STATISTICAL PREDICTABILITY OF

PACIFIC OCEAN SURFACE TEMPERATURE ANOMALIES

by

Clare Bertram Billing, Jr.

B.S., Massachusetts Institute of Technology
(1973)

M.A., Harvard University
(1975)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF

MASTER OF SCIENCE

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September, 1979

Signature of Author .....
Department of Meteorology, Sept. 1979

Certified by .....
Thesis Supervisor

Accepted by .....
Chairman, Departmental Committee on Graduate Students

STATISTICAL PREDICTABILITY OF

PACIFIC OCEAN SURFACE TEMPERATURE ANOMALIES

by

Clare Bertram Billing, Jr.

Submitted to the Department of Meteorology
in August, 1979 in partial fulfillment of the requirements
for the Degree of Master of Science.

ABSTRACT

Nonseasonal variability of the sea surface temperatures (SST) of the Equatorial and North Pacific Ocean is examined. The data set is derived from ship observations and covers 250 grid squares for the period 1949-March 1979 in monthly averages. The objective is examination of the basic time and space scales of the SST anomaly variations and examination of their predictability from the SST fields themselves.

An empirical orthogonal function (EOF) analysis of the anomaly fields shows a complex spatial structure with the variance spread over several modes. The first few modes are associated with physical and data set features of the SST variability. The time amplitudes associated with the primary modes show that the SST time variations exhibit first order autoregressive behavior with some higher order components and seasonal nonstationarity.

Linear statistical modeling of the EOF time series is used to evaluate predictability of the SST anomaly fields. Results show highly significant predictability at short lead times due to persistence in the anomaly fields and a much lower degree at 6 to 12 months lead time due to lag correlations of the primary EOF modes.

Thesis Supervisor : Reginald E. Newell
Title : Professor of Meteorology

ACKNOWLEDGEMENTS

I would like to thank Professor Reginald Newell and Dr. Alfredo Navato for their help and encouragement during the course of this work. Much of the initial processing and error screening of the data set was done by Dr. Navato. Isabelle Kole and Beverly Colby were invaluable in their drafting the figures and typing the text, respectively. I'm particularly grateful to Natalie for her constant moral support and to Jeffrey for his pleasant distractions.

This work and the author's Research Assistantship were supported by a grant from the Department of Energy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

The oceans, comprising over seventy percent of the earth's surface, receive and store great quantities of heat from the incoming solar radiation. Since the distribution of this heat and its release into the atmosphere is the major driving force of the general circulation, it should be expected that the properties of the ocean, especially as expressed in surface temperatures, play an important role in the weather and climate variations. The greater density and heat capacity of the oceans give their thermal and dynamic variations long natural time scales compared to that of the atmosphere. This long thermal persistence suggests that the oceanic temperature fluctuations, through surface heat transfer, produce a substantial effect on the large-scale atmospheric temperature and flow fields, thus influencing the short-term climate.

In addition to the direct thermal effect of the ocean on the atmosphere, the atmospheric variations affect the oceans. Through momentum transfer at the sea surface the winds drive the oceanic circulation thereby bringing about a redistribution of the thermal field. The speed and direction of the winds also influence the heat transfer between the air and sea. Thus, there exists a complex air-sea interaction process with multiple feedbacks governing the behavior of the oceans and atmosphere at large time and space scales. It is therefore essential for the understanding of and for the long-range forecasting of short-term climatic fluctuations to consider the interaction of the atmospheric fields with the oceans' surface temperature fields.

Because of this belief that the oceanic motions and thermal fields

strongly influence the weather and climate, it is hoped that sea surface temperatures (SST) and their fluctuations may be used to predict atmospheric flow fields and climate changes several months to years in advance. The exact nature, however, of the SST's influence is not understood. On the one hand, the ocean may be viewed as a low pass filter of the high frequency and relatively random atmospheric fluctuations. This describes the sea as having a stabilizing influence on the climate through thermal negative feedback. Alternatively, the ocean, through its thermal coupling to the atmosphere, may act as an initiator of short-term climate change. Since a combination of these two ideas is probably the correct view, understanding of both the seasonal and anomalous variations of SST fields should yield a better understanding of and to some degree predictions of atmospheric variations. Prediction of the SST fields themselves, if made possible from the increased understanding of their fluctuations and interaction with the atmosphere, will allow more effective predictions of the climate changes at greater lead times.

There are two contrasting but complimentary approaches in attempting to characterize and understand the variation of a naturally occuring system such as SST anomaly fields. The dynamical or deterministic method attempts to describe and predict the behavior of the system using physical laws which govern it. However, when these laws are sufficiently complex or not completely understood, the statistical approach is used to observe the system's behavior in time and attempt to identify its basic patterns empirically. Certainly the statistical behavior should reflect the physics and lead to its better understanding. In the present study the physical laws governing the SST fields and their interaction with the atmosphere are sufficiently complex and unformulated that an empirical approach is

appropriate.

The goal of this study is to characterize the major time and space features of monthly mean Pacific Ocean surface temperature anomaly fields, and examine their predictability. Although the important step of understanding the seasonal variations is not addressed, and their interaction with the nonseasonal anomalies in studying predictability is not considered, the approach here is felt to be a meaningful step in understanding the SST fluctuations and their role in large-scale ocean-atmosphere interaction. Specifically, the goal of this study is to:

(1) Present an analysis of the monthly mean SST fields for the period 1949 to 1976 for a large part of Pacific Ocean.

(2) Present a nonseasonal empirical orthogonal function analysis of the Pacific SST anomaly fields, and relate the primary empirical modes of variation to the system's physical behavior.

(3) Examine the time scales that characterize the above primary modes of variation and look at their interrelationships using time series analysis techniques.

(4) Evaluate the predictability of the SST anomaly fields from the SST's themselves using the principle modes, their variation in time, and their interaction in linear time series and multiple regression models.

II. BACKGROUND

a. SST and Large-scale Air-sea Interaction

It has long been recognized that ocean surface temperatures strongly influence the climate of coastal regions through direct heat transfer with the overlying air. The mild seasons of the northwestern U.S. coast and the cool climate of the Peru coast are prime examples. This simple view has been expanded, beginning with the studies of Helland-Hansen and Nansen in the early 1920's, to a more complex air-sea interaction scheme where the mean patterns of the SST interact with and influence the atmospheric patterns. It was, however, not until after the large-scale warming of the equatorial Pacific and anomalous events in the eastern North Pacific during 1957-58 that great interest developed in SST anomalies and their possible influence on the global climate. Namias (1959), in the first of a continuing series of case studies suggesting a close interaction between the oceans and the atmosphere, drew attention to the possible influence of the Pacific Ocean SST's on the large-scale North American climate. He related the anomalous warming of the eastern North Pacific to prevailing anomalies in the overlying atmospheric circulation and concluded, as well, that the anomalous SST's strongly influenced the atmospheric flow. At this time Bjerknes (1960) also began his investigations into the interaction of ocean temperatures and the atmospheric circulation in producing and maintaining anomalous ocean/atmosphere events.

In subsequent papers Namias (1963, 1965, 1969, 1970, 1971, 1973, 1974, 1975) and Namias and Born (1970) continued to show empirically the importance of North Pacific air-sea interaction processes in deter-

mining the thermal and circulation patterns of the ocean and atmosphere on monthly, seasonal, and climatic time scales. The major results of these studies were a statistical/observational description of the SST field variations in the central and eastern North Pacific, the attributing of their extensive and persistent anomalies to interaction with abnormal wind and weather systems, and evidence that anomalous SST's, by their interaction with and influence on the atmospheric flow, help determine the downstream climate of parts of North America. These studies, as well as those by White and Barnett (1972), Davis (1976, 1978), Reiter (1978), and Barnett and Preisendorfer (1978) have attempted to use observations to explain the physics of the air-sea interactions and with some success to specify and predict atmospheric pressure or temperature distributions from the SST patterns. They have, however, been unable to completely specify the precise effects of each component (advection, upwelling, heat exchange, etc.) in the interaction.

Other studies have considered the role of the Atlantic Ocean SST's in air-sea interaction. Bjerknes (1960, 1964) related Atlantic Ocean temperatures to atmospheric circulation and climate, and Ratcliffe and Murray (1970) looked at the use of lag associations between North Atlantic SST's and European sea level pressure in long-range weather forecasting. Weare (1977) and Haworth (1978) presented a statistical analysis of Atlantic SST field variations and Hsuing (1978) described their time scale behavior and presented a simple prediction model.

Another segment of the literature deals with equatorial Pacific air-sea interaction and specifically with the relationships among the Pacific trade winds, equatorial SST's, and the Hadley cell and Walker

circulation systems. Bjerknes (1961, 1966, 1969, 1972, 1974), in a series of studies analogous to those of Namias, addressed the problem of interaction of anomalous SST fields in the equatorial Pacific with the atmosphere. He postulated that events of large-scale eastern equatorial Pacific warming (El Niño) such as occurred in 1957-58, 1963-64, 1965-66, and 1972 are part of a global air-sea interaction system and that the warmer water exerts a profound effect on mid-latitude atmospheric flow. These "teleconnections" from the equatorial Pacific would therefore partially determine the short-term climate in various parts of the world. He suggested a scheme where weakening of the Southeast Trades along the eastern equatorial Pacific would decrease upwelling along the equator leading to a large-scale increase of the SST's. This warmer equatorial water over a wide area would drive the Hadley circulation at an increased rate through latent heat release, which would increase the rate of northward absolute angular momentum transport affecting both extratropical circulation and equatorial trades. These effects would then feedback on the SST as well as influence the downstream climate over North America and Europe. This hypothesis has been tested with numerical model experiments by Rowntree (1972) and statistically by Chui and Low (1979) and Barnett (1977b) with nonconclusive and contradictory results. Studies by Namias (1976), Newell and Weare (1976b), and Hickey (1975) have also looked into the influence of equatorial Pacific SST's on the atmosphere flow and downstream climate.

Wyrtki (1973, 1975) has related theoretically and observationally the SST anomalies to variations in the trade winds, sea level and the ocean current dynamics in the equatorial Pacific. Quinn (1974, 1976) has further related the phenomenon of equatorial Pacific warming, El

Niño, to the global fluctuation first described by Walker, the Southern Oscillation. This association has been further investigated in modelling studies by Julian and Chervin (1978) and statistically by Wyrtki et al (1976), Barnett (1977a, b, 1978) and Billing and Newell (unpublished).

b. Modelling of the SST Anomaly Fields

The ocean-atmosphere feedback system consists of the ocean affecting the atmosphere by sensible and latent heat transfer, and the atmosphere driving the ocean via surface momentum transfer. The thermal structure of the ocean surface is determined by the combination of the influences of incident solar radiation, ocean dynamics and heat transfer, and the air-sea interaction processes. The time and space patterns of the SST fields are therefore controlled by energy exchange at the sea surface, heat storage in the surface layer, horizontal advection, vertical advection by upwelling or sinking, and mixing processes. Any analysis, modeling, or prediction of SST anomaly fields therefore requires an understanding of all these processes and their interaction. However, incomplete knowledge may be supplemented with statistical modeling in arriving at useful specification and prediction techniques.

