SEA-SURFACE TEMPERATURE ESTIMATION

An Empirical Study of the Effect of Missing Data on Regression and Autocorrelation Analyses of Time Series of Data

C. J. Van Vliet • Research Report • 5 January 1965

U. S. NAVY ELECTRONICS LABORATORY, SAN DIEGO, CALIFORNIA 92152 • A BUREAU OF SHIPS LABORATORY

THE PROBLEM

Develop statistical, physical, and computer techniques for interpreting, summarizing, and extrapolating oceanic and meteorologic data for reliable estimation of the sound velocity distribution in the ocean. Specifically, determine the effect of random missing data and the effect of several long periods of missing data on the regression and autocorrelation analyses used in the estimation of sea-surface temperatures.

RESULTS

Analysis of records of sea-surface temperature, taken in the N. Atlantic and N. Pacific and up to 40 years in length, has shown that:

1. For many stations, the time series of sea-surface temperatures have missing temperatures scattered at random throughout the series. For each day there is a certain probability that the temperature will be missing. For such series, proper adjustments can be made in the computations of the regression and autocorrelation coefficients. The random deletion of data yields coefficients whose variances exceed those of a complete time series by an amount as predicted by the reduction in sample size.

2. For certain stations, there are an excessive number of longer sequences of missing data. For the time series considered, the increase in the variances of the regression coefficients attributable to this nonrandom missing data is twice the increase attributable to random missing data. Alternatively, for fractions of missing data greater than 0.2, time series with nonrandom missing data will have regression coefficient variances equal to those the same series with 0.15 more missing data would have, if all the missing data were random.

3. The effect of nonrandom, longer sequences of missing data on autocorrelation coefficients is less pronounced than for regression coefficients. The increase in the variances of autocorrelation coefficients attributable to nonrandom missing data is 1.2 times the increase attributable to random missing data. Alternatively, for fractions of missing data greater than 0.2, time series with nonrandom missing data will have autocorrelation coefficient variances equal to those the same series with 0.05 more missing data would have, if all the missing data were random.

RECOMMENDATIONS

1. Examine the nature of missing data in time series of sea-surface temperatures as to the randomness of occurrence in time. Then apply the appropriate results of this report in estimating the variances of regression coefficients and autocorrelation coefficients.

2. Perform an investigation similar to the present one on the effect of missing data for the regression problem but with several independent variables, namely, time, depth, and geographical location. The dependent variable will be water temperature.

3. Examine the effect of missing data on the short range prediction of sea-surface temperatures.

ADMINISTRATIVE INFORMATION

Work was performed under SR 004 03 01, Task 0586 (NEL L40551, formerly L4-5) by a member of the Computer Center. The report covers work from October 1963 to August 1964 and was approved for publication 5 January 1965.

The author wishes to express appreciation to E. R. Anderson of the NEL Oceanometrics Group for advice on the oceanographic aspects of the problem.

CONTENTS

TABLES

ILLUSTRATIONS

REVERSE SIDE BLANK

INTRODUCTION

The time-series analysis of sea-surface temperatures is of interest to oceanographers, meteorologists, and biologists. This study discusses methods used in an analysis of daily sea-surface temperatures. It is the first in a proposed series and is primarily concerned with the effect of missing data in certain regression and autocorrelation analyses. Only enough detail of these analyses will be included to ensure a degree of completeness to the present study.

Many time-series measurements have been made at various locations. In the eastern Pacific Ocean such measurements have been made by Canadian and American oceanographers at coastal, island, and ship locations for time periods up to 45 years. These data have been the subject of numerous papers including, among others, those of Pickard and McLeod[1] and Roden.[2,3] This study differs

[1] Pickard, G. L. and McLeod, D. C., "Seasonal Variation of Temperature and Salinity of Surface Waters of the British Columbia Coast," Journal of the Fisheries Research Board, Canada, v. 10, p. 125-145, 1953

[2] Roden, G. I., "Spectral Analysis of a Sea-Surface Temperature and Atmospheric Pressure Record off Southern California," Journal of Marine Research, v. 16, p. 90-95, 1958

[3] Roden, G. I., "On Nonseasonal Temperature and Salinity Variations Along the West Coast of the United States and Canada," California Cooperative Oceanic Fisheries Investigations. Reports, v. 8, p. 95-119, 1961

from those cited in that the original daily temperatures are used in the analysis without a preliminary smoothing by monthly averaging.

The purpose of time-series analysis is to isolate *trend, oscillation,* and *random elements,* which are defined as follows. Trend is a gradual increase or decrease in a system over a long period of time; an oscillation is a variation about the trend that occurs with more or less regularity over some time interval; and a random element is an unpredictable variation in the variable. If long term trend does not exist, then the primary need is the statistical fitting of some function to time series to represent the oscillatory element.

Several sets of daily sea-surface temperatures have been examined. Measurements were made at the two open ocean and four island or coastal locations shown in figure 1. To indicate how individual temperature measurements vary throughout the year, one year of measurements for each location is presented in figure 2. These years of temperatures are taken from records that vary in length from 7 to 40 years. Pertinent information about the stations yielding these records are included in table 1.

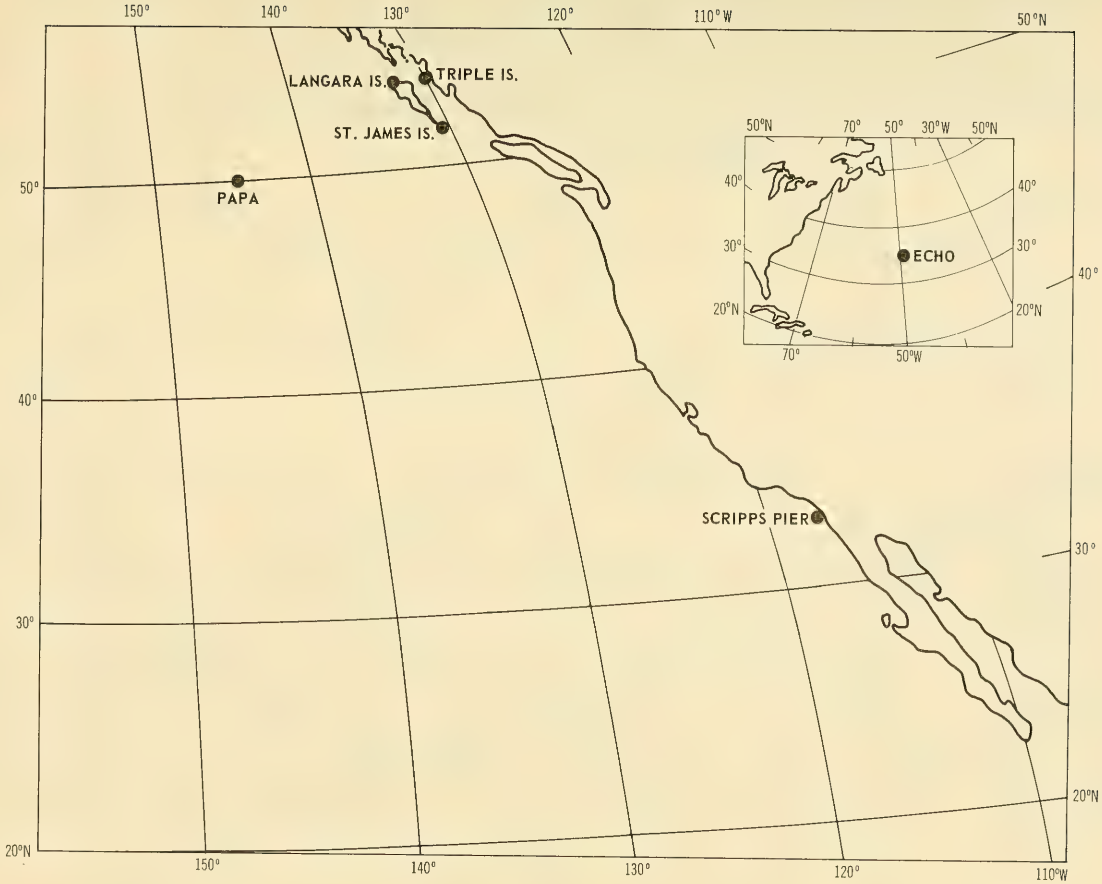

Figure 1. Geographical location of oceanographic stations.

