D
\end{aligned}$$

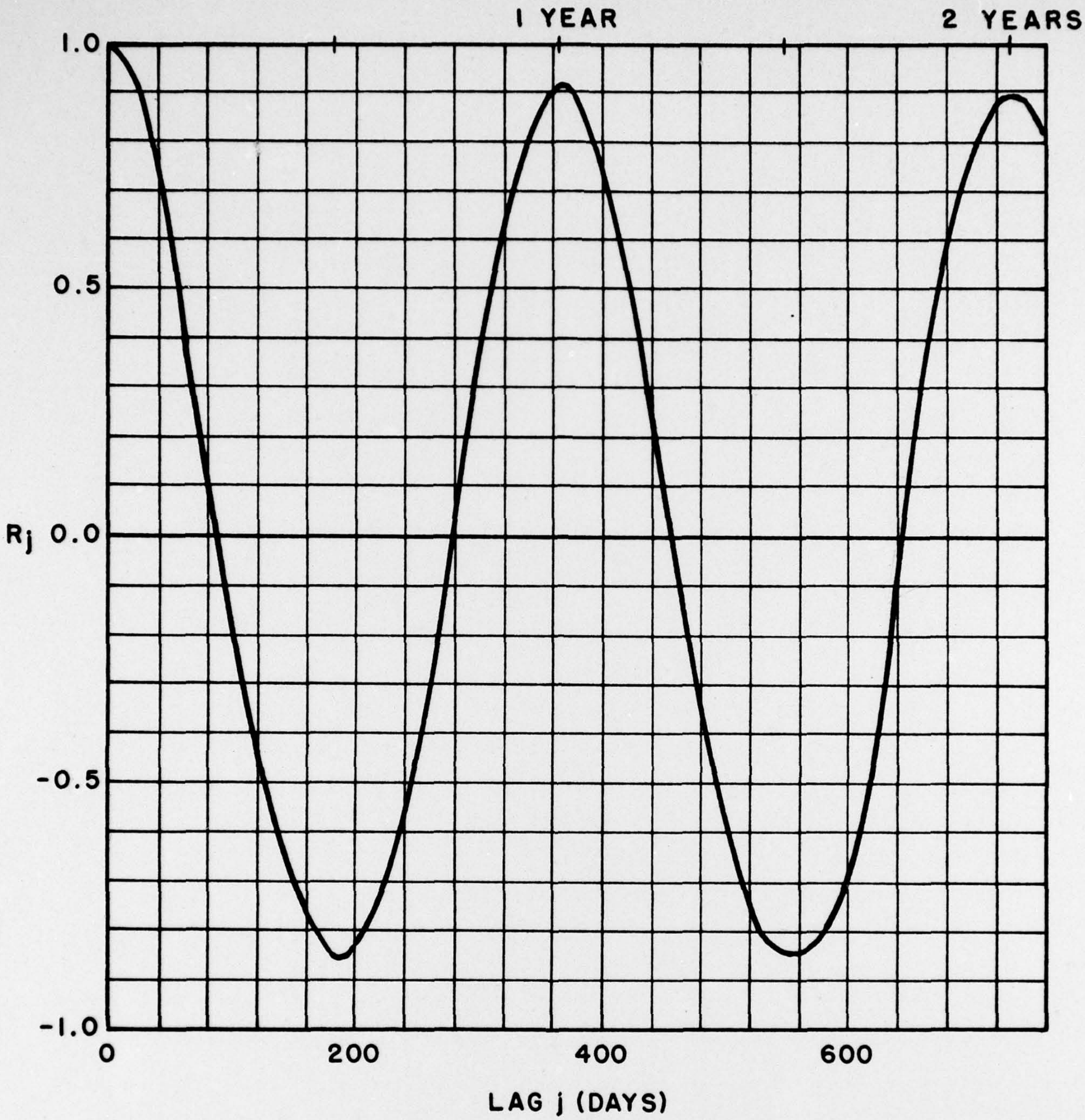
Having determined $\beta_0, \beta_1, \beta_2$ it is easy to obtain

$$\alpha = \sqrt{\beta_1^2 + \beta_2^2}, \quad \theta = \frac{365}{2\pi} \arctan \frac{-\beta_2}{\beta_1}. \qquad (3)$$

Before discussing the numerical results of fitting the PAPA data, consider the following. Although a visual observation of Figure 1 would suggest the fitting of Equation (1B) to the data, and is a reasonable technique in an investigative problem, it is perhaps enlightening to present the PAPA data in terms of the autocorrelation function

$$R_j = \frac{\operatorname{cov} X_i X_{i+j}}{\sqrt{\operatorname{var} X_i \operatorname{var} X_{i+j}}}, \qquad (4)$$
$$= \frac{(N-j)\Sigma X_i X_{i+j} - (\Sigma X_i)(\Sigma X_{i+j})}{\sqrt{[(N-j)\Sigma X_i^2 - (\Sigma X_i)^2][(N-j)\Sigma X_{i+j}^2 - (\Sigma X_{i+j})^2]}}, \text{ for lags } j = 0, 1, \cdots,$$

where all the summation signs mean $\Sigma = \sum_{i=1}^{N-j}$.