Wyrtki and Haberland (1968) and Clark (1972) in their papers on the heat balance in the North Pacific Ocean give a good description of the processes controlling SST. The heat balance equation for a volume of water in the surface layer with unit surface area and depth is

$$\int_0^t \int_0^h (\rho C_p \frac{\partial T}{\partial t}) dz dt = \int_0^t \int_0^h Q_H dz dt + \int_0^t (Q+Q_v) dt \qquad (1)$$

where T is the water temperature, Q the total air-sea surface heat transfer, Q_H the horizontal advection of heat, and Q_v the vertical heat trans-

fer by advection and mixing. The differential form of this equation gives the variation of the surface temperature with time.

The total heat transfer at the surface may be written as the sum of the absorbed incident color radiation flux Q_i, upward flux of long-wave radiation Q_ℓ latent heat transfer Q_e, and sensible heat transfer Q_s.

$$Q = Q_i - (Q_\ell + Q_e + Q_s) \tag{2}$$

These are functions of time, oceanic cloud cover, surface wind speed, and the air-sea temperature and vapor pressure difference. The horizontal advection of heat Q_H is given by

$$Q_H = - \rho C_p(\overline{V} \cdot \overline{\nabla T}) \tag{3}$$

where ρ and C_p are sea water density and specific heat at constant pressure, $\overline{V}$ the horizontal ocean surface velocity, and $\overline{\nabla T}$ the horizontal temperature gradient at the ocean surface. The vertical heat transfer Q_v would be given by a combination of Ekman pumping driven by surface wind stress and a statistical parameterization of the mixing processes.

Wyrtki and Haberland (1968) attribute the North Pacific SST variability mostly to heat gained from the solar radiation and its distribution by horizontal advection from the eastern and southern parts of the subtropical anticyclonic gyre to the Northwest. Cool water is advected

towards the California coast. Clark (1972) and Favorite and McClain (1973) through empirical studies showed that Eastern North Pacific SST anomalies could be explained quite well by relating them to anomalous heat transfer across the air-sea interface and to advection by anomalous wind-driven ocean surface currents. The previously discussed papers on El Niño have expressed the importance of vertical advection by Ekman upwelling in SST variations of the equatorial regions, but Haney et al (1979) in a dynamical-numerical study find this of minimal importance in other regions. White and Barnett (1972), Hurlburt, et al (1976), McCreary (1976), and Huang (1978) have carried out numerical modeling studies of the SST anomalies in the North and Equatorial Pacific, and have demonstrated the importance of all the above mentioned processes.

Frankignoul and Hasselman (1977), using the concept of stochastic climate models developed by Hasselman (1976), investigated the low frequency variability of the upper ocean. They presented a simple model for SST anomalies of random short-timescale atmospheric forcing balanced by a stabilizing linear feedback. This is represented in the heat balance equation below. T is the temperature anomaly, f the atmospheric stochastic forcing function, and λ a constant feedback factor.

$$\frac{dT}{dt} = f(T,T_a,\ q,U,R) - \lambda T \qquad (4)$$

The random function f depends on latent and sensible heat flux, net radiative flux, and heat exchange from the layer below. Advective processes are ignored. T_a, q, and U are the temperature, relative humidity, and

wind speed in the surface atmosphere. R is the radiation component. This relation is that of a first order autoregressive process with negative feedback, requiring the constant λ to be positive.

The authors indicate that this model explains the major features of mid-latitude SST anomalies, which have a predominantly low frequency red spectral characteristic. They do however recognize that a more detailed analysis including cross correlations with the principle modes of atmospheric forcing and SST field variations, along with consideration of large-scale advection, local advection and upwelling, and mixed layer deepening by wind actions is necessary for an improved picture. This is particularly important as expressed by Reynolds (1978) in modeling those areas of the ocean near strong currents or fronts where ocean processes have equal or greater influence on heat balance than atmospheric forcing. Such areas as the Gulf Stream region or the Eastern Equatorial Pacific are prime examples.

III. METHODS OF ANALYSIS

a. The Data Set

The data used in this study are monthly mean sea surface temperatures averaged into grid squares over the Equatorial and North Pacific Ocean derived from summaries of the marine weather observations of merchant ships. The data and its analysis are similar to that described by Weare, et al (1976) with extensions in time period, grid density, and amount of the Pacific covered. The data covers the period January 1949 to March 1979 in 363 monthly means of 250 5° X 5° and 5° X 10° grid squares. Figure 1 shows the data coverage from the American coast to 125°W in longitude and between 60°N and 30°S in latitude.

For the period 1949-62 monthly analysis of SST by Sette, et al (1968) were used. These had been carefully corrected for individual ship biases, gross errors, and analysis errors. Grid point values from this analysis were obtained on magnetic tape from NCAR from which 5° X 5° latitude-longitude grid averages were constructed. The initial guess field for Sette's analysis was the long term monthly means averaged in 2° squares, and the method was a version of the Fleet Numerical Weather Facility's Scalar Analysis Program for Randomly Spaced Data. Consequently the final analysis retains the character of the initial guess field (climatology) in areas of sparse data. This is a problem especially in equatorial regions since the number of observations per month increased from 5000 to 15,000 over the period. For the period 1963-73 a tape of 5° grid averages provided by L. F. Eber of the Southwest Fisheries Center was used for the Pacific east of 180°. The West Pacific data for this period were obtained from an interpolation of 5° grid point

averages from the Ten Day Marine Reports summary maps of the Japanese Meteorological Service. Figure 2 shows the percent of months during this period having SST data from the above sources in each 5° X 5° square. The clear area of the central Pacific have greater than 90 percent, the cross-hatched squares greater than half, and the solid squares less than half of the 167 months with data. Since 1973 data was obtained from analyses in the monthly publication of the Southwest Fisheries Center.

The above data sets were combined to form a 368 5° X 5° grid point analysis of the Pacific SST for the period covered. This was contracted to the 250 point data set by elimination of some squares with little data and the combination of some adjacent ones. Missing values prior to 1963 were filled in by the original analysis, basically by climatology as we stated. After this time they were filled in by our data analysis with mostly interpolation and extrapolation in time, sometimes by zonal interpolation and extrapolation, and rarely by meridional. Problems with the data set, making rigorous statistical evaluations of SST variability not always consistent with the required assumptions, are a consequence. In addition to the nonstationarity of the means and variances created by the discontinuities in the data set analysis techniques, problems arise from the unreliability of merchant ship measurements and reports, and the fact that the ships follow the major shipping lanes and not an even or random distribution.

Long term monthly means and standard deviations were then calculated from these analyses for the period 1/49 to 11/76 for each 5° X 5° or 5° X 10° grid square, and the anomalies to the climatological means derived.

b. Empirical Orthogonal Function Analysis

An eigenfunction or empirical orthogonal function (EOF) analysis was performed on the set of Pacific SST anomaly field data. Since the application of eigenfunction decomposition to meteorological fields was first introduced by Lorenz (1956) and Sellers (1957), there have been numerous papers describing its use in oceanography and meteorology. Verstraete (1978) has reviewed very nicely the method and its applications. Derivations of the method are presented by Brier and Meltesen (1976) and Kutzbach (1967) as well as in the above mentioned papers. Barnett (1977a, 1978), Davis (1976, 1978), Barnett and Hasselman (1979), Weare, et al (1976), and Hsiung (1978) have applied EOF analysis to their data in a manner similar to that which has been done here. In this study an EOF representation of the SST anomaly fields provide the following:

(1) An efficient and objective representation of the variance in the SST anomaly field, with information on its distribution in space and time. The resulting patterns of the primary modes of variation may also give some physical insight into the mechanisms which drive the SST variations, although this is not guaranteed.

(2) A method of filtering errors or unwanted variation from the representation of the SST field. This is accomplished by retaining only those modes, when reconstructing the field, that are of interest to the purpose of the study or that are felt to contain significant contribution to the field (not random noise).

(3) The most efficient, in the least squares sense, set of predictors for variations in the SST field. This is actually true only for representation of the field by the predictors at zero lag, since only then are the time series coefficients of the modes uncorrelated.

The EOF analysis therefore enables fields of highly correlated data to be

represented by a small number of orthogonal functions and their corresponding time coefficients. These functions furthermore are not of a predetermined form as in Fourier analysis, but depend on interrelationships within the data.

The EOF approach is described here briefly. Further mathematical details may be obtained from the above references or from a text on linear multivariate statistical analysis. The SST anomaly field, T(x,t), is represented by a matrix of the SST departures from the monthly means at each grid point for each month. Space dependence is noted by the subscript x and time dependence by t. Each data point is simply the departure of the SST from the climatological mean for each month and grid point. No additional weighting or standardization is applied as in some other studies. This is felt to be the best and simplest first approach, and is best for error recognition in the data and and for physical interpretation of the modes of variation. The goal of the EOF approach is to represent the data field, T(x,t), as the sum of the products of the orthogonal spatial functions of $f_m(x)$ and their associated time varying amplitudes $c_m(t)$. They are chosen to be orthogonal and to therefore most efficiently approximate the field in the sense that the mean square error is minimum

$$< [T(x,t) - \sum_{m}^{N} f_m(x) c_m(t)]^2 > \quad = \min. \tag{5}$$

This is accomplished by forming the variance-covariance matrix by multiplying the anomaly data matrix by its transpose,

$$C_{MM}(x,x') = T_{MN}(x,t)\ T_{NM}(x',t), \tag{6}$$

where M is the number of grid points and N the number of months. This symmetric matrix is then diagonalized to derive the eigenvalues, λ_m, and the eigenfunctions of $f_m(x)$. The matrix of eigenfunctions is that orthogonal matrix which diagonalizes the variance-covariance matrix,

$$F'_{MM}\ T_{MN}\ T'_{NM}\ F_{MM} = \Lambda_M, \tag{7}$$

where F is the matrix of eigenfunctions and Λ the diagonal matrix of eigenvalues. The eigenfunctions, or EOF modes, are constrained to be orthonormal and thus represent independent spacial modes of variation,

$$\sum_x^N f_m(x)\ f_n(x) = \delta_{mn}\ ,\ \delta_{mn} = \begin{cases} 0, & m \neq n \\ 1, & m = n \end{cases} \tag{8}$$

The percent of the total variance of the SST field represented by each mode is given by

$$\sigma_m = \lambda_m \ / \sum_k \lambda_k \tag{9}$$

The time dependent coefficients, $c_m(t)$, for each mode are then calculated by summing over all grid points the product of the SST departures for each month and the eigenfunction value for the mode,

$$c_m(t) = \sum_x^m T(x,t) f_m(x), \tag{10}$$

where M is the total number of grid points and m the mode number. The time series of coefficients for each mode are uncorrelated with each other at zero lag, with their means all zero and their variances M times the corresponding eigenvalue λ_m,

$$\sum_{t}^{M} c_m(t)\ c_n(t) = M\ \lambda_m\ \delta_{mn}, \tag{11}$$

The original SST anomaly field can then be reconstructed using those modes with the largest λ_m. These are the "principle" modes of variation of the field. Choosing the particular modes for use in reconstructing or analyzing the SST field provides a way of filtering out scales of variation which are due to data errors or are of little interest to the particular study.

In this study an EOF analysis was performed on the set of 250 Pacific Ocean SST anomaly grid points over the 335 month period January 1949 to November 1976. The variance-covariance matrix was diagonalized and eigenfunctions calculated by the EISPACK matrix decomposition routines, and the time series coefficients calculated as described above. Table 1 shows the eigenvalues of and the percent of the nonseasonal SST variance explained for each of the 25 most important EOF modes. "Most important" refers to those modes with the largest eigenvalues.

c. Time Series Analysis

The time series of the coefficients, $c_m(t)$, for each EOF mode represents the time variation of each orthogonal function of space. The contribution of a particular mode to the SST variation of a region in space, at a particular time, can be determined by multiplying the value of the space function at the area by the time series value for that mode and month. Since the time series are uncorrelated at zero lag, they provide the most efficient representation of the time fluctuations of the SST field. This is particularly useful in statistical modeling and prediction of the SST anomaly fields since a few of the most important orthogonal time series may be used instead of the whole set of highly correlated series from the individual grid points.