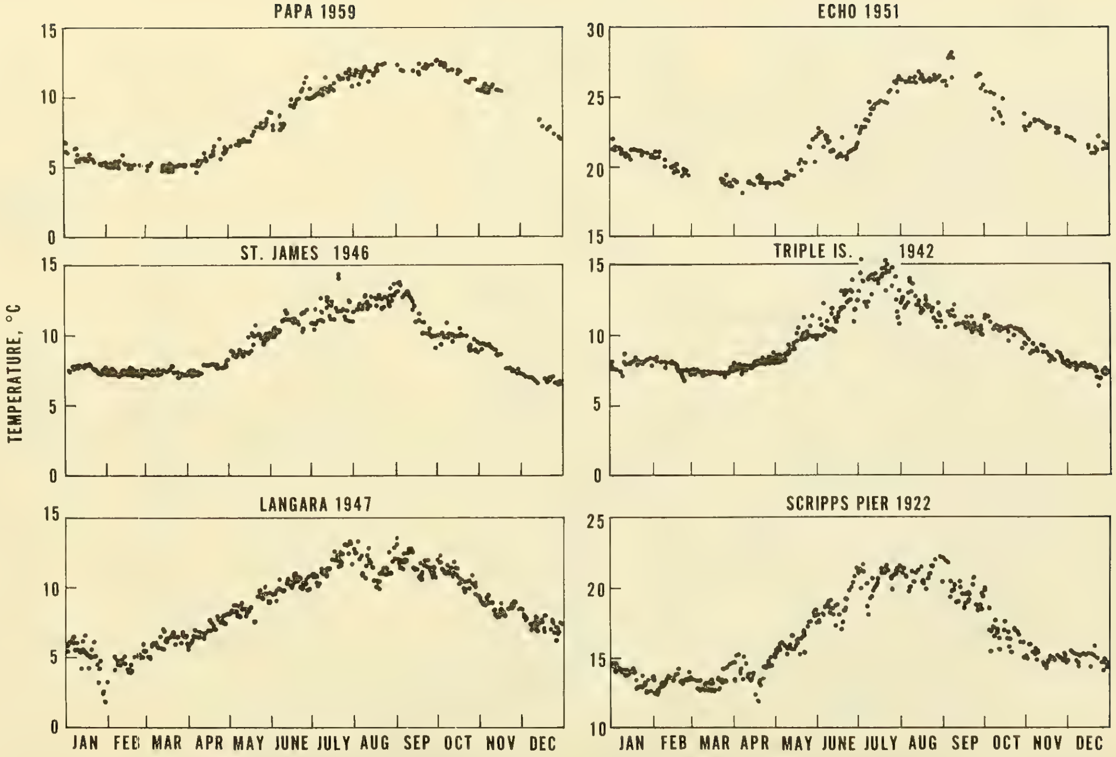

Figure 2. Sea-surface temperatures as a function of time for selected years of data.

TABLE 1. SEA-SURFACE TEMPERATURE TIME SERIES

Location	Time Period	Number Days	Number Daily Observations	Percent Possible Observations
Weather Ship "PAPA" 50°N 145°W North Pacific	1/56 - 1/63 7 yr	2557	1690	66
Weather Ship "ECHO" 35°N 48°W North Atlantic	9/49 - 9/56 7 yr	2557	1533	60
St. James Island 52°N 131°W North Pacific	1/40 - 1/61 21 yr	7671	6180	81
Triple Island 54°N 131°W North Pacific	1/40 - 1/61 21 yr	7671	7244	95
Langara Island 54°N 133°W North Pacific	1/41 - 1/61 20 yr	7304	6402	88
Scripps Pier 33°N 117°W North Pacific	1/21 - 1/61 40 yr	14610	14352	98

REGRESSION AND AUTOCORRELATION ANALYSES

A visual observation of the data suggests statistically fitting some theoretical function which oscillates with a period of one year. Further justification is provided by the autocorrelation function

$$C_k = \mathrm{COV}(T_i T_{i+k})/\mathrm{VAR}(T_i), \text{for lags } k = 0, 1, 2, \ldots. \qquad (1)$$

In this application the variable T_i is the sea-surface temperature on day i, T_{i+k} is the temperature k days later, and COV and VAR are the covariance and variance of the variables as indicated. Computation of the autocorrelation function yields peaks whose magnitudes and spacings strongly indicate the existence of an annual oscillation in the time series.

The simplest model consisting of an oscillatory function with period one year is

$$T' = \beta_0 + \alpha \sin\,[2\pi(D-\theta)/365] + \epsilon \qquad (2A)$$

$$= \beta_0 + \beta_1 \sin\,(2\pi D/365) + \beta_2 \cos\,(2\pi D/365) + \epsilon \qquad (2B)$$

where D is time measured in days from some arbitrary origin and T' is the fitted value of the surface temperature. Fitting the function of equation (2B) to the observed surface temperatures T using the method of least squares yields estimates of the regression coefficients β_0, β_1 and β_2 and an estimate of the variance of ϵ. The amplitude α and phase θ can be obtained from β_1 and β_2. The quantity ϵ is the random, or error, or residual term.

If the residuals $T - T'$ are examined visually or by computation of the autocorrelation function of the residual time series, a fairly strong semiannual oscillation is discovered for some of the stations. This suggests the model

$$T' = \beta_0 + \beta_1 \sin(2\pi D/365) + \beta_2 \cos(2\pi D/365) \tag{3}$$

$$+ \beta_3 \sin(4\pi D/365) + \beta_4 \cos(4\pi D/365) + \epsilon$$

The addition of semiannual oscillatory terms to the regression equation improves the fit obtained with the annual terms. Tests of significance of sums of squares attributable to annual and semiannual oscillations are performed using the appropriate F-ratios.

Computation of the autocorrelation functions of the residual time series after equation (3) has been fitted to the series of sea-surface temperatures yields the plots of figure 3. These residuals are themselves autocorrelated, although no additional oscillatory terms exist. The least squares method is valid if (1) the error between the true regression curve and the observed value is distributed independently of the independent variables with zero mean and constant variance; and (2) ideally, successive errors are distributed independently of one another. Actually, the problem of using the method of least squares when the error terms are autocorrelated has been solved if the ϵ's follow certain autoregressive processes.[4] The autocorrelated residuals should affect the distributions of the regression coefficients. As will be seen later, the effect on the variance of the regression coefficients is negligible.