AUTOCORRELATION FUNCTION R_j FOR THE ORIGINAL STATION PAPA DATA.

FIGURE 2

The autocorrelation function as defined has the value $R_o = 1$ for lag zero, meaning that the time series is perfectly correlated with itself. If the PAPA time series repeated itself exactly every year, the autocorrelation for $j = 1, 2, \cdots$, years would also be $R_j = 1$, since there would be perfect correlation whenever the time series and the lagged time series were exactly in phase. Since the data do not repeat themselves exactly every year, the autocorrelation function can never be one except at $j = 0$, but as is seen from Figure 2, the autocorrelation function has very large positive peaks at yearly intervals. This is a definite indication that there are oscillations with period one year in the PAPA data.

Fitting Equation (1B) to the data yields the regression equation

$$X' = 8.54 - 3.43 \sin\frac{2\pi}{365}D - 1.79 \cos\frac{2\pi}{365}D, \qquad (5)$$
$$= 8.54 + 3.89 \sin\frac{2\pi}{365}(D - 154.6).$$

where the units of X' are degrees Centigrade, and day D=1 is January 1.

As measures of the goodness of fit of this equation to the actual data we can consider three related quantities.

1. The multiple correlation coefficient between the observed and fitted X's.

2. The variance of the observations about the fitted curve.

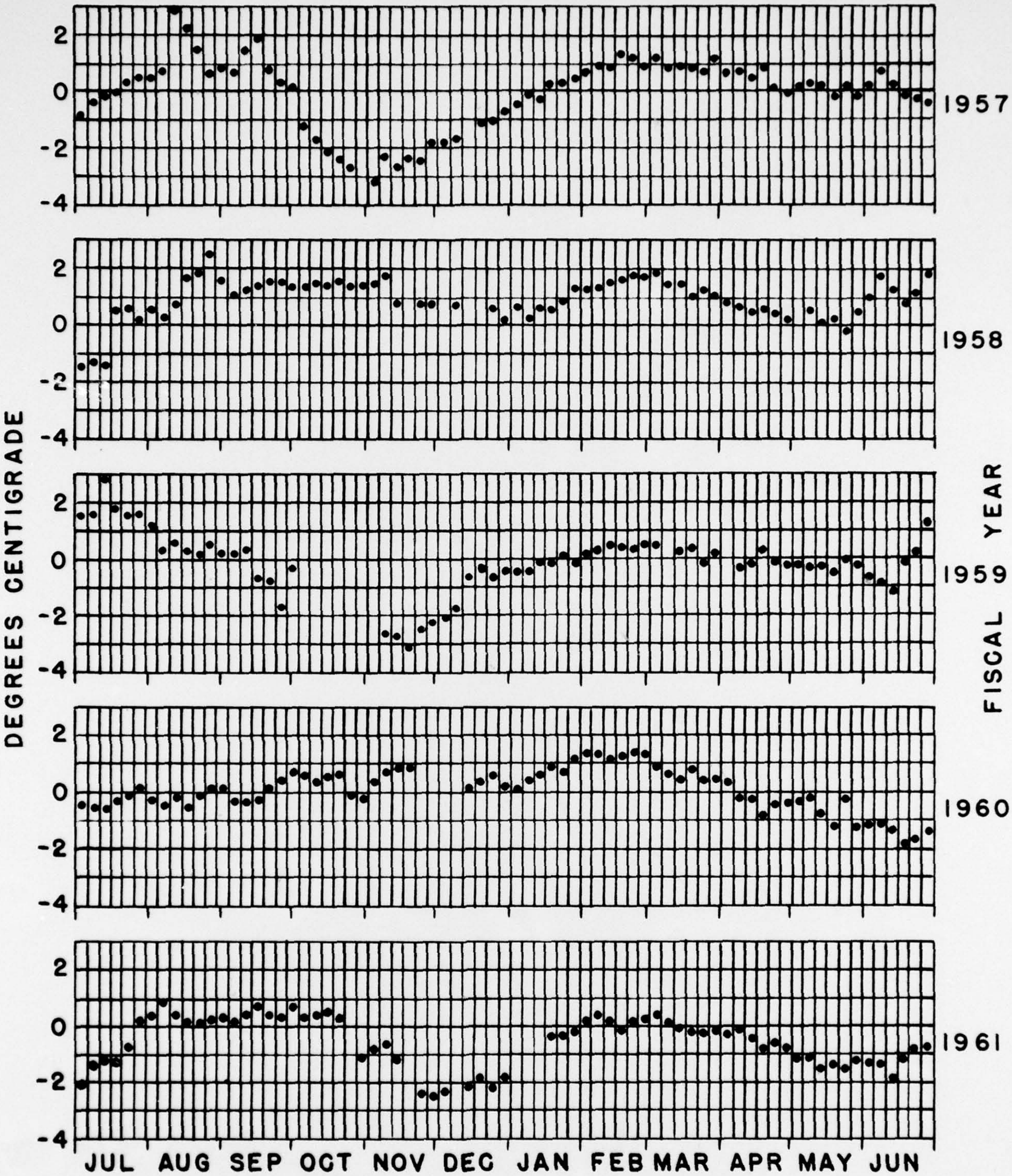
3. The F-ratio of the regression sum of squares to the unexplained sum of squares.