These EOF time series were analyzed to characterize the time variation of the principle modes of the SST field fluctuations. Their autocorrelation and partial autocorrelation functions were calculated and used to determine stochastic models for their variation. The techniques used are described in Box and Jenkins (1976). Similar analyses for the characterization, modeling, and prediction of sea surface temperature time series have been performed by Namias and Born (1970) and Reynolds (1978) for the North Pacific Ocean and Hsiung (1978) for the Atlantic Ocean.

The autocorrelation function of the process generating the time series is defined below as the standardized autocovariance at time lags k. E[] denotes expected value and μ the theoretical mean of the process.

$$\gamma_k = \mathrm{cov}[c_t, c_{t+k}] = E[(c_t-\mu)(c_{t+k}-\mu)] \qquad (12)$$

$$\rho_k = \gamma_k \Big/ \sqrt{E[(c_t-\mu)^2]E[(c_{t+k}-\mu)^2]} = \gamma_k / \sigma_c^2 \qquad (13)$$

For a stationary process $\sigma_c^2 = \gamma_0$ is constant at all times, therefore

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad \text{with} \quad \rho_0 = 1 \qquad (14)$$

The autocorrelation function is estimated from a finite time series of length N by,

$$r_k = \frac{b_k}{b_0} \quad , \text{ where } b_k = \frac{1}{N} \sum_{t=1}^{n-k} (c_t-\bar{c})(c_{t+k}-\bar{c}) \qquad (15)$$

The variance of the estimated autocorrelation function for a stationary normal process at lags k greater than some value q beyond which the function should have died out is given by Bartlett's approximation

$$\text{var}\,[r_k] \simeq \frac{1}{N} \{1 + 2 \sum_{v=1}^{q} r_v^2 \} \qquad (16)$$

The square root of this is the large lag standard error used in evaluating the assumptions that the autocorrelation has died out for $k > q$. Care must be taken in interpreting the individual autocorrelations because large covariances can exist between them distorting the visual appearance of the estimates.

The partial autocorrelation function is given by,

$$\hat{\phi}_{kk} = |R_k^*| \Big/ |R_k| \qquad (17)$$

where R_k is the k X k symmetric matrix of autocorrelations, r_k, and R_k^* is the R_k with the last column replaced by $(r_1 \ldots r_n)^{tr}$. For example,

$$\hat{\phi}_{11} = r_1$$

$$\hat{\phi}_{22} = \begin{vmatrix} 1 & r_1 \\ r_1 & r_2 \end{vmatrix} / \begin{vmatrix} 1 & r_1 \\ r_1 & 1 \end{vmatrix} = \frac{r_2 - r_1^2}{1 - r_1^2} \qquad (18)$$

The k^{th} partial autocorrelation is the k^{th} autoregression coefficient estimated by fitting the time series to a k^{th} order process. Thus, a first order process will have a value for $\hat{\phi}_{11}$ with the rest being ideally zero. The standard deviation of the estimates of $\hat{\phi}_{kk}$ where k is greater than the order of the process is given by Quenouile's approximation,

$$\text{S.E.}\,[\hat{\phi}_{kk}] \simeq \frac{1}{\sqrt{N}}, \quad k \geq p + 1 \qquad (19)$$

where N is the time series length and p the order of the process.

By inspection of the autocorrelation and partial autocorrelation estimates it is possible to recognize patterns which could be explained by some stochastic model. In this study we evaluate fits to a first or second autoregressive model, with the time series assumed to be stationary and nonseasonal,

$$c_i = \phi_1 c_{i-1} + a_i \qquad \text{(1st order)}$$

$$c_i = \phi_1 c_{i-1} + \phi_2 c_{i-2} + a_i \qquad \text{(2nd order)} \qquad (20)$$

where ϕ_i are the autoregressive parameters and a_i the random (white noise) input.

Also in this study the interaction of the primary modes of variation are evaluated by cross correlation estimates at nonzero time lags. Under the assumption that the pair of time series are realizations of a hypothetical population pair of the time series, called a bivariate stochastic process, the estimates of the cross correlation function at lag k are given by

$$r_{xy}(k) = \frac{b_{xy}(k)}{s_x s_y}, \quad k = \pm 0, \pm 1, \ldots \tag{21}$$

where

$$b_{xy} = \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(y_{t+k} - \bar{y}) , \quad k = \pm 0, \pm 1, \ldots$$

and

$$s_x = \sqrt{b_{xx}(0)} \quad , \quad s_y = \sqrt{b_{yy}(0)} \ .$$

The variance of the cross correlations is given by another Bartlett approximation

$$\text{var}\,[r_{xy}(k)] \simeq (n - k)^{-1} \sum_{v=-\infty}^{\infty} \rho_{xx}(v)\rho_{yy}(v) \tag{22}$$

On the hypothesis that the two processes have zero cross correlation, equation (22) gives the variance of the estimates at lags k.

IV. CHARACTERISTICS OF THE SST FIELD

a. Seasonal and Annual Characteristics

The mean features of the Pacific SST's have been presented previously by Weare, et al (1976), Davis (1976), in various papers by Namias and Bjerknes, and in other papers and atlases. Presented here are the averages and standard deviations of the SST data set described in Section IIIa for a large part of the equatorial and North Pacific Ocean for the period January 1949 to November 1976. This analysis is more extensive in time and in coverage of the Pacific Ocean than either that of Davis or Weare, et al. Additionally it is valuable to compare the EOF modes with the average characteristics of the data set on which the analysis was done.

The long-term monthly means for January and July over the 27 year period are shown in Figures 3 and 4, and the long term annual means in Figure 5. These are in general agreement with other analyses and atlases except that the band of cool equatorial upwelled water is not shown as prominantly as it is in some atlases, such as that presented by Tabata (1975). From these maps we can see some of the features commonly presented for the Pacific SST's. The central North Pacific has basically a zonal distribution of temperatures implying that heat exchange at the atmosphere-ocean boundary and seasonal radiation effects are the major seasonal and monthly determinants of the surface temperatures. Dynamic influences in the SST distribution is quite evident in the equatorial regions where warmer water appears to be on the western side and cooler in the east off the coast of South America. Also noted is the large seasonal difference in temperatures between winter and summer in the

North Pacific and less of a change in the tropics. The pool of warmer water in the western equatorial region is thought by some to be due to the effect of westward transport of water along the equator, however, it is probably simply the contrast to the cool upwelled water in the Eastern equatorial region which creates the effect of a warm pool.

Figures 6 and 7 show the standard deviations about the monthly means for January and July, and Figure 8 the long term SST anomaly standard deviations for the whole period. It should be stressed that these represent variations other than the average seasonal pattern. These maps show that the greatest variations in the Pacific SST anomaly field are in the northwest Kuroshio Current region east of Japan, the central equatorial upwelling region, and most significantly the South American coastal region. This agrees with the finding of Weare, et al (1976) in the whole Pacific and Davis (1976) in the North Pacific, for monthly averaged SST anomaly analyses. The large variance in the southern equatorial region at 25°S and 130°W is thought to be an artifact of the sparse and often bad data of the region. It may be due to the discontinuity in the data sources in 1962, since sparse data areas would be most affected by the change in analysis method. There are some variations in the standard deviation patterns from month to month indicating that the assumption of variance stationarity of the time series may not be valid. This could present problems in SST prediction by nonseasonal time series which are continuous for all seasons. Predictability may be quite different in the different seasons as has been shown by Davis (1976, 1978) and as may be expected from differences in the climatic regimes.

b. Time and Space Scales of the SST Anomalies

Empirical orthogonal functions were calculated for the set of 250 Pacific SST deviations from the monthly means for the 335 months in the period 1/49 to 11/76. The eigenvalues were ordered from largest to smallest and only those eigenfunctions corresponding to the largest 25 of the total 250 were calculated. Figure 1 shows the percent and cumulative percent of total variance in the SST field explained by these 25 most important EOF modes. Figure 9 shows the log of the eigenvalues plotted against the ordered mode number. This type of plot is often used as a way of determining which modes have significant information over random noise. Such methods are controversial, and it is doubtful that the SST data satisfy the necessary conditions for them to be valid. We can however see that the slope of the log relationship changes greatly at about mode 8 and that the decrease in variance explained by the higher modes is roughly exponential or even less steep. This would indicate that the modes after about 8-10 have little significant information above noise. In addition, inspection of the maps of the higher order modes shows variations in small space scales that are impossible to interpret physically. The SST field variations are therefore very complex, and the data set noise appears to blend in with natural space scales such that the field is not dominated by only a few principle modes of variation, as is the case for Pacific surface pressure anomalies in Davis (1976). The above results are in agreement with the EOF analysis of Weare, et al (1976), but comparison with that of Davis (1976) for the North Pacific shows that the addition of the equatorial regions requires more modes to explain an equivalent percent of variance.

Figures 10-15 show the six most dominant (largest variance explained) EOF spatial modes of the Pacific SST anomaly field, and figures 16-18 show

the variation of their amplitudes in time for the length of the data records. Plus and minus signs indicate regions in the spatial function which are correlated or anticorrelated. Like signs indicate correlation in the function. It should be stressed however that correlation of areas in a particular EOF mode does not necessarily mean that they are correlated in the total SST field. This is only so when the field is behaving with the particular mode a dominant factor. Contour lines in the EOF maps are in units of $\pm$ 0.1. As discussed previously, the spatial modes are orthogonal and the time series are uncorrelated at zero lag. The dashed section of the time series plots in the years 1977-1979 are extensions beyond the time period for which the EOF's were calculated. These coefficients are calculated in the same way as the originals by equation 10 of the section IIIb, using the same modes but the newer SST anomalies. The time series extended in this manner, without recalculation of the EOF's are not guaranteed to remain uncorrelated unless the modes are highly stable. The extrapolated time series coefficients may be additionally inconsistent with the earlier ones since in their calculation missing or bad SST data points were given zero weights instead of that of adjacent points in time or space. This amounted to approximating missing values with climatology, or zero anomaly. This is particularly noted in the second mode's time series for reasons which will be discussed later.

The EOF's calculated here show great similarity to those of Weare, et al (1976), indicating stability of the primary modes to extensions in time and space. However, there are some differences between the two analyses. The first mode, which here accounts for 18.4% of the variance, corresponds strongly in space and time to the El Niño phenomenon as it does in Weare, et al. The spatial function has a large component in the equa-

torial upwelling and Peru Current region, and the time series is highly correlated with the 2.5°S zonal mean Pacific SST anomalies, which have been used by Newell and Billing (unpublished) as an index of El Niño. Mode 3(6.9% of variance) in this study corresponds to Weare's second mode and along with the northern part of our mode 1 to Davis's (1976) primary mode. This mode may physically relate to the variation in the North Pacific SST which is thought to be related to the surface high pressure pattern and that has been lag correlated with North American weather regimes by Namias. As found by Davis (1976), some areas with high anomaly standard deviations in Figure 7 have their variance divided among several different modes indicating the dynamical complexity of the SST variations. Further physical interpretation of the modes has not been done at this time. It should be stressed that such interpretation may not be totally valid since the decomposition procedure is strictly statistical and that there is no guarantee of any relation to the actual physical mechanisms.