[4] Anderson, R. L., "The Problem of Autocorrelation in Regression Analysis," American Statistical Association. Journal, v. 49, p. 113-129, March 1954

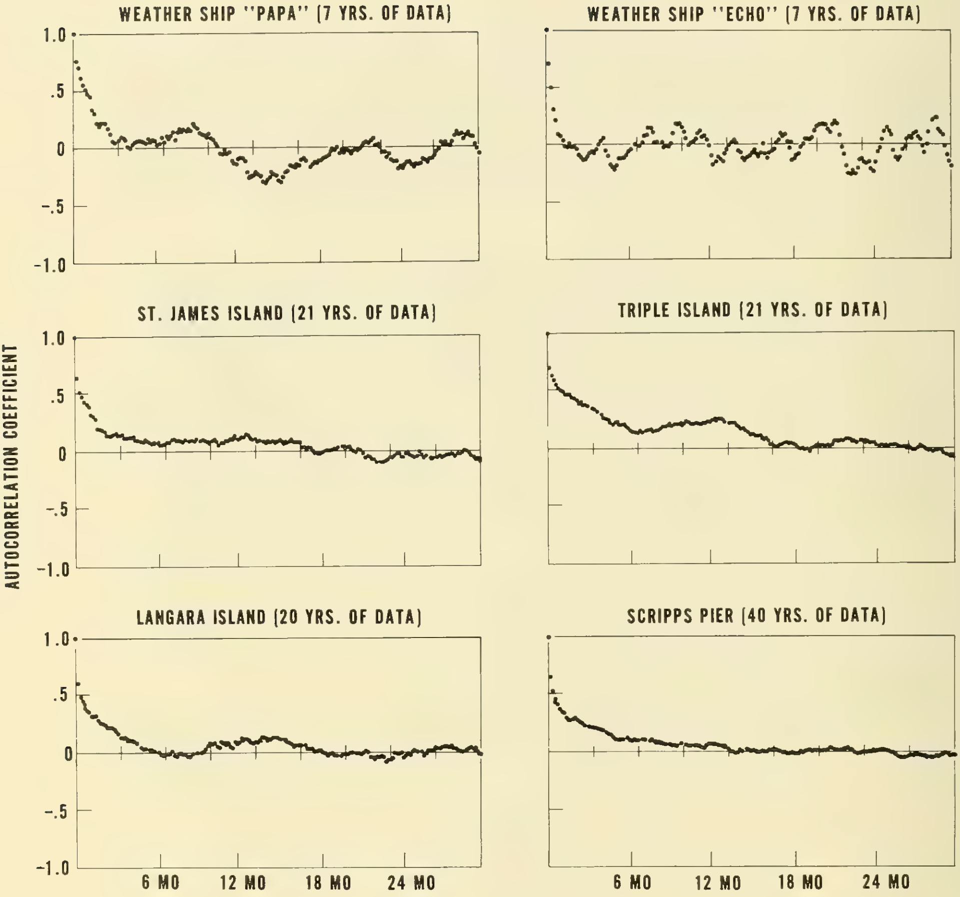

Figure 3. Autocorrelation coefficients for residual time series.

If equation (3) is fitted to individual years of data, the analysis provides in β_o an estimate of the yearly average of the surface temperature. The sequence of β_o's can be examined for the existence of trend in the time series. Testing for trend using either the theory of runs or the autocorrelation coefficient with lag unity indicates that no long term trend exists in any of the time series under consideration. Details of the trend analysis will be presented in a subsequent report.

MISSING DATA

The six locations have data missing in amounts varying from 2 percent to 40 percent of the number of possible observations. Intuitively it would seem that, for the types of analyses attempted, a fairly large fraction of randomly distributed missing data can be tolerated. It is the purpose of this report to examine quantitatively the effect of various fractions of missing data.

Although the expression *missing data* has been used thus far in the discussion, it is worthwhile now to comment on this usage. Conceivably, in a statistical problem, missing data can result in nothing more drastic than a sample of smaller size than planned. This might well be the case in a regression analysis in which the residuals are independently distributed with equal variances, and the missing data are uniformly or randomly distributed throughout the ranges of the independent variables. On the other hand, missing data in an extreme case can invalidate an experiment.

Table 1 shows that the stations with the largest fractions of missing data are the weather ships, partly because of the exclusion of temperatures if the ships are off station. Data may also be missing at either weather ships or shore locations because of bad weather or equipment failure. As an indication of the nature of occurrence of the missing data for stations with fairly large fractions of missing data, consider figure 4. Shown are histograms of the frequency of missing temperature sequences for 7 years of PAPA data and 21 years of St. James Island data. Station PAPA was selected from the two open ocean locations since the bathythermograph observations were made by oceanographers and were considered to be more accurate than for station ECHO. St. James Island was chosen from among the island and coastal locations since it had the largest fraction of missing data of these stations.

Except for a few long periods of missing data for each station, as indicated in figure 4, the missing data days are distributed very much as though at random. That is, given the appropriate probability of there being data on a day, the distributions by length of data-present sequences and data-missing sequences are like those expected. More specifically, the computed histograms shown in figure 4 result from randomly generated time series with two controlling conditional probabilities. The first conditional probability used for figure 4A is 0.76, which is the probability that a temperature will be observed, given that a temperature was observed the previous day. The second conditional probability used is 0.51, which is the probability that a temperature will be observed, given that a temperature was not observed the previous day. The corresponding probabilities for figure 4B are 0.89 and 0.59. These conditional probabilities agree quite well with the physical situation that missing data sequences occur infrequently, but once they occur they persist longer than can be explained by a single probability.

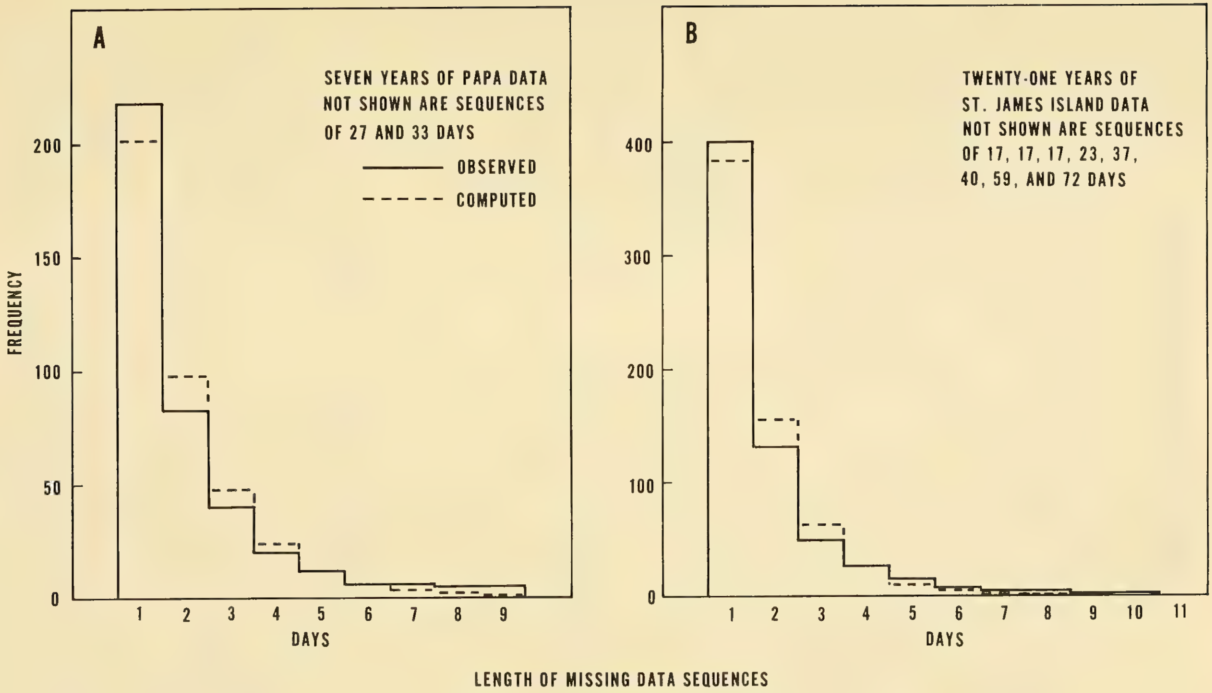

Figure 4. Histograms of the frequency of missing temperature sequences.