Considering these quantities in turn, the square of the multiple correlation coefficient between the observed surface temperature X and the fitted surface temperature X^1 for the PAPA data is $R^2 = 0.863$, from which R=0.929. The quantity R is a measure of the strength of the linear relationship between the two variables. A value of R as large as 0.929 implies that Equation (5) provides quite a good fit to the PAPA data.

The variance of the observations about their mean value is 8.904. The variance of the observations about the fitted curve is 1.220. The relationship between these variances and R^2 is observed to be

$$R^2 = 1 - \frac{1.220}{8.904} = 0.863,$$

which leads one to express $100R^2$ as being the percentage of variance



STATION PAPA DEVIATIONS AFTER ANNUAL OSCILLATION HAS BEEN REMOVED.

FIGURE 3

explained by fitting the regression curve. The standard deviation of the observations about the regression curve is $\sqrt{1.220} = 1.105°C$.

Finally, the total sum of squares of the PAPA observations about their mean is 10,114.6. The reduction in sum of squares due to fitting Equation (5) is 8729.2, which leaves an unexplained sum of squares of 1385.4. With the proper normality assumptions, the last two sums of squares are independently Chi-square distributed with 2 and 1133 degrees of freedom, respectively, so that the F-ratio

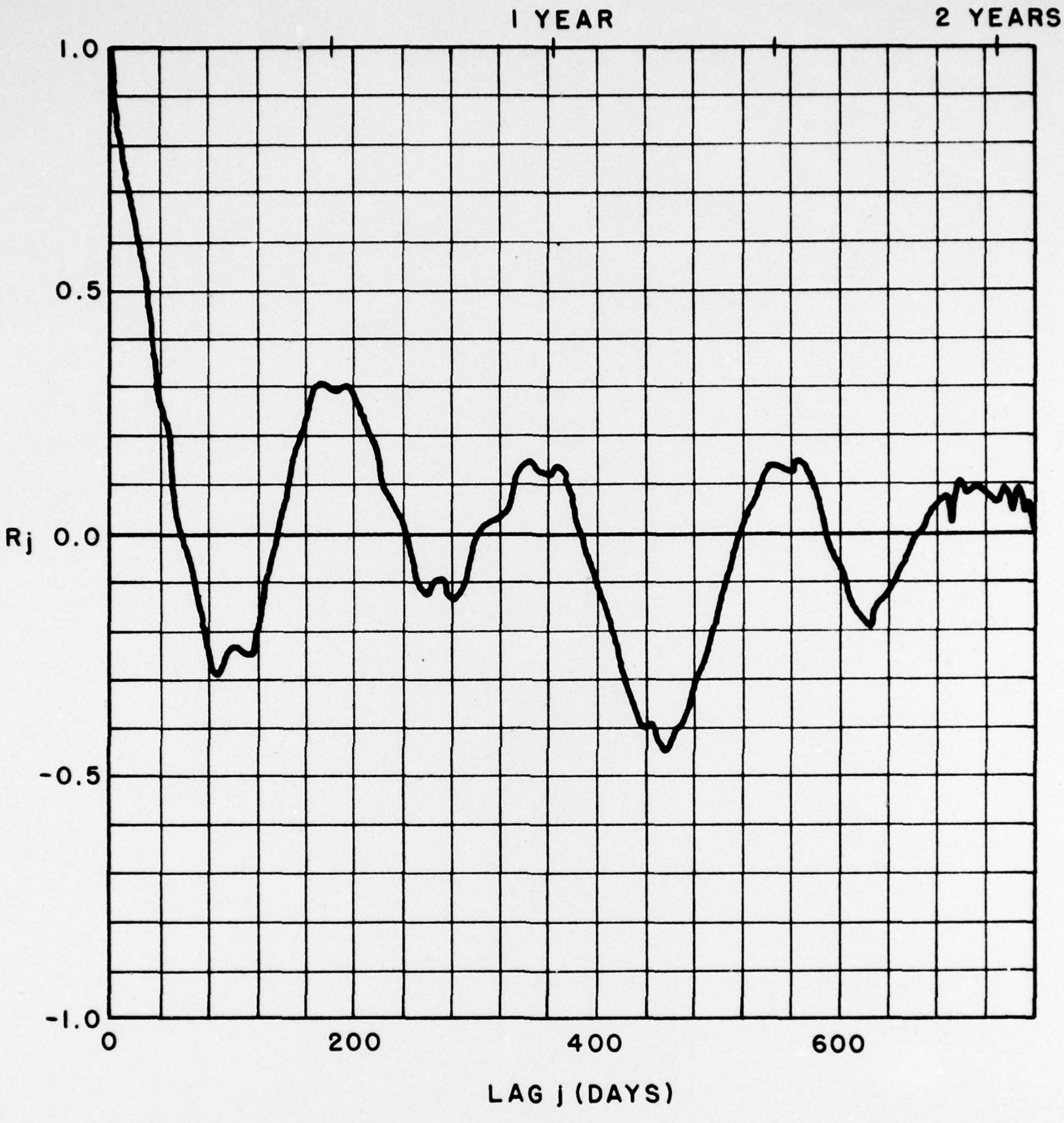
$$F = \frac{8729.2 \div 2}{1385.4 \div 1133} = 3569$$

can be written and interpreted. This value of F is highly significant since the 1% critical value of F for the above degrees of freedom is $F_{0.01} = 4.62$. One concludes that the regression Equation (5) significantly reduces the total sum of squares. The analysis to follow will show that the unexplained sum of squares can be reduced further in a meaningful way.

The clue that an annual oscillation existed in the PAPA data was provided by visual observation of the original data, or of the autocorrelation function. Statistical justification for fitting a given function to data is provided a-posteriori by the F-ratio analysis, or its equivalent.