Of special interest is the second mode (9.9%), which upon inspection of its spatial pattern appears to be primarily the result of bad data. Strong centers of action are in the southern equatorial region where data reports are sparse. The time series of this mode shown in figure 16 has a relatively low variance until 1963, where there is a shift in mean and an increased variance. This is consistent with the shift in the data sources noted in section IIIa. The usefulness of the EOF decomposition is shown in this example by the fact that a data set problem is isolated in one mode and can be eliminated in any reconstructions of the SST field. The extrapolation (dashed line) of this time series drops in mean and variance approaching that of before 1963. This is probably because some data points which were commonly missing throughout the whole time series are treated

in the extension calculation as zero SST anomaly, as was done in the analysis of the data before 1963. Thus the contribution of those areas of the Pacific to the variance of the mode are not included as they were in the 1963 to 1976 analyses.

The time behavior of the SST anomaly field fluctuations is examined by calculation of the autocorrelation and partial autocorrelation functions of the primary EOF modes' time series. Information on the order of regressivity, estimates of the regression model parameters, and calculation of the time scales of the time series may be derived from these plots. Namias and Born in evaluating the time coherence of North Pacific SST's found their time series to be basically first order autoregressive (simple Markov). They further broke the analysis into seasons and found winter to have higher persistence (larger time scale in the Markov process) due to the deeper mixed layer than the summer. Frankignoul and Hasselman (1976) have stochastically modeled SST time series by a damped first order autoregressive process and Reynolds (1978) has identified those SST grid squares in the North Pacific whose time series exhibit first or second order autoregressivity.

Figures 19-23 are plots of the autocorrelation and partial autocorrelation functions for the first five EOF time series shown earlier in figures 16-18. From these plots models may be fitted to the time series and parameters calculated. The two standard error significance levels for these functions may be calculated from equations 16 and 19 of section IIIc. This value is 0.11 for the autocorrelations, under the assumption of a zeroeth order random process (white noise), and is the same for the partial

autocorrelations for all autoregressivity order assumptions. The large lag two standard error levels for the autocorrelations vary with the time series and with the assumption of the lag number beyond which the autocorrelation should have died out. These estimates are valid only under the assumption that the time series are derived from stationary normal processes. This assumption approximately holds for the shown time series except for the highly nonstationary mode 2. From inspection of the plots the first mode appears to have basically first order autoregressive behavior with superimposed higher order characteristics. The fall off of the autocorrelations is not strictly exponential with a faster decrease and a dip for a time below zero. Higher order effects, although not statistically significant, are seen centered at 48 and 78 months in lag. These deviations from first order Markov (persistence only) behavior are probably representations of the time character of the El Niño fluctuation. A longer time series is needed to resolve significantly such lower frequency variations. The second mode, as should be expected from viewing the time series plot, shows highly nonstationary behavior due to the data source shift. Modes 3 and 5 show classic first order autoregressive behavior. Mode 4 could be modelled as either first or second order autoregressive with superimposed seasonal (12 month) component. The second order parameter of the partial autocorrelation is significant at the 2 S.E. level as is the first seasonal peak of the autocorrelations. The residual seasonal component in what should be nonseasonal anomaly data may be due to the nonstationarity in variance among the seasons which was mentioned earlier.

From the estimated autocorrelation values, r_k, the following regression model and time scale parameters are calculated and shown in Table 2

along with the mode number and variance, $\hat{\sigma}_c^2$, for each time series:

(1) The first and second order autoregressive parameters for the time series model fit estimated by the Yule-Walker equations given in Box and Jenkins (1976).

$$\hat{\phi}_{11} = r_1$$

$$\hat{\phi}_{21} = \frac{r_1(1-r_2)}{1-r_1^2} \tag{23}$$

$$\hat{\phi}_{22} = \frac{r_2-r_1^2}{1-r_1^2}$$

(2) The time scale parameter of the time series assuming a first order autoregressive process with the autocorrelation parameters $\rho_k = \rho^k$ for lag k.

$$\rho_k = \exp(-k/\tau_n)$$

$$\therefore \quad \tau_n = -1/\ell n\ \rho_1 \simeq -1/\ell n\ r_1 \tag{24}$$

(3) The time scale parameter calculated from the autocorrelation estimates, r_k, with no assumption of order. This value is that used by Davis (1976) and is an integral time scale determining the time period required to gain a new "degree of freedom" in the time series.

$$\hat{\tau}_n = \sum_{i=-\infty}^{\infty} \rho_n^2 \simeq \sum_{i=-100}^{100} r_n^2 \tag{25}$$

The two time scale parameters may be used to test how well the autocorrelation functions follow first order behavior as well as providing an idea of the EOF modes time scale and degrees of freedom. There is fairly good agreement between the two time scales except for mode 2, which clearly does not follow simple autoregression. Figure 24 shows plots of modes' 1 and 3 autocorrelation functions compared with what a first order process theoretically would yield based on τ_n and $\hat{\tau}_n$ as the time constants.

The partial autocorrelation functions are the estimates of the k^{th} order autoregression coefficient in a fit to a k^{th} order process. These are calculated as described in section IIIc from the autocorrelation values, and can give an indication of the order of the process governing the time series. For modes 1, 3, and 5 the partial autocorrelations do not vary significantly from zero so that these modes are best modeled by a first order process. Mode two has some significant deviations in the function which are interpreted as the result of the data discontinuities. Mode 4 shows first and second order significance in its partial autocorrelations as well as the seasonal component which is more significantly indicated in the autocorrelation function.

V. PREDICTABILITY OF THE SST FIELD

The predictability of the Pacific Ocean SST anomaly field from past values of the SST's themselves is studied by evaluating linear statistical hindcast models. The methods used are similar to those used by Lorenz (1956), by Davis (1976, 1978) in his study of North Pacific SST and sea level pressure anomaly predictability, and by Barnett (1977) in his statistical modeling of El Niño SST events in the equatorial Pacific. More general and advance treatments of statistical prediction methods for geophysical fluid systems and fields are given by Davis (1977), Barnett and Preisendorfer (1978), and Barnett and Hasselman (1979). Although these papers extend the earlier treatments of linear and analogue models, prediction and hindcast skill, and artificial predictability to make them more statistically rigorous and to provide proper tests of significance of the skills, their methods are still somewhat controversial. They were therefore not used in this initial study of Pacific SST predictability.

Since the most important EOF modes and their associated time series may be combined to reconstruct a filtered version of the original SST anomaly fields, prediction of each time series for some future month will yield a prediction for the entire field. The linear prediction model for the time and space dependent temperature field, T(x,t) is of the form

$$\hat{T}(x,\tau) = \sum_{m=1}^{M} c_m(t) f_m(x) \qquad (26)$$

where $c_m(t)$ and $f_m(x)$ are the M orthogonal functions of time and space derived from the EOF analysis used in the reconstruction. The coefficients $c_m(t)$ are predicted by a linear regression model based on past values of each $c_m(t)$ time series itself or on an a priori selection of the other orthogonal time series,

$$\hat{c}_m(t) = \sum_{i=1}^{N} \{d_{im}\, c_i(t-k_i)\} \quad , \tag{27}$$

where the model coefficients d_{im} can be found directly by least squares methods, and k_i are the time lags of each predictor time series. Since the time series $c_m(t)$ are not guaranteed to be uncorrelated at other than zero lag, this prediction model is not the most efficient possible. Another eigenvector analysis must then be performed on the $d_{im} \cdot c_i(t-k_i)$ to produce an orthogonal set of predictors. This is not done in the present study, but should be considered as a possible improvement on the model of equations 26 and 27.

The SST field whose predictability is evaluated in this study is that derived from the reconstruction from the EOF modes 1, 3, 4, 5, 6. This is a filtered version of the actual SST field data with the contribution of higher order modes (≥ 7) considered to be either noise or scales of motion not of interest. Mode 2 is eliminated because it is mostly representative of the discontinuity in the original data sources and not the physical processes. Its elimination therefore allows a more homogeneous data set without loss of true prediction skill. It should be stressed that these choices of the filtered data set and the predictions were made a priori to any evaluation of correlation and skill. Since the predictability evalu-

ated here is actually that of hindcast models, where values of the field within the data set from which the sample covariances and prediction coefficients were calculated, the true predictability is less than the evaluated hindcast skill by the value of artificial predictability. The hindcast skill as defined by Lorenz (1956) and Davis (1976) is the fraction of the predictand fields variance explained by the hindcast model. This is

$$S_H = \left[1 - \frac{\overline{<(T-\hat{T})^2>}}{\overline{<T^2>}}\right] \quad , \tag{28}$$

where the angle brackets denote ensemble average. The hindcast skill can also be calculated from the sum of the skill in hindcasting each EOF time series used in the reconstruciton, weighted by the construction of each mode to the variance of the predictand field. The ratio of each eigenvalue to the sum of the eigenvalues of those modes making up the predictand field is used for these weights. The skill for predicting each time series is given by

$$S_{Hm} = \sum_{i=1}^{M} \frac{<c_m(t)c_i(t-k_i)>^2}{<c_m^2(t)><c_i^2(t-k_i)>} \tag{29}$$

The number of predictors need to be kept small, a priori, in order to minimize the artificial skill created by chance relationships among the predictor and predictand time series. The calculated hindcast skill above exceeds the "true" predictability skill obtainable from a perfect knowledge of the original population, by a quantity directly proportional to the num-

ber of predictors and their integral time scales, τ_i, and inversely proportional to the total number of sample data points N. This can be calculated for each time series predicted and for the reconstructed whole field by the same weighting used in calculating hindcast skill. The artifical skill as given in Davis (1976) is

$$S_A = \frac{1}{N} \sum_{i}^{M} \frac{\tau_i^*}{\Delta t} \quad , \tag{30}$$

where $\frac{\tau_i^*}{\Delta t}$ equals the average time scale calculated from the interrelationships of the predictor time series with the predictand. This is given by

$$\hat{\tau}_i = \sum_{k=100}^{100} r_{0k}\, r_{ik} \tag{31}$$

where the r_{ik} are the autocorrelations of the predictand and predictors. In actual practice the time scales are approximated by τ, the average time scale of all time series considered for each prediction. In evaluating the degree of predictability for each of the models used, the hindcast skill is compared to the artificial skill. Davis (1976) has considered that S_H being greater than the S_A for the prediction is evidence of significant skill. We will use this simple and probably inadequate criteria, as well, instead of the more rigorous significance testing of the preliminary report by Barnett and Hasselman (1979).

Four types of hindcast models with increasing degrees of complexity were evaluated for the prediction of the filtered SST field described above. This predictand field represents 38% of the total original SST

field's variance as can be seen from Table 1. Although this appears to be a small fraction, it represents the major part of the significant large scale SST variation of interest, with variance from data set inaccuracies eliminated. The models for hindcasting each time series making up the filtered SST field are:

(1) First order autoregressive modeling of each time series. This assumes that each time series is approximately first order and future values are estimated using the exponential decay of the first order autocorrelation value. This method has been evaluated by Hsiung (1978) in hindcasting Atlantic SST anomalies, but no attempt was made to consider whether the resultant "measurable skill over climatology" was due to true or artificial predictability. This method relies totally on persistence in the SST time series and provides the "optimal" prediction between climatology (zero skill) and strict persistence of the field.