MODEL FOR MISSING DATA

It is proposed that the effect of missing data be evaluated in the following manner. There exist series of sea surface temperatures for which there are no missing data (Scripps Pier), or almost none (Triple Island), over periods of several years. Complete series of length up to 12 years can be selected from each of these sources. The few missing temperatures for Triple Island are filled in by adding to interpolated values random normal deviates having the appropriate variance. The complete series remains unchanged thereafter. It is thought necessary to consider two stations, whose time series of sea-surface temperatures have slightly different characteristics with respect to residual variability and significance of semiannual oscillatory terms, in order to avoid decisions which might be too dependent on the characteristics of a single station.

Regression and autocorrelation analyses are performed on the complete series. Estimates of the variances of the regression coefficients β are available from the matrix inverse to that of the coefficients in the normal equations of the least squares analysis. The estimates of the variances used assume independent, equal variance residuals. These variances are attributable to the residual variability of observations about the true regression curve.

The 40 years of Scripps Pier residuals with very few missing observations provide an estimate of the variance of the near-zero autocorrelation coefficients. The autocorrelation function for Scripps Pier was computed out to a lag of 1800 days, an arbitrary figure slightly over 10 percent of the total sample length. The standard deviation of the autocorrelation coefficients with lags from 400 days (end of the initial decay of the function) to 1800 days is $\sigma_c = 0.293$. This estimate of σ_c is considered to be the best available measure of the random variability of the near-zero autocorrelation coefficients of sea-surface temperature anomalies. It is based on a large sample of the autocorrelation coefficient, and the maximum lag involved is still only a small fraction of the total time series length.

Missing data days are randomly introduced into a complete time series using computer-generated, uniformly distributed random numbers. Any desired fraction of missing data can be introduced by associating with each daily temperature one of the random numbers with range 0 to 1. If the random number has a value greater than the desired fraction, the temperature is retained; if not, the temperature is deleted. Although two probabilities are used to generate each of the computed histograms of figure 4, it is more convenient in the analyses below to use single probabilities yielding the same fractions of missing data. The resulting computed histograms decrease more rapidly as a function of length of sequence than do those of figure 4, but the analysis used is fairly insensitive to the shape of the histograms. Since the gross characteristics of the time series are similar for all stations, any deletion of temperatures from complete Scripps Pier and Triple Island time series yields sample time series which are like those with naturally missing larger fractions of data, and which have whatever weaknesses are implied by the missing data. In the remainder of this paper, the name *sample time series* refers to a *complete time series* with data deleted by the above described process.

MONTE CARLO APPROACH

For a sample time series, the harmonic and autocorrelation analyses can be performed just as for a complete series, the proper adjustments being made in the computations. The regression and autocorrelation coefficients obtained from a sample time series are different from those obtained from the corresponding complete series. If many sample time series with the same fraction of missing data are independently generated from the same complete time series, and if regression and autocorrelation analyses are performed for each sample time series, then the variabilities in the resulting coefficients will measure the effect of missing data. The generation of many such time series to give estimates and confidence limits for parameters is an example of the technique which has been given the name Monte Carlo.

The major interest in the β's as statistical variables is the variability from sample to sample of their deviations from some true, but unknown, values. For an integral number of years of complete time series, the four estimates of the variances of β_1, β_2, β_3, and β_4, as obtained from the inverse matrix, are equal. Figures 5A, 5B, 6A, and 6B are histograms of the differences between the β's of 120 independently generated sample time series and the corresponding β's of the complete time series for 7 years of Scripps Pier and Triple Island data. The β's are uncorrelated and have equal variances. Because of the effective increase in sample size, the differences have been grouped for the four β's. For figures 5A and 6A the sample time series average 50 percent missing data; for figures 5B and 6B they average 20 percent missing data. The histograms are presented to demonstrate that the differences are symmetrically distributed about zero, and to show the dispersion.

Figure 7 is a plot of β_2 vs β_1 for a sample time series. It is included to demonstrate that the β's are uncorrelated. Table 2 presents the correlation coefficients R for all combinations of β's for the 7-year sample time series with 50 percent of the data missing. The 5 percent critical value of R is 0.179. One of the twelve R's barely exceeds this value. This not unlikely event has prior probability 0.34.

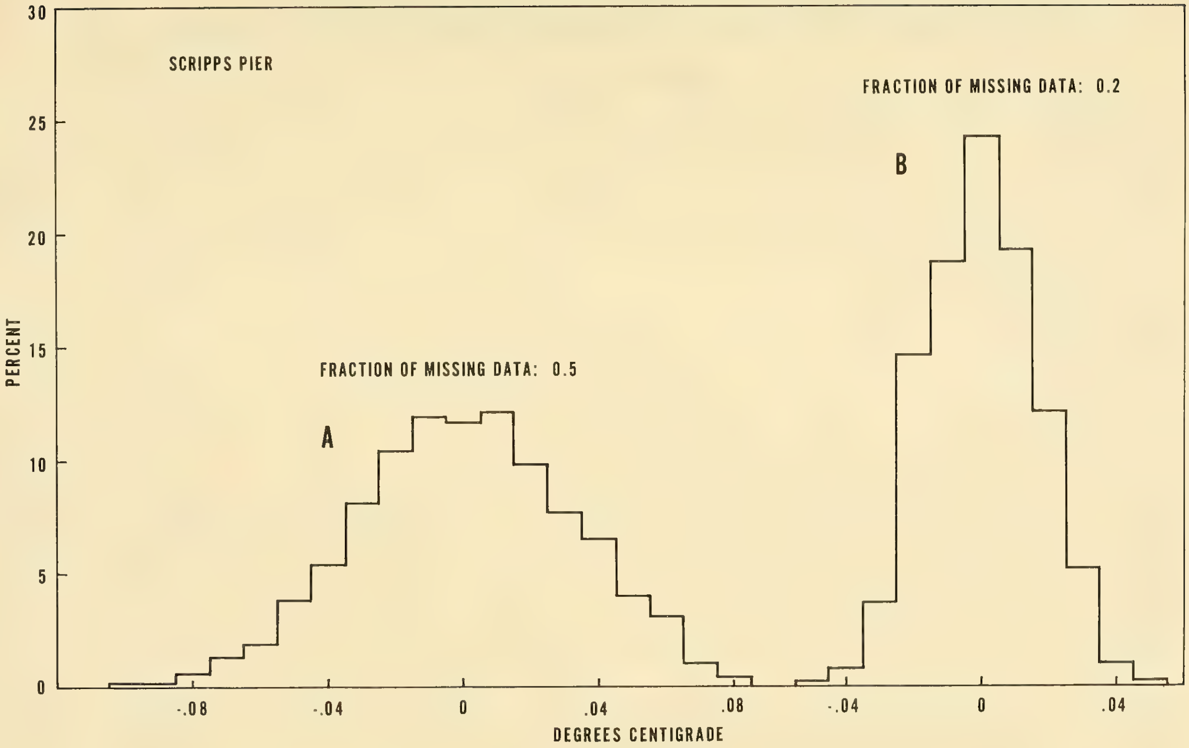

Figure 5. Histograms of differences between regression coefficients of sample time series and complete time series. Scripps Pier.