ANALYSIS OF THE STATION PAPA DATA: REMOVAL OF THE SEMIANNUAL VARIATION.

The fit of Equation (5) to the PAPA data can perhaps best be displayed visually by plotting the deviations of the observations from the fitted curve, as has been done in Figure 3. This permits an expended vertical scale over that of Figure 1. Again the median observation in each successive five day interval has been plotted. A feature of the deviation plots is the existence for each fiscal year of two fairly well defined maxima at roughly one-half year intervals separated by fairly well defined minima. For 1960 the first maximum occurs in the last 5 days of 1959, Figure 3 provides visual evidence of the existence of some semiannual oscillation in the PAPA data.
as an alternate clue to the existence of the semiannual variation, we again make use of the autocorrelation function as was done in Figure 2. The autocorrelation function of the devia tions of Figure 3 is computed and displayed in Figure 4. This figure shows quite clearly that a semiannual oscillation exists in the PAPA data, although the peaks in Figure



AUTOCORRELATION FUNCTION R_j FOR THE STATION PAPA DATA. ANNUAL OSCILLATION HAS BEEN REMOVED

FIGURE 4

4 are smaller in magnitude than those of Figure 2.

Having demonstrated the existence of some sort of a semiannual oscillation in the data, it is now convenient to revert to the original PAPA data and remove the annual and semiannual oscillations simultaneously. Analogous to the work above, the simplest function containing both annual and semiannual oscillatory terms is

$$X' = \beta_0 + \alpha_1 \sin\frac{2\pi}{365}(D-\theta_1) + \alpha_2 \sin\frac{4\pi}{365}(D-\theta_2), \tag{6A}$$

$$= \beta_0 + \beta_1 \sin\frac{2\pi}{365}D + \beta_2 \cos\frac{2\pi}{365}D + \beta_3 \sin\frac{4\pi}{365}D + \beta_4 \cos\frac{4\pi}{365}D. \tag{6B}$$

The quantity $Q = \sum (X-X')^2$ is minimized with respect to the five β's. That is, the equations

$$\frac{\partial Q}{\partial \beta_i} = 0, \text{ for } i = 0, 1, 2, 3, 4,$$

are solved simulatneously. Analogous to Equations (2), there are five linear equations in the β's to be solved simultaneously.