(2) Modeling of each time series by multiple autoregression on five previous values in order. This is similar to (1) except no assumption of the processes are made and the actual correlation coefficients at the chosen lags are used in the hindcast model. Coefficients for the predictors are determined by least squares methods as mentioned earlier.

(3) Multiple regression modeling of each time series using all five time series considered as predictors. The lag times are the same for all predictors. This method and that in (4) are designed to take advantage of lag cross correlations among the five EOF time series used.

(4) Modeling similar to that of (3) except that the time lags for each predictor are selected from visual inspection of significant cross correlations among the five time series. These cross correlations with the 2 S.E. significance levels are given in Figures 25-27. From this it can been seen that there is some significant cross-correlation among the time series so that an

improvement in skill over autocorrelation models is possible. These cross-correlations also may allow significant prediction of parts of the SST based on the earlier variations at other parts and consequently an understanding of the physical processes involved in the SST fluctuations.

Tables 3 through 6 show the hindcast and artificial skills of each prediction method for each EOF time series modeled. Table 7 shows the combined skills for hindcasts of the total filtered SST field. All skills were calculated for the entire time period of the data set. The artificial skills are given by S_A for each mode and method, and below them are given the hindcast skills for each lead time in months. Those hindcast skills which are not "significant" by the defined criterion are enclosed in parentheses. The true prediction skills may be estimated by subtracting the artificial skills from the hindcast skills for those hindcasts considered "significant". Inspection of these tables indicates that all the time series are predicted quite well by methods based on persistence for the first few months of lead times. After that predictability is quite low and insignificant for all but mode 3. The significant lag correlation of this mode with the first one allows high predictability at larger leads in the regression models. It appears that for prediction of each time series little is gained by using autoregression with five sequential leads over the assumption of a first order Markov process. In fact the increase of the predictors from one to five drastically increases the artificial skill. Perhaps a better method would be simply using the autocorrelation r_k for the k^{th} lead time. Using regression modeling of the five time series does not improve predictability at low lead times and only slightly does at the higher leads. Any improvement

is due mostly to the increased predictability of mode 3 and to a much lower extent modes 5 and 6. Using selected lead times based on the cross correlation of the time series improves hindcast skill of the field only slightly while artificial skill is greatly increased by the increase in possible predictors that may be chosen. The high success of the sequential method compared to this is largely due to the large persistence component in the predictability. Again note the comparison of skills for the four methods in Figure 7. The success of the regression methods may be increased somewhat by performing the additional eigenvector analysis at the desired lags in order to have a most efficient model. This probably would not be great, though, because of the lack of significant cross correlations as was shown in Figures 25-27.

VI. CONCLUSIONS

In order to increase the understanding of the fluctuations of the Pacific Ocean and their role in air-sea interaction processes, the time and space scales of the Pacific SST anomalies have been investigated. Long term monthly statistics were calculated for the set of monthly SST values in 250 grid points over 335 months. An empirical orthogonal function analysis was performed on the set of nonseasonal SST covariances and physical interpretation of the primary modes of variation made. Time scales and stochastic and regression models were evaluated for the time series of amplitudes of the primary EOF modes. The predictability of each time series and of the whole reconstructed field is then evaluated for each of four modeling techniques. Specific conclusions and observations are as follows.

(1) The analysis of the Pacific SST and calculated statistics are for the most part in agreement with those in the literature. The major nonseasonal fluctuations are in the equatorial region, the South American coastal region, and the Kuroshio current regions of the Pacific.

(2) The primary EOF modes of this analysis of the equatorial and Northern Pacific basically agrees with the others in the literature which have smaller space and time coverage. This indicates stationarity of the primary modes, which is important in the consideration of predictability. The first mode corresponds to the SST fluctuation in the equatorial Pacific (El Niño events) and the third to North Pacific fluctuations discussed by Davis and Namias.

(3) A major difference in the present EOF analysis from previous ones is the apparent isolation of a data set problem in one of the primary modes. (Mode 2 accounting for 9.9% of the variance). The nonstationarity in the data created by the change in data sources in 1963 is concentrated in this one mode which can be eliminated from reconstructions. This shows the value in EOF analysis in filtering the field to include only the variations of interest.

(4) The SST anomaly fluctuations as represented by nonseasonal EOF times series show time scales of the order of 3-10 months. This persistence is particularly long for the equatorial region. The EOF times series basically show first order autoregressive behavior with some smaller components of higher order and seasonal behavior in some modes.

(5) First order autoregressive modeling of the primary modes of the SST anomaly variations gives highly significant predictive skill of the whole filtered field for lead times of a few months. Some additional significant skill is gained at the higher lead times for a regression modeling of each time series by the others; however, this is small in magnitude because of the very little significant lag cross-correlations among the modes. Some potential for greater predictability of portions of the SST field at high lead times is seen (i.e., the significant correlation of mode 1 leading mode 3.)

(6) The importance of limiting the number of predictors and doing so a priori is indicated by the relative hindcast and artificial skills seen in comparing the autoregressions with 1 and 5 predictors and in the extension of the regression model to selecting the time leads based on the cross-correlations.

REFERENCES

Barnett, T., 1977a: The principal time and space scales of the Pacific trade wind fields. J. Atmos. Sci., 34, 221-236.

Barnett, T., 1977b: An attempt to verify some theories of El Niño. J. Phys. Oceanogr., 7, 633-647.

Barnett, T., 1978: Prediction of ocean/atmosphere variables in the equatorial Pacific Ocean. SCOR WG55 Meeting, San Francisco, Calif.

Barnett, T. and R. Preisendorfer, 1978: Multifield analog prediction of short-term climate fluctuations using a climate state vector. J. Atmos. Sci., 35, 1771-1787.

Barnett, T. and K. Hasselman, 1979: Techniques of linear prediction with application to oceanic atmospheric fields in the tropical Pacific. (unpublished)

Bjerknes, J., 1960: Ocean temperatures and atmospheric circulation. WMO Bulletin, Vol. IX, 151-157.

Bjerknes, J., 1961: El Niño study based on analysis of ocean surface temperatures, 1935-1957. Inter. Am. Tropical Tuna Comm. Bull., Vol. V, 219-272.

Bjerknes, J., 1964: Atlantic air-sea interaction. Adv. in Geophys., 10, 1-82.

Bjerknes, J., 1966: A possible response of the atmosphere Hadley circulation to equatorial anomalies of ocean termperature. Tellus, 18, 516-524.

Bjerknes, J., 1969: Atmospheric teleconnections from the equatorial Pacific. Mon. Wea. Rev., 97, 163-172.

Bjerknes, J., 1972: Large-scale atmospheric response to the 1964-65 Pacific equatorial warming. J. Phys. Oceanogr., 2, 212-217.

Bjerknes, J., 1974: Atmospheric teleconnections from the equatorial Pacific during 1963-67. Final Report, NSF Grant GA 27754, Univ. of Calif.

Box, G. and G. Jenkins, 1976: Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, Calif.

Brier, G. and G. Meltesen, 1976: Eigenvector analysis for prediction of time series. J. Appl. Meteor., 15, 1307-1312.

Chui, W. and A. Low, 1979: A preliminary study of the possible statistical relationship between the tropical Pacific sea surface temperatures and the atmospheric circulation. Mon. Wea. Rev., 107, 18-25.

Clark, N., 1972: Specification of sea surface temperature anomaly patterns in the eastern North Pacific. J. Phys. Oceanogr., 2, 391-404.

Davis, R., 1976: Predictability of sea surface temperature and sea level pressure anomalies over the North Pacific Ocean. J. Phys. Oceanogr., 6, 249-266.

Davis, R., 1977: Techniques for statistical analysis and prediction of geophysical fluid systems. Geophys. Astrophys. Fluid Dyn., 8, 245-277.

Davis, R., 1978: Predicability of sea level pressure anomalies over the North Pacific Ocean. J. Phys. Oceanogr., 8, 233-246.

Favorite, F. and R. McClain, 1973: Coherence in transpacific movements of positive and negative anomalies of sea surface temperature, 1953-60. Nature, 244, 139-143.

Frankignoul, C. and K. Hasselman, 1976: Stochastic climate models, Part 2, Application to sea-surface temperature anomalies and thermocline variability. Tellus, 29, 289-305.

Haney, B., W. Shriver, K. Hurt, 1979: A dynamical-numerical study of the formation and evolution of large-scale ocean anomalies. J. Phy. Oceanogr., (in press).

Hasselman, K., 1976: Stochastic climate models, Part 1, Theory, Tellus, 28, 473-485.

Haworth, C., 1978: Some relationships between sea surface temperature anomalies and surface pressure anomalies. Quart. J. R. Met. Soc., 104, 131-146.

Hickey, B., 1975: Relationship between fluctuations in sea level wind stress and s.s.t. in the equatorial Pacific. J. Phys. Oceanogr., 5, 460-475.

Hsiung, J., 1978: Statistical prediction of Atlantic sea-surface temperatures using empirical orthogonal functions. M.S. Thesis, M.I.T.

Huang, J., 1978: Numerical simulation studies of oceanic anomalies in the North Pacific Basin, I. J. Phys. Oceanogr., 8, 755-788.

Hurlburt, et al, 1976: A numerical simulation of the onset of El Niño. J. Phys. Oceanogr., 6, 621-631.

Julian, P. and R. Chervin, 1978: A study of the southern oscillation and Walker circulation phenomenon. Mon. Wea. Rev., 106, 1433-1451.

Kutzbach, J., 1967: Empirical eigenvectors of sea-level pressure, surface temperature, and precipitation complexes over North America. J. Appl. Meteor., 16, 791-802.

Lorenz, E., 1956: Empirical orthogonal functions and statistical weather prediction. Sci. Report No. 1, Statistical Forecasting Project, Dept. of Met., M.I.T.

Lorenz, E., 1969: Studies of atmospheric predictability. Final Report, Stat. Forecasting Project, Dept. of Met., M.I.T.

McCreary, J., 1976: Eastern tropical ocean response to changing wind systems with application to El Niño. J. Phys. Oceanogr., 6, 634-645.

Namias, J., 1959: Recent seasonal interactions between North Pacific waters and the overlying atmospheric circulation. J. Geophys. Res., 64, 631-646.

Namias, J., 1963: Large-scale air-sea interactions over the North Pacific from summer 1962 through the subsequent winter. J. Geophys. Res., 68, 6171-6186.

Namias, J., 1965: Macroscopic association between mean monthly sea-surface temperatures and the overlying winds. J. Geophys. Res., 70, 2307-2318.

Namias, J., 1969: Seasonal interactions between the North Pacific Ocean and the atmosphere during the 1960's. Mon. Wea. Rev., 97, 173-192.

Namias, J., 1970: Macroscale variations in sea-surface temperatures in the North Pacific, J. Geophys. Res., 75, 565-582.

Namias, J., 1971: The 1968-69 winter as an outgrowth of sea and air coupling during antecedent seasons. J. Phys. Oceanogr., 1, 65-81.

Namias, J., 1973: Thermal communication between the sea surface and the lower troposphere. J. Phys. Oceanogr., 3, 373-378.

Namias, J., 1974: Longevity of a coupled air-sea-continent system. Mon. Wea. Rev., 102, 638-648.

Namias, J., 1975: Stabilization of atmospheric circulation pattern by sea surface temperatures. J. Mar. Res., 33, 53-60.