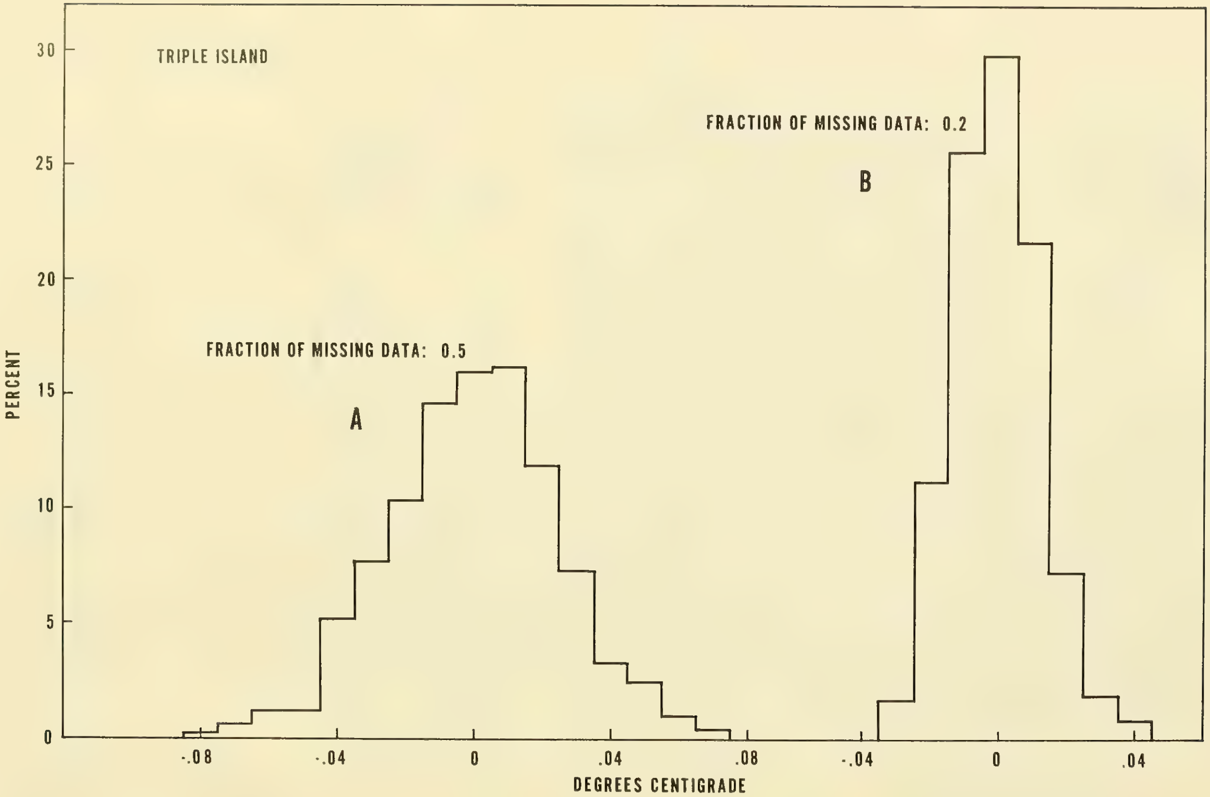

Figure 6. Histograms of differences between regression coefficients of sample time series and complete time series. Triple Island.

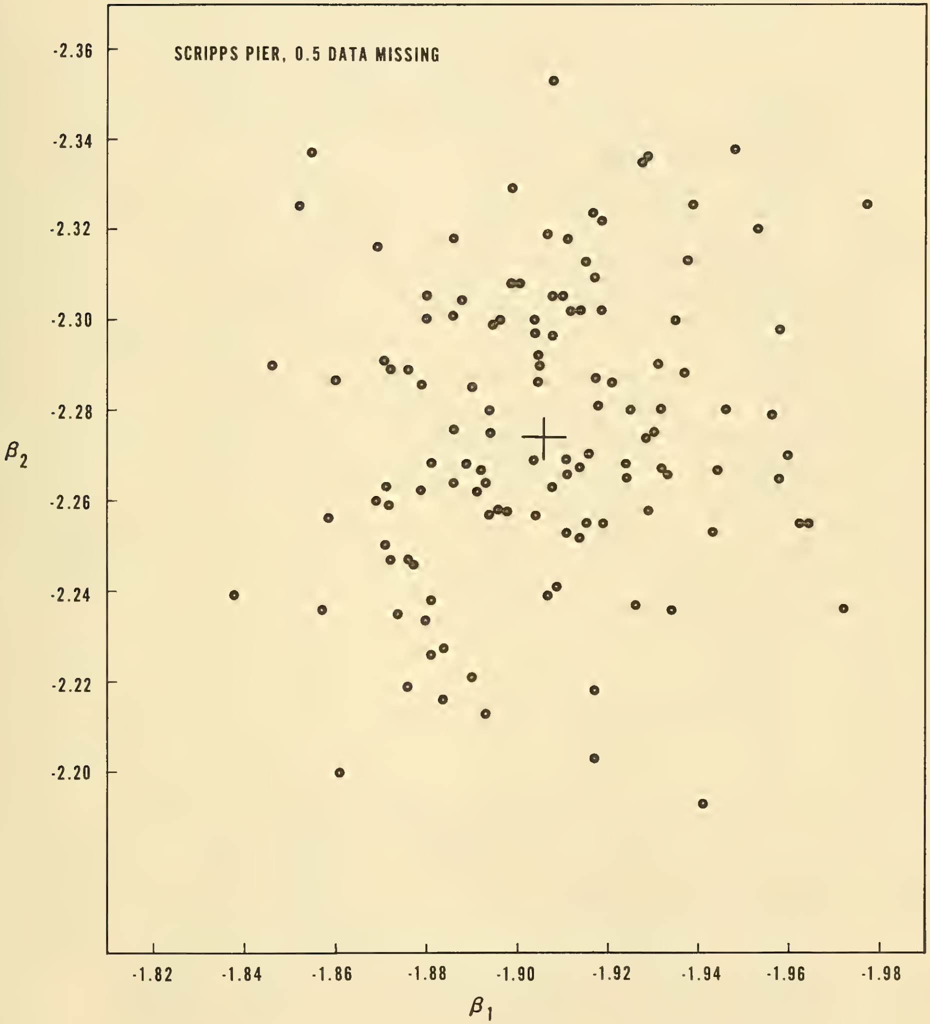

Figure 7. Plot of regression coefficients β_1 versus β_2 for sample time series. Scripps Pier, 0.5 data missing.

TABLE 2. CORRELATION COEFFICIENTS BETWEEN SAMPLE β'S, 50 PERCENT OF DATA MISSING

Variables	Correlation Coefficients	
	Scripps Pier	Triple Island
β_1 , β_2	0.139	0.187
β_1 , β_3	-0.151	0.023
β_1 , β_4	-0.117	-0.043
β_2 , β_3	0.111	0.061
β_2 , β_4	-0.148	-0.075
β_3 , β_4	-0.155	0.026

The Monte Carlo technique has been applied to data for the combinations of two stations, Scripps Pier and Triple Island; for three lengths of series, 4, 7, and 12 years; and for fractions f of data missing in the range 0.05 to 0.70. With respect to regression coefficients, figures 8A, 8B, and 8C display the results of these analyses in a normalized form. The quantity plotted is the ratio Q of the variances of the β's attributable to missing data to the variances of the β's attributable to residual variability. This ratio is the fractional increase in the variance of the β's attributable to missing data. Each point is based on 120 sample time series.