The minimum of Q obtained by simulteneously fitting with annual and semiannual terms is the same minimum that would be obtained if a fit of annual terms were followed by a separate fit of semiannual terms to the resulting deviations, obvious changes of origin being made. The two separate regression equations added together would be the same as the simultaneous regression equation.

The amplitudes and phases are given by

$$\alpha_1 = \sqrt{\beta_1^2+\beta_2^2}, \quad \theta_1 = \frac{365}{2\pi}\arctan\frac{-\beta_2}{\beta_1}, \tag{7}$$
$$\alpha_2 = \sqrt{\beta_3^2+\beta_4^2}, \quad \theta_2 = \frac{365}{4\pi}\arctan\frac{-\beta_4}{\beta_2}.$$

Fitting Equation (6B) to the data yields

$$X' = 8.49 - 3.30 \sin\frac{2\pi}{365}D - 1.92 \cos\frac{2\pi}{365}D + 0.750 \sin\frac{4\pi}{365}D - 0.405 \cos\frac{4\pi}{365}D, \tag{8}$$
$$= 8.49 + 3.82 \sin\frac{2\pi}{365}(D-152.0) + 0.85 \sin\frac{4\pi}{365}(D-14.4),$$

where X is in degrees Centigrade and D=1 is January 1.

The square of the multiple correlation coefficient between the observed surface temperature X and this new fitted X^1 is R^2=0.903, and R=0.950, an increase from R^2=0.863, and R=0.929, when only the annual oscillation was fitted.

The variance of the observations about the fitted curve is 0.865, a decrease from 1.220. The corresponding standard deviation is 0.930°C, a decrease from 1.105°C.

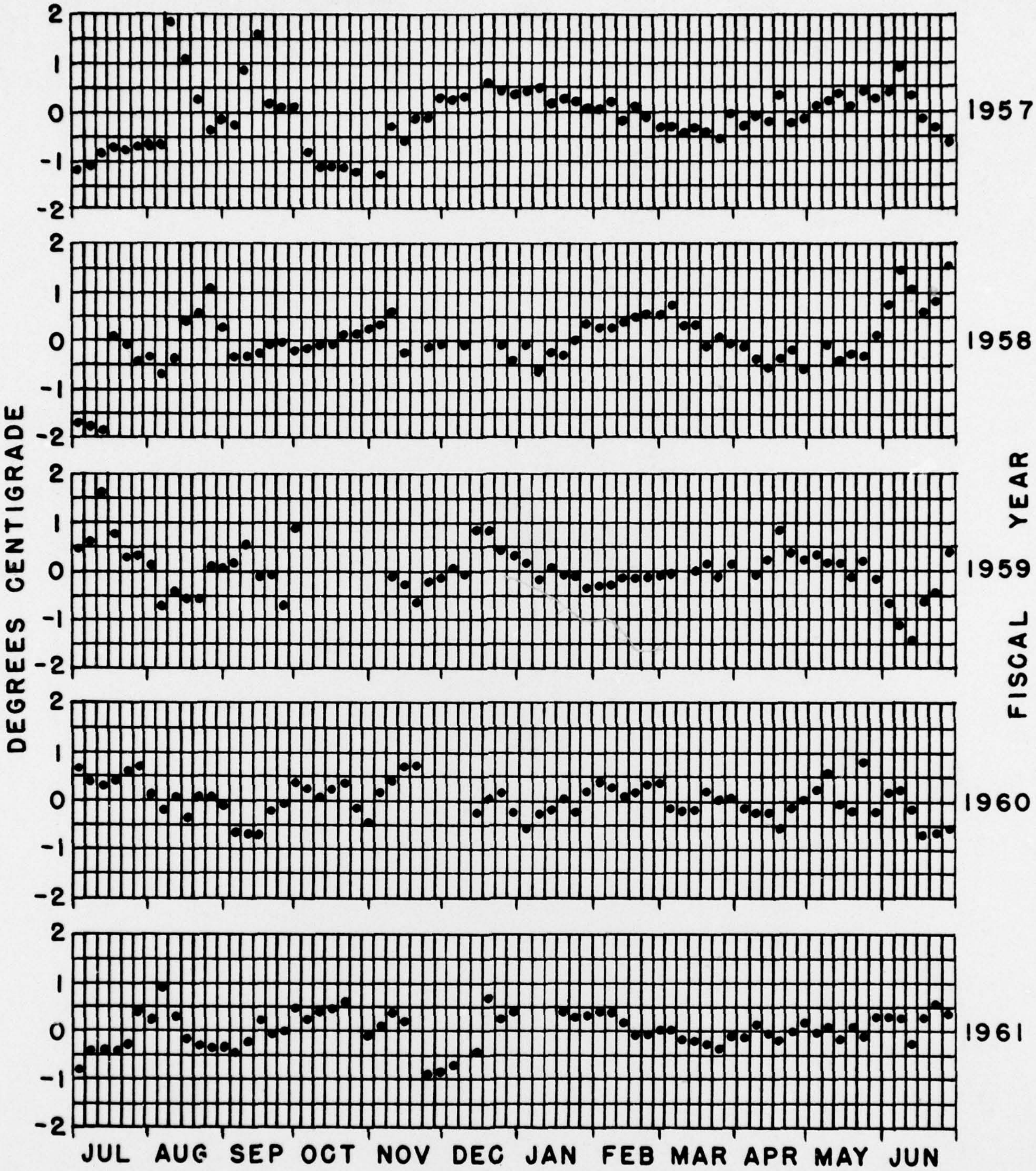
More importantly, the additional reduction in sum of squares due to fitting the semiannual oscillation is 403.0, leaving an unexplained sum of squares of 982.4, and resulting in an F-ratio of 232, still highly significant at the 1% level. At this point it would seem appropriate to display a table similar to an Analysis of Variance Table in order to test the significance of the reduction in sum of squares due to each additional regression variable. This is done in Table 1.