Namias, J., 1976: Some statistical and synoptic characteristics associated with El Niño, J. Phys. Oceanogr., 6, 130-138.

Namias, J. and R. Born, 1970: Temporal coherence in North Pacific sea-surface temperature patterns. J. Geophys. Res., 75, 5952-5955.

Newell, R. and B. Weare, 1976a: Ocean temperatures and large-scale atmospheric variations. Nature, 262, 40-41.

Newell, R. and B. Weare, 1976b: Factors governing tropospheric mean temperature. Science, 194, 1413-1414.

Quinn, W., 1974: Monitoring and predicting El Nino Invasions, J. Appl. Meteor., 13, 825-830.

Quinn, W., 1976: An improved approach for following and predicting equatorial Pacific changes and El Niño. Proc. NOAA Climate Diagnostics Workshop, Washington, D.C.

Ratcliffe, R. and R. Murray, 1970: New lag associations between North Atlantic sea temperature and European pressure applied to long-range weather forecasting. Quart. J. R. Met. Soc., 96, 226-246.

Reiter, E., 1978: The interannual variability of the ocean-atmosphere system. J. Atmos. Sci., 35, 349-370.

Reynolds, R., 1978: Sea surface temperature anomalies in the North Pacific Ocean. Tellus, 30, 97-103.

Rowntree, P., 1972: The influence of tropical east Pacific ocean temperatures on the atmosphere. Quart. J. R. Met. Soc., 99, 394-395.

Sette, O., J. Laurs and L. Eber, 1968: Monthly mean charts of sea surface temperatures in North Pacific Ocean, 1949-62. Bur. Commer. Fish. Circ., No. 258, i-vi.

Sellers, W., 1957: A statistical-dynamic approach to numerical weather prediction. Sci. Report No. 2, Stat. Forecasting Pro., M.I.T.

Tabata, S., 1976: The General Circulation of the Pacific Ocean. Part II - Thermal Regime and influence on the climate. Atmosphere, 14, 1-27.

Verstraete, M., 1978: Empirical orthogonal function analysis of the free air temperature field over the African continent. M.S. Thesis, M.I.T.

Weare, B., A. Navato and R. Newell, 1976: Empirical orthogonal analysis of Pacific sea surface temperatures. J. Phys. Oceanogr., 6, 671-678.

Weare, B., 1977: Empirical orthogonal analysis of Atlantic Ocean surface temperatures. Q. J. Roy. Met. Soc., 103, 467-478.

White, W. and T. Barnett, 1972: A servomechanism in the ocean/atmosphere system of the mid-latitude North Pacific. J. Phys. Oceanogr., 2, 372-381.

Wyrtki, K., 1973: Teleconnections in the equatorial Pacific Ocean. Science, 180, 66-68.

Wyrtki, K., 1975: El Niño - the dynamic response of the equatorial Pacific Ocean to atmospheric forcing. J. Phys. Oceanogr., 5, 450-459.

Wyrtki, K., et al, 1976: Predicting and observing El Niño. Science, 191, 343-346.

Wyrtki, K. and J. Haberland, 1968: On the redistribution of heat in the North Pacific Ocean. J. Oceanogr. Soc. Japan, 24, 220-233.

TABLE 1

Eigenvalues and Percent Variance Explained for Each Mode

Mode #	λ	% var	Cum %
1	3278.2	18.4	18.4
2	1762.2	9.9	28.3
3	1231.2	6.9	35.2
4	916.3	5.2	40.4
5	687.2	3.9	44.3
6	640.0	3.6	47.9
7	539.2	3.0	50.9
8	412.8	2.3	53.2
9	387.3	2.2	55.4
10	358.5	2.0	57.4
11	329.7	1.9	59.3
12	275.0	1.5	60.8
13	258.1	1.5	62.3
14	230.4	1.3	63.6
15	220.4	1.2	64.8
16	208.9	1.2	66.0
17	206.9	1.2	67.2
18	183.3	1.0	68.2
19	176.2	1.0	69.2
20	172.2	1.0	70.2
21	167.5	0.9	71.1
22	146.4	0.8	71.9
23	138.8	0.8	72.7
24	136.0	0.8	73.5
25	128.8	0.7	74.2

TABLE 2

Time Series Parameters for Each EOF Mode

Mode #	$\hat{\sigma}_e^2$	$\hat{\phi}_{11}$	$\hat{\phi}_{21}$	$\hat{\phi}_{22}$	τ_n	$\hat{\tau}_n$
1	32.8	.906	.880	.029	10.1	8.2
2	17.6	.928	.809	.128	13.4	49.7
3	12.3	.791	.760	.039	4.3	4.6
4	9.2	.764	.889	-.164	3.7	4.0
5	6.9	.685	.671	.020	2.6	3.5
6	6.4	.631	.552	.125	2.2	3.1
7	5.4	.560	.428	.236	1.7	3.4
8	4.1	.637	.642	-.008	2.2	2.2
9	3.9	.525	.491	.065	1.6	2.2
10	3.6	.596	.564	.055	1.9	1.9
11	3.3	.590	.506	.142	1.9	3.4
12	2.8	.472	.430	.090	1.3	1.7
13	2.6	.529	.529	.000	1.6	1.4
14	2.3	.527	.528	-.004	1.6	1.7
15	2.2	.540	.506	.062	1.6	2.2
16	2.1	.533	.535	-.004	1.6	2.0
17	2.1	.480	.477	.006	1.4	2.6
18	1.8	.493	.495	-.004	1.4	1.3
19	1.8	.519	.502	.033	1.5	1.6
20	1.7	.537	.534	.005	1.6	1.6
21	1.7	.618	.532	.140	2.1	2.8
22	1.5	.397	.390	.017	1.1	1.2
23	1.4	.280	.288	-.029	0.8	0.8
24	1.4	.462	.461	.002	1.3	1.4
25	1.3	.336	.323	.038	0.9	1.0

TABLE 3

Hindcast Skill Assuming 1st Order Autoregressive Model on r_1

	Mode 1	Mode 3	Mode 4	Mode 5	Mode 6
S_A	.025	.014	.012	.010	.009
mos. lead					
1	.821	.626	.584	.469	.398
2	.679	.400	.301	.225	.118
3	.549	.253	.147	.127	.092
4	.431	.171	.065	.069	.038
5	.321	.117	.027	.035	.023
6	.223	.084	.018	.020	.014
7	.142	.059	.015	.011	(.005)
8	.084	.037	.016	(.005)	(.002)
9	.036	.025	.014	(.002)	(.002)
10	(.005)	.016	.014	(.002)	(.001)
11	(-.021)	(.011)	.016	(.001)	(.001)
12	(-.036)	(.009)	(.009)	(.001)	(.001)
13	(-.042)	(.004)	(.005)	(.000)	(.000)
14	(-.038)	(.002)	(.003)	(.000)	(.000)
15	(-.039)	(.001)	(.001)	(.000)	(.000)
16	(-.039)	(.000)	(-.001)	(.000)	(.000)
17	(-.037)	(.000)	(-.001)	(.000)	(.000)
18	(-.033)	(.000)	(-.001)	(.000)	(.000)

TABLE 4

Hindcast Skill for Autoregression Using Five Sequential Lead Times

	Mode 1	Mode 3	Mode 4	Mode 5	Mode 6
S_A	.127	.071	.061	.054	.047
mos. lead					
1	.842	.632	.597	.479	.417
2	.741	.434	.284	.244	.245
3	.642	.300	.130	.159	.133
4	.536	.222	.062	.104	.070
5	.416	.163	(.036)	.064	.069
6	.300	.128	(.050)	(.040)	.067
7	.207	.096	(.061)	(.027)	(.039)
8	.151	(.063)	.068	(.015)	(.039)
9	(.101)	(.043)	.066	(.008)	(.041)
10	(.075)	(.033)	.066	(.009)	(.044)
11	(.054)	(.022)	.065	(.007)	(.038)
12	(.045)	(.017)	.063	(.009)	(.028)
13	(.045)	(.011)	(.050)	(.006)	(.016)
14	(.049)	(.003)	(.041)	(.007)	(.011)
15	(.054)	(.002)	(.030)	(.007)	(.009)
16	(.050)	(.002)	(.027)	(.007)	(.010)
17	(.049)	(.002)	(.027)	(.010)	(.080)
18	(.051)	(.002)	(.040)	(.010)	(.009)

TABLE 5

Hindcast Skill for Regression with Sequential Lead Times

	Mode 1	Mode 3	Mode 4	Mode 5	Mode 6
S_A	.070	.055	.046	.042	.042
mos. lead					
1	.837	.647	.588	.475	.413
2	.718	.453	.278	.244	.245
3	.601	.333	.123	.174	.173
4	.494	.299	.048	.127	.109
5	.381	.286	.022	.096	.106
6	.270	.283	.034	.085	.099
7	.164	.294	.053	.067	.063
8	.093	.301	.067	.056	.064
9	(.053)	.307	.073	.059	.070
10	(.043)	.300	.086	.069	.069
11	(.037)	.286	.083	.060	.057
12	(.049)	.267	.084	.053	.051
13	(.063)	.229	.067	(.036)	(.031)
14	(.069)	.189	.050	.054	.045
15	(.082)	.159	(.029)	.059	.049
16	(.085)	.150	(.041)	(.040)	(.036)
17	(.100)	.134	.055	.047	(.033)
18	(.112)	.110	.057	.068	(.019)

TABLE 6

Hindcast Skill for Regression with Selected Lead Times

	Mode 1	Mode 3	Mode 4	Mode 5	Mode 6
S_A	.185	.186	.189	.188	.188
mos. lead					
1	.839	.659	.596	.501	.426
2	.721	.504	.309	.292	.272
3	.601	.416	(.164)	.228	(.187)
4	.480	.382	(.090)	(.184)	(.146)
5	.369	.357	(.060)	(.154)	(.146)
6	.308	.349	(.057)	(.138)	(.142)
7	.222	.341	(.079)	(.140)	(.111)
8	(.181)	.331	(.088)	(.116)	(.111)
9	(.135)	.326	(.097)	(.116)	(.108)
10	(.118)	.327	(.116)	(.119)	(.114)
11	(.112)	.286	(.121)	(.115)	(.098)
12	(.127)	.267	(.114)	(.115)	(.091)

TABLE 7

Hindcast Skill for Whole Filtered Field

Model	1st order	Auto	Seq. Reg.	Sel. Reg.
S_A	.02	.09	.06	.19
mos. lead				
1	.68	.70	.69	.70
2	.48	.53	.53	.54
3	.36	.41	.40	.43
4	.26	.33	.33	.35
5	.19	.25	.26	.28
6	.13	.19	.20	.25
7	.08	.13	.15	.21
8	.05	.10	.12	(.18)
9	.02	(.07)	.10	(.16)
10	(.01)	(.06)	.10	(.16)
11	(-.01)	(.04)	.09	(.14)
12	(-.01)	(.04)	.09	(.15)