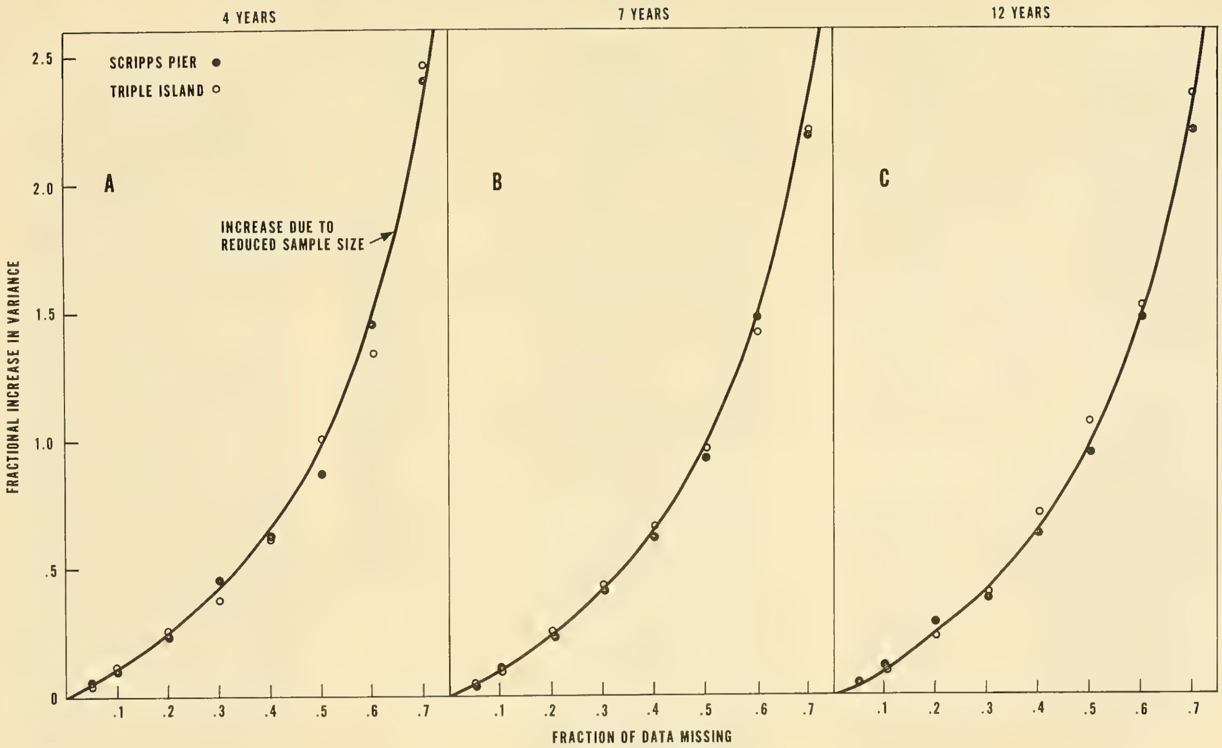

Figure 8. Fractional increase in variance of regression coefficients due to random missing data.

The ratio attributable to reduced sample size is shown by the continuous curve. The variances of β's based on samples from the same population are inversely proportional to sample size. If N is the sample size for the complete time series, $N(1 - f)$ is the size of the sample time series. For the β's

$$\sigma_s^2/\sigma_c^2 = 1/(1 - f)$$

where the subscripts s and c refer to sample and complete time series, respectively. The increase in variance attributable to reduced sample size is

$$Q = 1/(1 - f) - 1 = f/(1 - f)$$

The variances of the β's for the complete time series are computed as though the residuals are independent. The variances for the sample time series reflect the influence of the autocorrelated residuals. The empirical ratios of figure 8 lie almost on the theoretical curve, which assumes independent residuals. Thus, it is concluded that the combination of autocorrelated residuals and random deletion of data yields regression coefficients whose variances are as expected simply on the basis of sample size.

NONRANDOM MISSING DATA

As indicated in figure 4, there is an excessive number of longer sequences of missing data days. These sequences occur in the poor weather months October to March, inclusive. In addition there are several sequences of 5 to 10 days each, which are in excess of the number of such sequences expected by chance. It has been demonstrated above that randomly distributed missing data affect the variance of the regression coefficients just as though the sample size were smaller. It is necessary to determine if the longer sequences lead to the same result.

To approximate the time series yielding figure 4, sample time series have been generated in which longer sequences of data have been deleted in a random manner during the poor weather months. Then, individual temperatures are deleted at random from the remaining days until certain arbitrary fractions of missing data are obtained. Table 3 contains the number of longer sequences deleted for three series lengths and for three fractions of missing data. Analysis for 4-year series length was not attempted for the smallest missing fraction. The Monte Carlo technique is applied using 120 independently generated sample time series for each station and each combination of series length and fraction deleted.

TABLE 3. NUMBER OF LONGER SEQUENCES DELETED

Total Fraction Missing	Period Length (days)	Series Length		
		4 Years	7 Years	12 Years
0.16	28		1	2
	6		4	5
0.34	28	1	2	3
	6	5	7	14
0.51	28	2	3	5
	6	5	11	19

The normalized results are displayed in figure 9. The same theoretical curve has been plotted as in figure 8. The arbitrary dashed curve has twice the ordinate of the solid curve. Because of the much longer sequences deleted, and perhaps because of compromises necessary in constructing table 3, the scatter of points in figure 9 is greater than in figure 8.

The dashed curve has been fitted conservatively. It indicates that the fractional increase in the variance of the

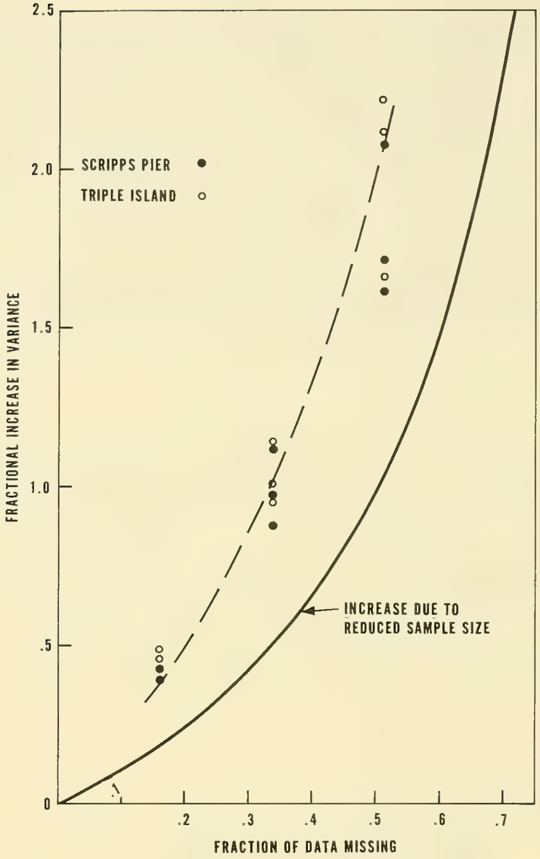

Figure 9. Fractional increase in variance of regression coefficients due to nonrandom missing data.

β's attributable to nonrandom missing data is twice the increase attributable to random missing data. This suggests caution in estimating the variances of β's, or of residuals, for time series when the missing data occurs in sequences longer than those occurring by chance.

The dashed curve can be interpreted another way. For fractions of missing data greater than 0.2, the dashed curve lies about 0.15 unit to the left of the continuous curve. When applicable, this quantity 0.15 should be added to the actual fraction of missing data. Then conclusions about nonrandom missing data can be made as for random missing data, but with the larger fraction of missing data used.

THE AUTOCORRELATION COEFFICIENT

The effect of missing data on autocorrelation coefficients will now be considered. The results are perhaps not as straightforward to evaluate as for regression coefficients, but are more encouraging as far as tolerating nonrandom missing data. Figure 10 presents the results of Monte Carlo analyses of autocorrelation coefficients similar to those for regression coefficients. The autocorrelation coefficients are for the time series of residuals remaining after the regression analyses have been performed. The same combinations of stations, series length, and fractions deleted are used. The variances of autocorrelation coefficients are averaged for lags from 10 to 100 in steps of 10. Assuming the variances are inversely proportional to series length, the average variances are normalized to an arbitrary series length of one year. In figure 10, results for random missing data are plotted as circles; results for nonrandom missing data are plotted as triangles.