TABLE 1. PAPA ANALYSIS

Variable	Sum of Squares	Degrees of Freedom	Mean Square	F Ratio
Total	10114.6	1135		
Annual	8729.2	2	4365.	3569
Unexplained	1385.4	1133	1.223	
Semiannual	403.0	2	201.5	232
Unexplained	982.4	1131	0.869	

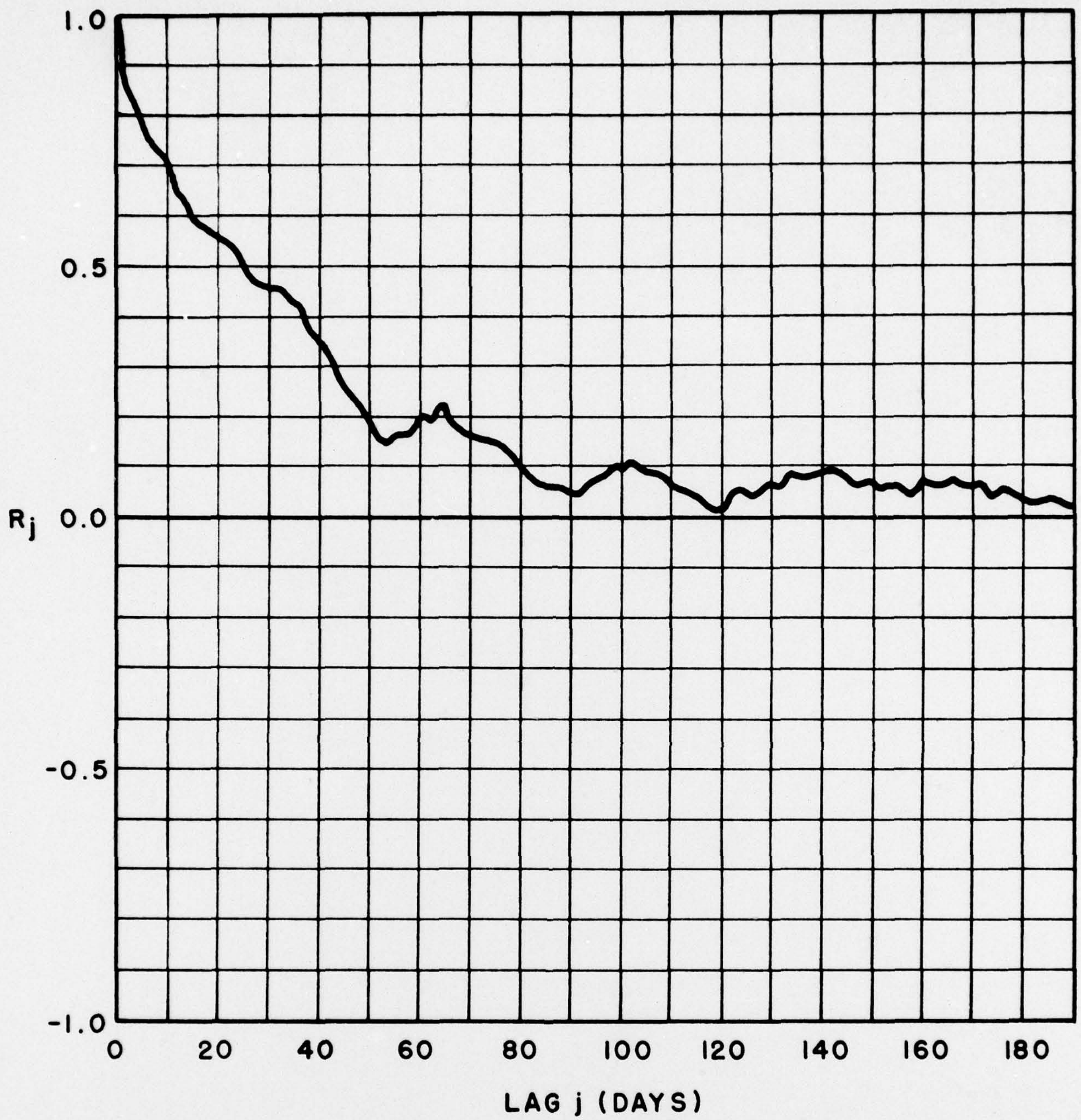
(Significance Levels: $F_{0.05}$ = 3.00, $F_{0.01}$ = 4.62)