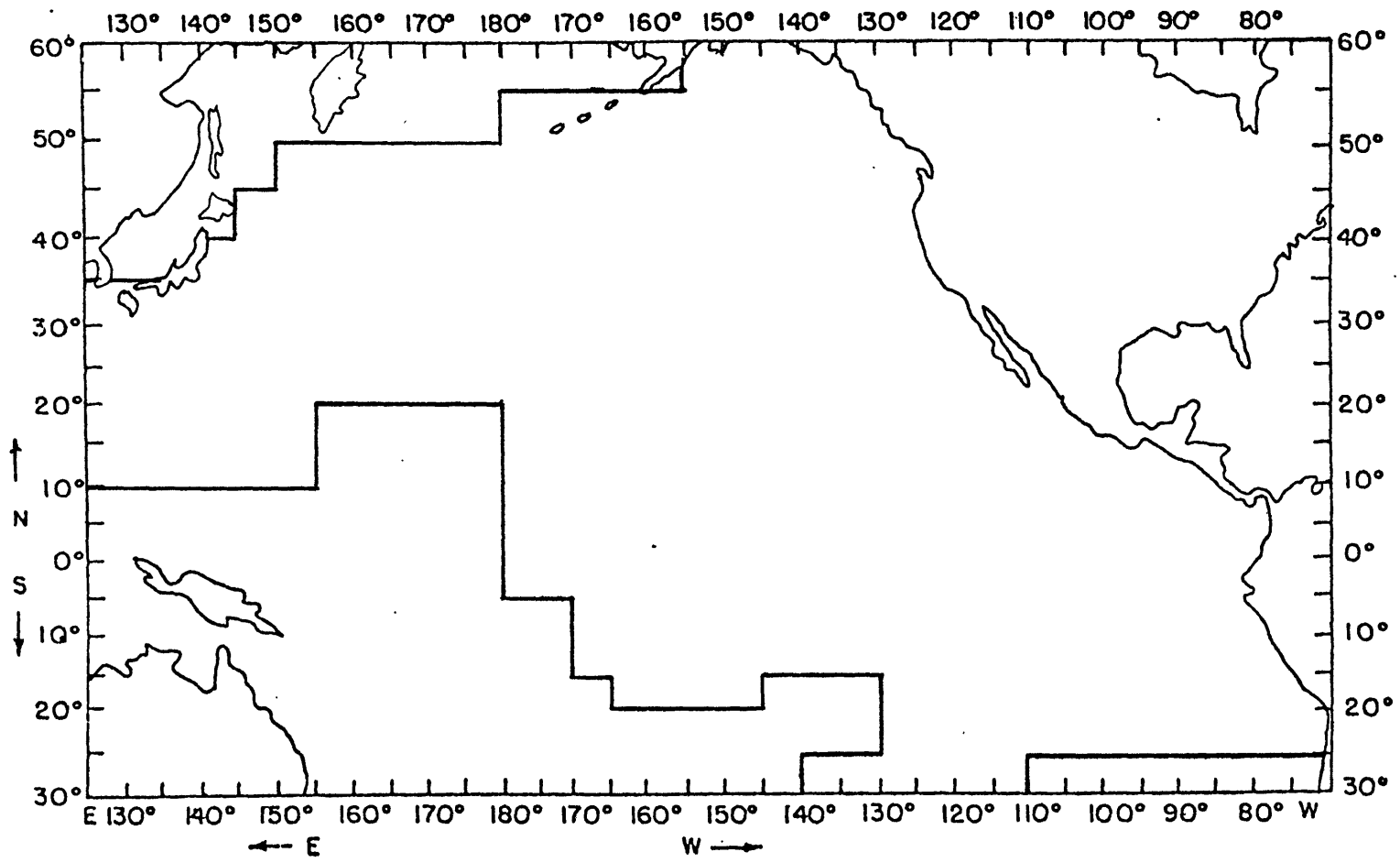


FIGURE 1. Data Set Coverage of the Equatorial and North Pacific Ocean

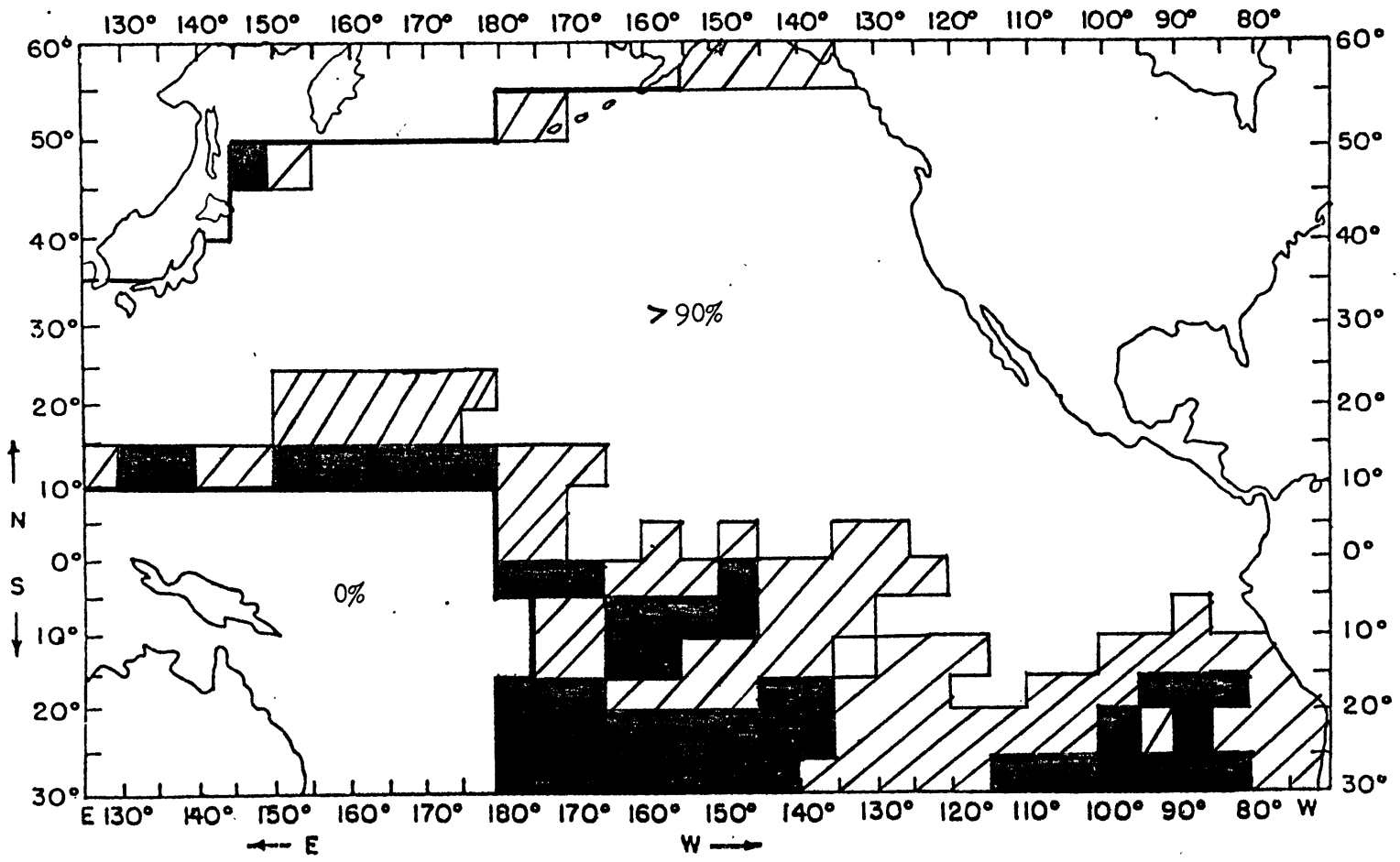
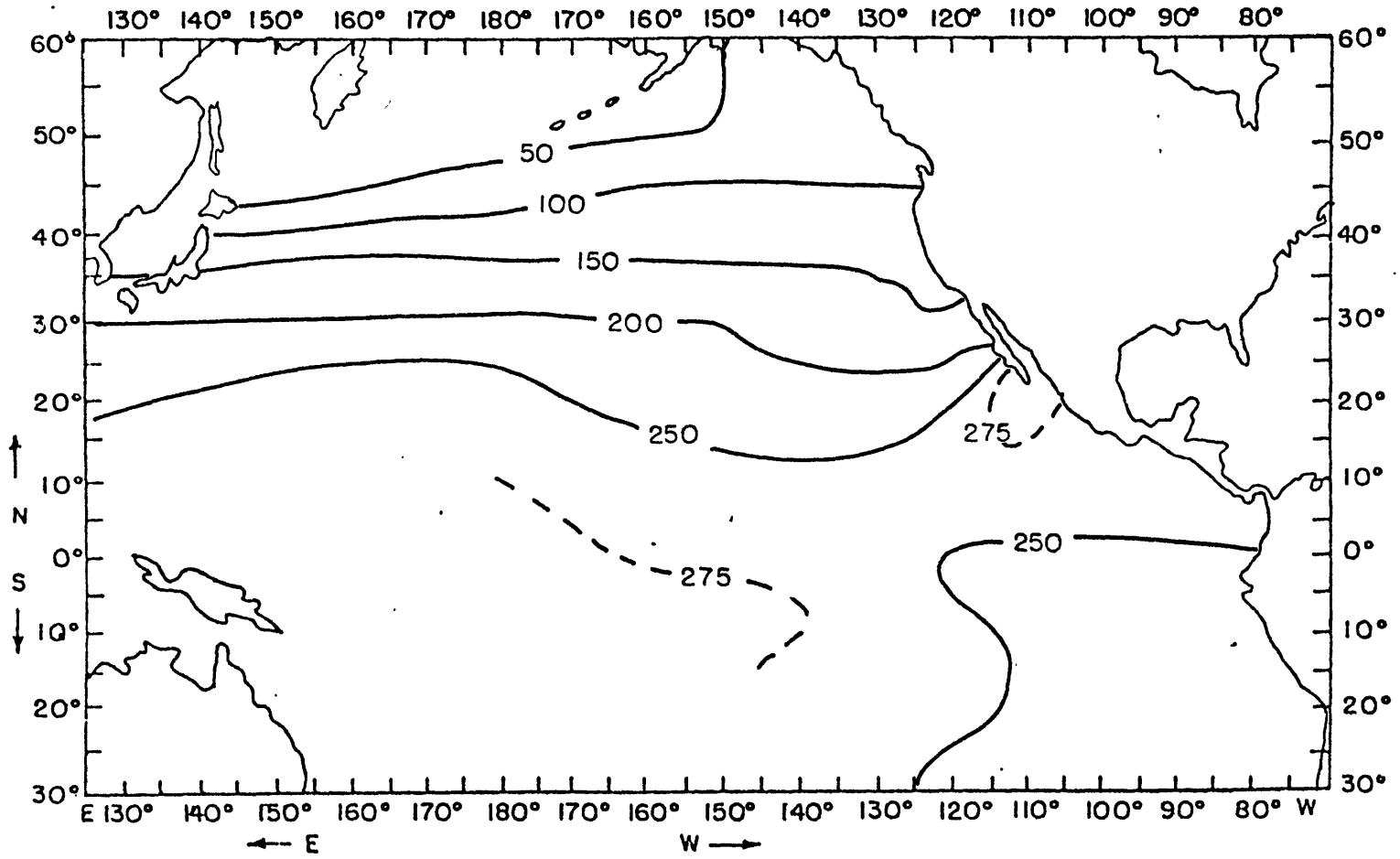
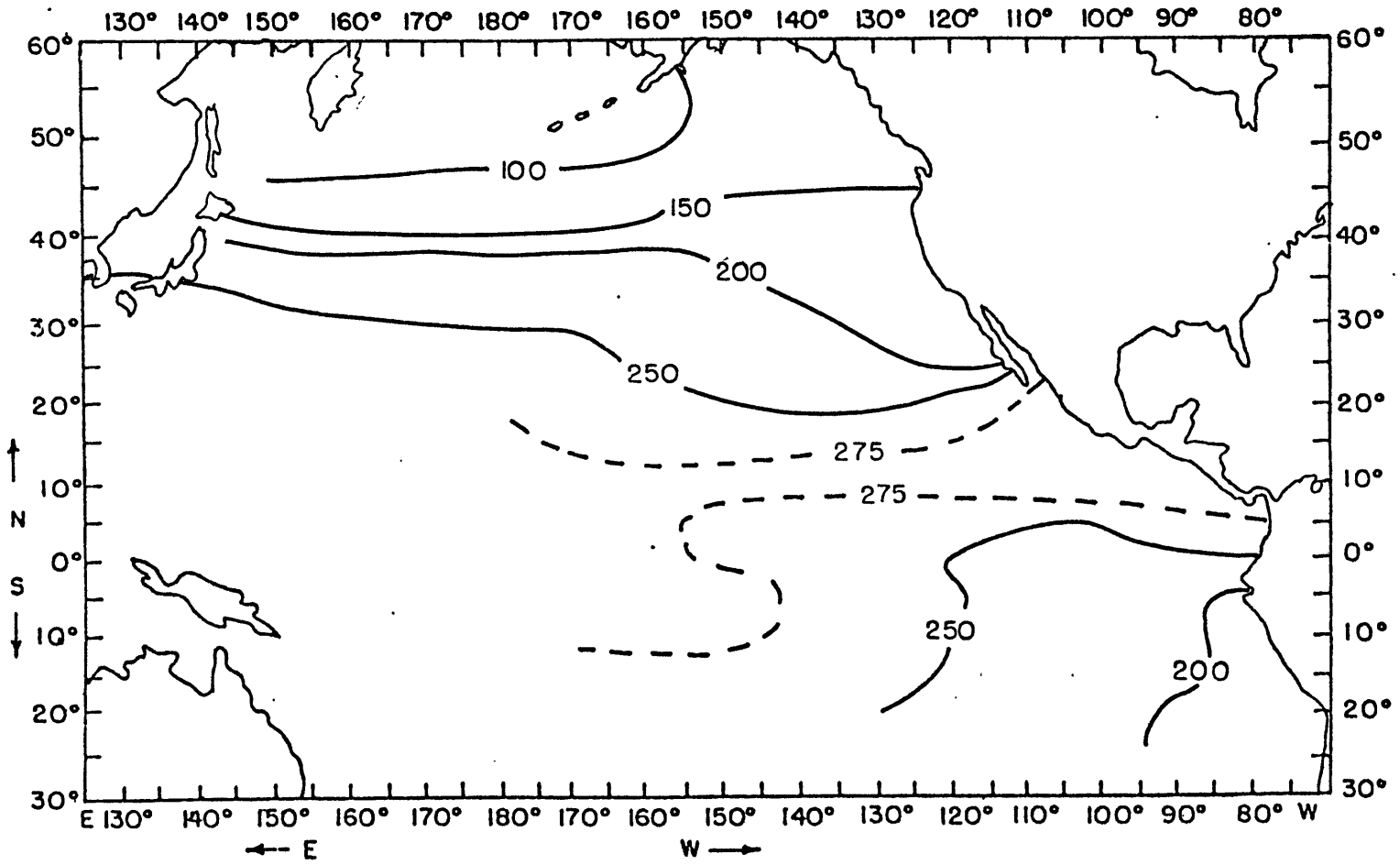