The variance of near-zero autocorrelation coefficients based on 40 years of Scripps Pier data is 0.000859. Normalized to the series length of one year, the variance is 0.0344. This quantity is plotted as the dashed line at the top of figure 10.

Somewhat arbitrary curves have been fitted to the two sets of points. The effect of missing data on the variance of autocorrelation coefficients results in

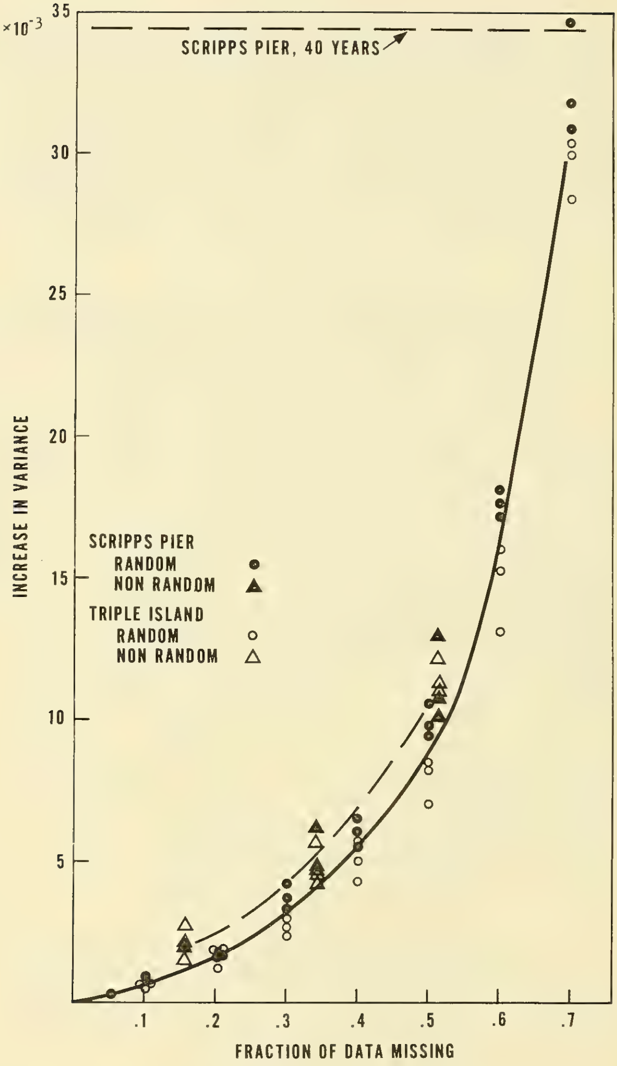

Figure 10. Increase in variance of autocorrelation coefficients due to random and nonrandom missing data.

curves similar to those for regression coefficients. The major difference is that the magnitude of the effect of introducing nonrandom missing data is much less in the case of autocorrelation coefficients. The ratio of ordinates averages about 1.2 instead of the 2.0 for the regression coefficient case. The dashed curve is about 0.05 unit to the left of the continuous curve rather than 0.15 unit. If the analyses of regression and autocorrelation coefficients are of equal importance, then the limitations on nonrandom missing data are determined by the regression coefficient results above.

A comparison of the variance determined from the 40 years of Scripps Pier, near-zero autocorrelation coefficients with the variance of the Monte Carlo analysis indicates that 70 percent of the data may be randomly missing before the two variances are equal.

COMMENTS AND CONCLUSIONS

In the analysis above, certain compromises are made with computer techniques and computing times required: (1) The distribution of missing day sequences based on a simple use of random numbers will never agree exactly with the observed distribution of missing day sequences for a given station. Nevertheless, the techniques used provide good initial estimates of the effect of missing data. (2) The use of 120 Monte Carlo runs per case is a compromise between computer time required and the apparent rate of convergence to a limit of the parameters estimated.

It is concluded that random missing data in a time series result in regression coefficients whose variances increase over those of a complete time series by an amount as predicted by the reduction in sample size. However, the presence of longer sequences of nonrandom missing data may have a pronounced effect in estimating regression coefficients. Specifically, if variances of regression coefficients are estimated in the usual manner, on the average these estimates must be adjusted upwards. Roughly, the increase in variance due to missing data must be doubled. Alternatively, for fractions of missing data greater than 0.2, time series with nonrandom missing data will have regression coefficient variances equal to those the same series with 0.15 more missing data would have, if all the missing data were random.

The effect of nonrandom missing data on autocorrelation coefficients is less pronounced. The increase in their estimated variance need be only 20 percent. Alternatively, for fractions of missing data greater than 0.2, time series with nonrandom missing data will have autocorrelation coefficient variances equal to those the same series with only 0.05 more missing data would have, if all the missing data were random.

RECOMMENDATIONS

Almost all time series of sea-surface temperatures contain missing data. The nature of this missing data as to randomness of occurrence in time should be examined before regression and autocorrelation analyses are performed. The appropriate results of this report should be applied in estimating the variances of regression and autocorrelation coefficients.

A similar investigation of the effect of missing data should be performed for the regression problem with several independent variables, namely time, depth and geographical location. The dependent variable will be water temperature.

The results of this report apply to the long range estimation of sea-surface temperatures. An examination should be made of the effect of missing data on the short range (a few weeks or months) prediction of sea-surface temperatures.

Navy Electronics Lab., San Diego, Calif.
Report 1256

SEA-SURFACE TEMPERATURE ESTIMATION, by C. J. Van Vliet, 32 p. 5 Jan 65.

UNCLASSIFIED

Many time series of sea-surface temperatures feature random missing data or long periods of missing data. Methods are presented that enable one to correct for the effects of the missing data in computations of regression and autocorrelation coefficients.

1. Oceanographical data - Statistical analysis
2. Statistical analysis - Applications

I. Van Vliet, C. J.

SR 004 03 01, Task 0586
(NEL L40551, formerly L4-5)

This card is UNCLASSIFIED

Navy Electronics Lab., San Diego, Calif.
Report 1256

SEA-SURFACE TEMPERATURE ESTIMATION, by C. J. Van Vliet, 32 p. 5 Jan 65.

UNCLASSIFIED

Many time series of sea-surface temperatures feature random missing data or long periods of missing data. Methods are presented that enable one to correct for the effects of the missing data in computations of regression and autocorrelation coefficients.

1. Oceanographical data - Statistical analysis
2. Statistical analysis - Applications

I. Van Vliet, C. J.

SR 004 03 01, Task 0586
(NEL L40551, formerly L4-5)

This card is UNCLASSIFIED

Navy Electronics Lab., San Diego, Calif.
Report 1256

SEA-SURFACE TEMPERATURE ESTIMATION, by C. J. Van Vliet, 32 p. 5 Jan 65.

UNCLASSIFIED

Many time series of sea-surface temperatures feature random missing data or long periods of missing data. Methods are presented that enable one to correct for the effects of the missing data in computations of regression and autocorrelation coefficients.

1. Oceanographical data - Statistical analysis
2. Statistical analysis - Applications

I. Van Vliet, C. J.

SR 004 03 01, Task 0586
(NEL L40551, formerly L4-5)

This card is UNCLASSIFIED

Navy Electronics Lab., San Diego, Calif.
Report 1256

SEA-SURFACE TEMPERATURE ESTIMATION, by C. J. Van Vliet, 32 p. 5 Jan 65.

UNCLASSIFIED

Many time series of sea-surface temperatures feature random missing data or long periods of missing data. Methods are presented that enable one to correct for the effects of the missing data in computations of regression and autocorrelation coefficients.