On the basis of the above table one concludes that the regression Equation (8) reduces the total sum of squares significantly more than does Equation (5), so that annual and semiannual oscillations should be removed from the PAPA data simultaneously prior to further analysis.



STATION PAPA DEVIATIONS AFTER ANNUAL AND SEMIANNUAL OSCILLATIONS HAVE BEEN REMOVED

FIGURE 5



AUTOCORRELATION FUNCTION R_j FOR THE STATION PAPA DATA. ANNUAL AND SEMIANNUAL OSCILLATIONS HAVE BEEN REMOVED.

FIGURE 6

ANALYSIS OF THE STATION PAPA DATA; RESIDUAL VARIABILITY

As before, the fit of Equation (8) to the PAPA data can be displayed visually by plotting the deviations of the observations from the fitted curve, as has been done in Figure 5. Alternatively, the autocorrelation function of these deviations has been plotted in Figure 6. Neither Figures 5 nor 6 indicates the presence of any oscillation with period less than one year in the deviations after the annual and semiannual oscillations have been removed. No further analysis of the unexplained sum of squares will be attempted in this memorandum. Future analysis of the residual variability will be outlined in the final section of this memorandum.

ANALYSIS OF STATION DELTA AND STATION ECHO DATA

The methods developed above for the analysis of the PAPA data have been applied to the DELTA and ECHO data. Similar results have been obtained. The results for the three stations are presented below in tabular form. Table 2 presents the regression coefficients, the multiple correlation coefficients, and the standard deviations of the observations about the regression curves. Tables 3 and 4 are Analysis of Variance tables corresponding to Table 1 for the PAPA data.

TABLE 3. DELTA ANALYSIS

Variable	Sum of Squares	Degrees of Freedom	Mean Square	F Ratio
Total	9285	1287		
Annual	6203	2	3101	1293
Unexplained	3082	1285	2.398	
Semiannual	570	2	285	146
Unexplained	2512	1283	1.958	

TABLE 2

Oscillation Fitted	Station	Regression Coefficients					Amplitudes		Phases		Multiple Correlation Coefficient		StDev. About Reg.
		β_0	β_1	β_2	β_3	β_4	a_1	a_2	θ_1	θ_2	R^2	R	σ
Annual	DELTA	17.79	-2.81	-1.35	—	—	3.12	—	156.7	—	.668	.817	1.547
	ECHO	22.06	-3.66	-1.67	—	—	4.02	—	157.7	—	.885	.941	1.010
	PAPA	8.54	-3.43	-1.79	—	—	3.89	—	154.6	—	.863	.929	1.105
Annual and Semiannual	DELTA	17.85	-2.73	-1.20	.806	.378	2.98	.89	158.6	-12.8	.729	.854	1.396
	ECHO	22.06	-3.56	-1.67	.766	.191	3.94	.79	157.0	-7.1	.922	.960	.835
	PAPA	8.49	-3.30	-1.92	.750	-.405	3.82	.85	152.0	14.4	.903	.950	.930

$$X' = \beta_0 + \beta_1 \sin\frac{2\pi}{365}D + \beta_2 \cos\frac{2\pi}{365}D + \beta_3 \sin\frac{4\pi}{365}D + \beta_4 \cos\frac{4\pi}{365}D,$$

$$= \beta_0 + a_1 \sin\frac{2\pi}{365}(D-\theta_1) + a_2 \sin\frac{4\pi}{365}(D-\theta_2).$$

Units of regression coefficients, amplitudes and standard deviations are degrees Centigrade. Units of phases are days.