1. Oceanographical data - Statistical analysis
2. Statistical analysis - Applications

I. Van Vliet, C. J.

SR 004 03 01, Task 0586
(NEL L40551, formerly L4-5)

This card is UNCLASSIFIED

INITIAL DISTRIBUTION LIST

CHIEF, BUREAU OF SHIPS
CODE 210L CODE 345B
CODE 240C (2)
CODE 320 CODE 334
CODE 360 (3)
CODE 370
CHIEF, BUREAU OF NAVAL WEAPONS
DLI-3
DLI-31
FASS
RU-222
RUDC-2
RUDC-11
CHIEF, BUREAU OF YARDS AND DOCKS
CHIEF OF NAVAL PERSONNEL
PERS 11B
CHIEF OF NAVAL OPERATIONS
OP-07T OP-716C
OP-71
OP-76C
OP-03EG
OP-09B5
CHIEF OF NAVAL RESEARCH
CODE 416
CODE 466
CODE 468
COMMANDER IN CHIEF US PACIFIC FLEET
COMMANDER IN CHIEF US ATLANTIC FLEET
COMMANDER OPERATIONAL TEST AND
EVALUATION FORCE
DEPUTY COMMANDER OPERATIONAL TEST -
EVALUATION FORCE, PACIFIC
COMMANDER CRUISER-DESTROYER FORCE,
US ATLANTIC FLEET
US PACIFIC FLEET
COMMANDER TRAINING COMMAND
US PACIFIC FLEET
COMMANDER SUBMARINE DEVELOPMENT
GROUP TWO
FLEET AIR WINGS, ATLANTIC FLEET
SCIENTIFIC ADVISORY TEAM
US NAVAL AIR DEVELOPMENT CENTER
NADC LIBRARY
US NAVAL MISSILE CENTER
TECH. LIBRARY, CODE NO 3022
PACIFIC MISSILE RANGE /CODE 3250/
US NAVAL ORDNANCE LABORATORY
LIBRARY
US NAVAL ORDNANCE TEST STATION
PASADENA ANNEX LIBRARY
CHINA LAKE
US NAVAL WEAPONS LABORATORY
KXL
PUGET SOUND NAVAL SHIPYARD
USN RADIOLOGICAL DEFENSE LABORATORY
DAVID TAYLOR MODEL BASIN
APPLIED MATHEMATICS LABORATORY
/LIBRARY/
US NAVY MINE DEFENSE LABORATORY
US NAVAL TRAINING DEVICE CENTER
CODE 365H, ASW DIVISION
USN UNDERWATER SOUND LABORATORY
LIBRARY (3)
ATLANTIC FLEET ASW TACTICAL SCHOOL
USN MARINE ENGINEERING LABORATORY
US NAVAL CIVIL ENGINEERING LAB.
L54
US NAVAL RESEARCH LABORATORY
CODE 2027
US NAVAL ORDNANCE LABORATORY
CORONA
USN UNDERWATER SOUND REFERENCE LAB.
BEACH JUMPER UNIT TWO
US FLEET ASW SCHOOL
US FLEET SONAR SCHOOL
USN UNDERWATER ORDNANCE STATION
OFFICE OF NAVAL RESEARCH
PASADENA
USN WEATHER RESEARCH FACILITY
US NAVAL OCEANOGRAPHIC OFFICE (2)
US NAVAL POSTGRADUATE SCHOOL
LIBRARY (2)
DEPT. OF ENVIRONMENTAL SCIENCES
OFFICE OF NAVAL RESEARCH
LONDON
BOSTON
CHICAGO
SAN FRANCISCO

FLEET NUMERICAL WEATHER FACILITY
US NAVAL ACADEMY
ASSISTANT SECRETARY OF THE NAVY R-D
ONR SCIENTIFIC LIAISON OFFICER
WOODS HOLE OCEANOGRAPHIC INSTITUTION
INSTITUTE OF NAVAL STUDIES
LIBRARY
AIR DEVELOPMENT SQUADRON ONE /VX-1/
DEFENSE DOCUMENTATION CENTER (20)
DOD RESEARCH AND ENGINEERING
TECHNICAL LIBRARY
NATIONAL OCEANOGRAPHIC DATA CENTER (2)
NASA
LANGLEY RESEARCH CENTER
COMMITTEE ON UNDERSEA WARFARE
US COAST GUARD
OCEANOGRAPHY - METEOROLOGY BRANCH
ARCTIC RESEARCH LABORATORY
WOODS HOLE OCEANOGRAPHIC INSTITUTION
US COAST AND GEODETIC SURVEY
MARINE DATA DIVISION /ATTN-22/
US WEATHER BUREAU
US GEOLOGICAL SURVEY LIBRARY
DENVER SECTION
US BUREAU OF COMMERCIAL FISHERIES
LA JOLLA DR. AHLSTROM
WASHINGTON 25, D. C.
POINT LOMA STATION
WOODS HOLE, MASSACHUSETTS
HONOLULU-JOHN C MARR
LA JOLLA, CALIFORNIA
HONOLULU, HAWAII
STANFORD, CALIFORNIA
POINT LOMA STA-J. H. JOHNSON
ABERDEEN PROVING GROUND, MARYLAND
REDSTONE SCIENTIFIC INFORMATION
CENTER
BEACH EROSION BOARD
CORPS OF ENGINEERS, US ARMY
DEPUTY CHIEF OF STAFF, US AIR FORCE
AFRST-SC
STRATEGIC AIR COMMAND
HQ AIR WEATHER SERVICE
UNIVERSITY OF MIAMI
THE MARINE LAB. LIBRARY (3)
COLUMBIA UNIVERSITY
HUDSON LABORATORIES
LAMONT GEOLOGICAL OBSERVATORY
DARTMOUTH COLLEGE
THAYER SCHOOL OF ENGINEERING
RADIOPHYSICS LABORATORY
RUTGERS UNIVERSITY
CORNELL UNIVERSITY
OREGON STATE UNIVERSITY
DEPARTMENT OF OCEANOGRAPHY
UNIVERSITY OF SOUTHERN CALIFORNIA
ALLAN HANCOCK FOUNDATION
UNIVERSITY OF WASHINGTON
DEPARTMENT OF OCEANOGRAPHY
FISHERIES-OCEANOGRAPHY LIBRARY
NEW YORK UNIVERSITY
DEPT OF METEOROLOGY - OCEANOGRAPHY
UNIVERSITY OF MICHIGAN
DR. JOHN C. AYERS
UNIVERSITY OF ALASKA
GEOPHYSICAL INSTITUTE
UNIVERSITY OF RHODE ISLAND
NARRAGANSETT MARINE LABORATORY
YALE UNIVERSITY
BINGHAM OCEANOGRAPHIC LABORATORY
FLORIDA STATE UNIVERSITY
OCEANOGRAPHIC INSTITUTE
UNIVERSITY OF HAWAII
HAWAII INSTITUTE OF GEOPHYSICS
ELECTRICAL ENGINEERING DEPT
A-M COLLEGE OF TEXAS
DEPARTMENT OF OCEANOGRAPHY
THE UNIVERSITY OF TEXAS
DEFENSE RESEARCH LABORATORY
HARVARD UNIVERSITY
SCRIPPS INSTITUTION OF OCEANOGRAPHY (4)
MARINE PHYSICAL LAB
UNIVERSITY OF CALIFORNIA
ENGINEERING DEPARTMENT
UNIVERSITY OF CALIFORNIA, SAN DIEGO
SIO
THE JOHNS HOPKINS UNIVERSITY
APPLIED PHYSICS LABORATORY
INSTITUTE FOR DEFENSE ANALYSIS