TABLE 4. ECHO ANALYSIS

Variable	Sum of Squares	Degrees of Freedom	Mean Square	F Ratio
Total	13625	1530		
Annual	12065	2	6032	5902
Unexplained	1561	1528	1.022	
Semiannual	493	2	246	351
Unexplained	1068	1526	.700	

(Significance Levels: $F_{0.05} = 3.00$, $F_{0.01} = 4.62$)

CONCLUSIONS

The most important discovery described in this memorandum is the highly significant reduction in sums of squares in surface temperatures attributable to the fitting to the data of annual and semiannual oscillatory terms. Considering Tables 1, 3 and 4, the large F-ratios corresponding to fitting semiannual terms even after the highly significant annual terms have been fitted, indicates that both annual and semiannual oscillations should be removed from surface temperature data before statistical analysis of the anomalies is performed.

Turning attention to Table 2, which concerns the coefficients and fit of the regression curves, a feature of the table is the similarity for the three stations of terms within each of the columns except β_0 and θ_2. These similarities exist even though the stations are in different oceans and even though there are differences in the quality of the recorded data. It would appear that the major differences among stations is accounted for by the differences in absolute magnitudes reflected by the β_0. The differences among stations of the amplitudes are relatively minor, as are the differences of the phases, although it is the differences among the phases θ_2 which determine the fine details of the shapes of the individual best fitting curves.

One concludes that the annual variation in sea-surface temperatures is well represented by the sum of two very simple functions, a sine function with period one year and a sine function with period one-half year. Although an annual oscillation in surface temperatures was certainly expected, the existence of the semiannual oscillation was not expected.

OUTLINE OF FUTURE WORK

This memorandum has demonstrated the existence of annual and semiannual oscillations in surface temperature data for one Pacific Ocean and two Atlantic Ocean stations. When these oscillations are removed from the data there remain time series consisting of the deviations from the regression curves. Referring to Figure 6, which shows the autocorrelation function for these deviations for the PAPA data, there are indications of a persistence of, say, seventy-five days in this data. The values of autocorrelation out to this point are probably significant. This persistence also appears in Figure 5, showing the deviations themselves, in that successive deviations do not vary randomly about the axis, but are above or below the axis for fairly long periods of time. It is questionable that the oscillations in the autocorrelation function of Figure 6 are meaningful much beyond 100 days. They are due to nothing more than the characteristics of the particular five year sample available for analysis.

If a persistence of two or three months exists in a time series, then it should be possible to use this persistence to predict into the future. Present plans envisage a prediction formula of the form

$$Y'_{t+k} = C_0 Y_t + C_1 Y_{t-1} + \cdots + C_n Y_{t-n} ,$$

where Y is the surface temperature anomaly. This equation indicates that a temperature anomaly k days in the future will be predicted as a linear combination of the n+1 most recent observations. With k and n as parameters, the C_i will be determined by minimizing $\Sigma(Y_{t+k} - Y'_{t+k})^2$ for a known portion of a time series. The ultimate evaluation of the prediction formula will be based on how well it predicts the unknown (to the predictor) future of the time series. As mentioned early in the memorandum, the composite prediction formula would combine the above anomaly prediction with trend and annual oscillation terms.

Further investigation should be made, for other regions of the ocean, into the existence of the apparently well defined semiannual oscillation in surface temperatures. The use of this oscillation as a statistical tool is justified, even if a physical interpretation is difficult. The stations discussed in this memorandum are in the open ocean in deep water. Two other situations of interest are that of a coastal station in shallow water and the intermediate situation of an ocean station in relatively shallow water. Long term records exist which will be analyzed for these situations in the same manner as discussed in this memorandum. There are 45 years of data taken at a coastal station, namely Scripps pier in La Jolla, California. There are some 20 years of data taken at each of two stations on Queen Charlotte Island, LANGARA and ST. JAMES, which because of their exposed nature can be considered as ocean stations in shallow water.

As is always the case, it would be desirable to obtain much longer time series to use in investigations for trend, and for more extensive auto-correlation type analyses. The long records mentioned above should help solve this problem.