FIGURE 2. Percent of Months During 1963 - 1976 Having SST Data



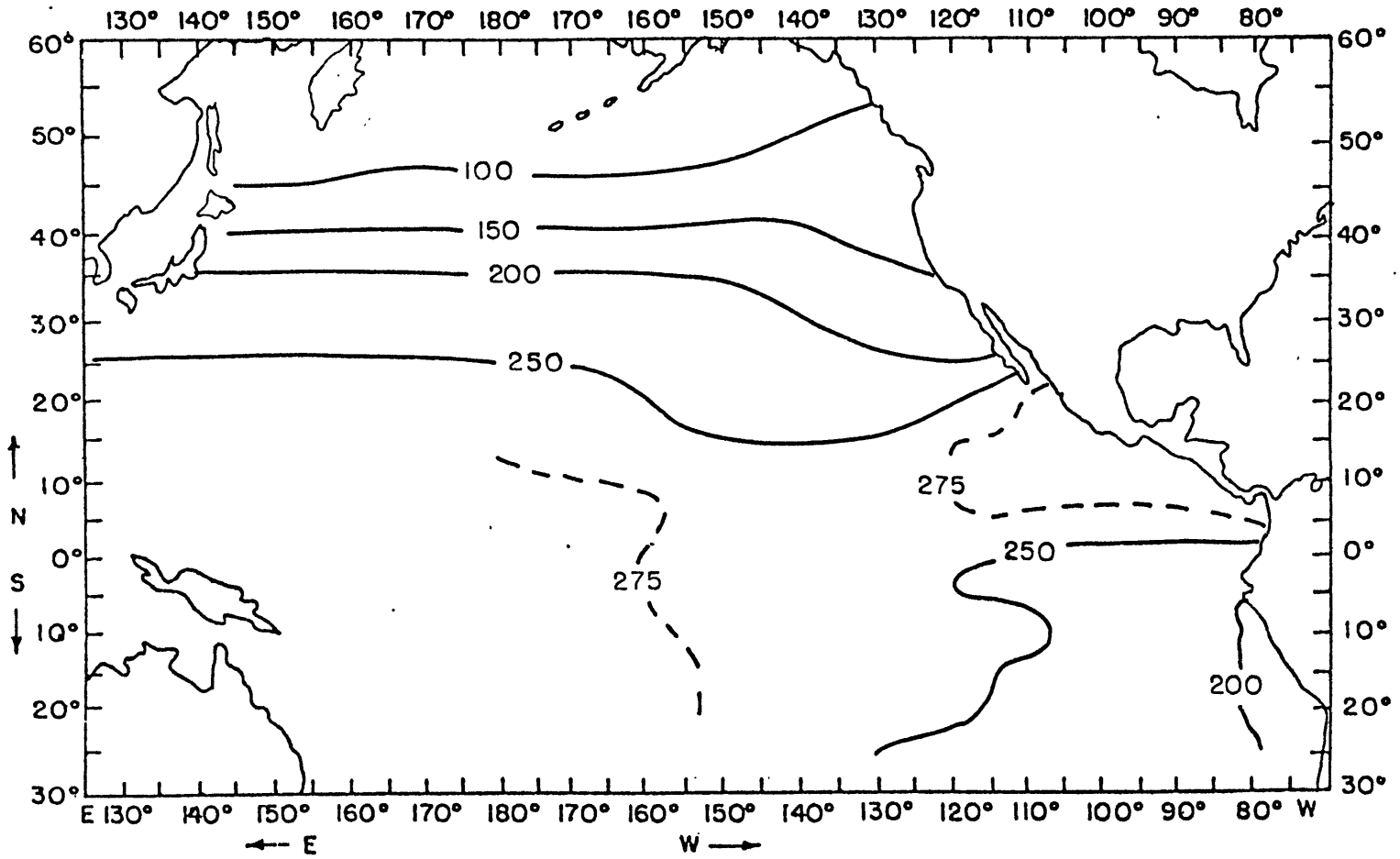
1/10 DEG C

FIGURE 3. Pacific SST Long Term Monthly Means for January



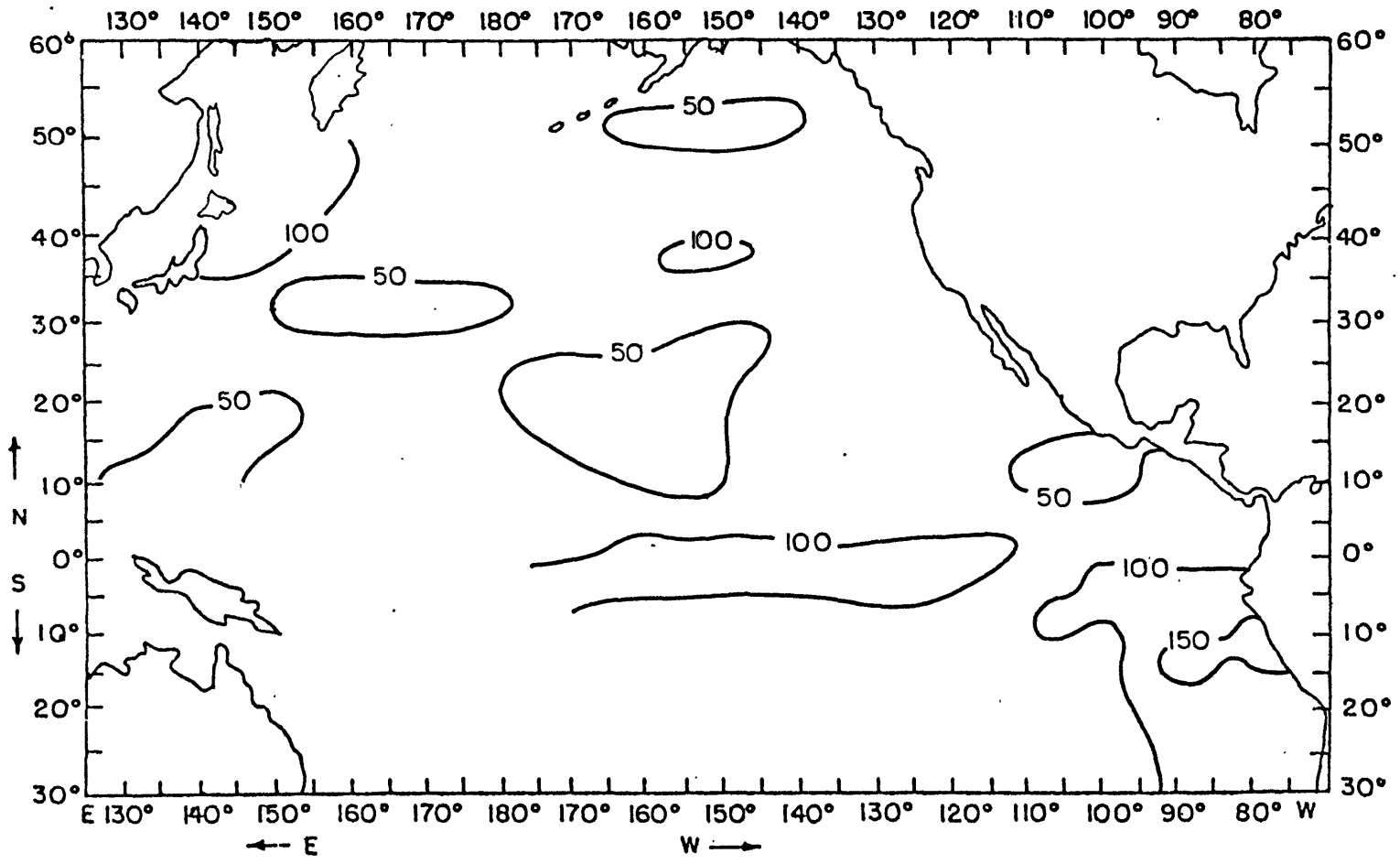
1/10 DEG C

FIGURE 4. Pacific SST Long Term Monthly Means for July



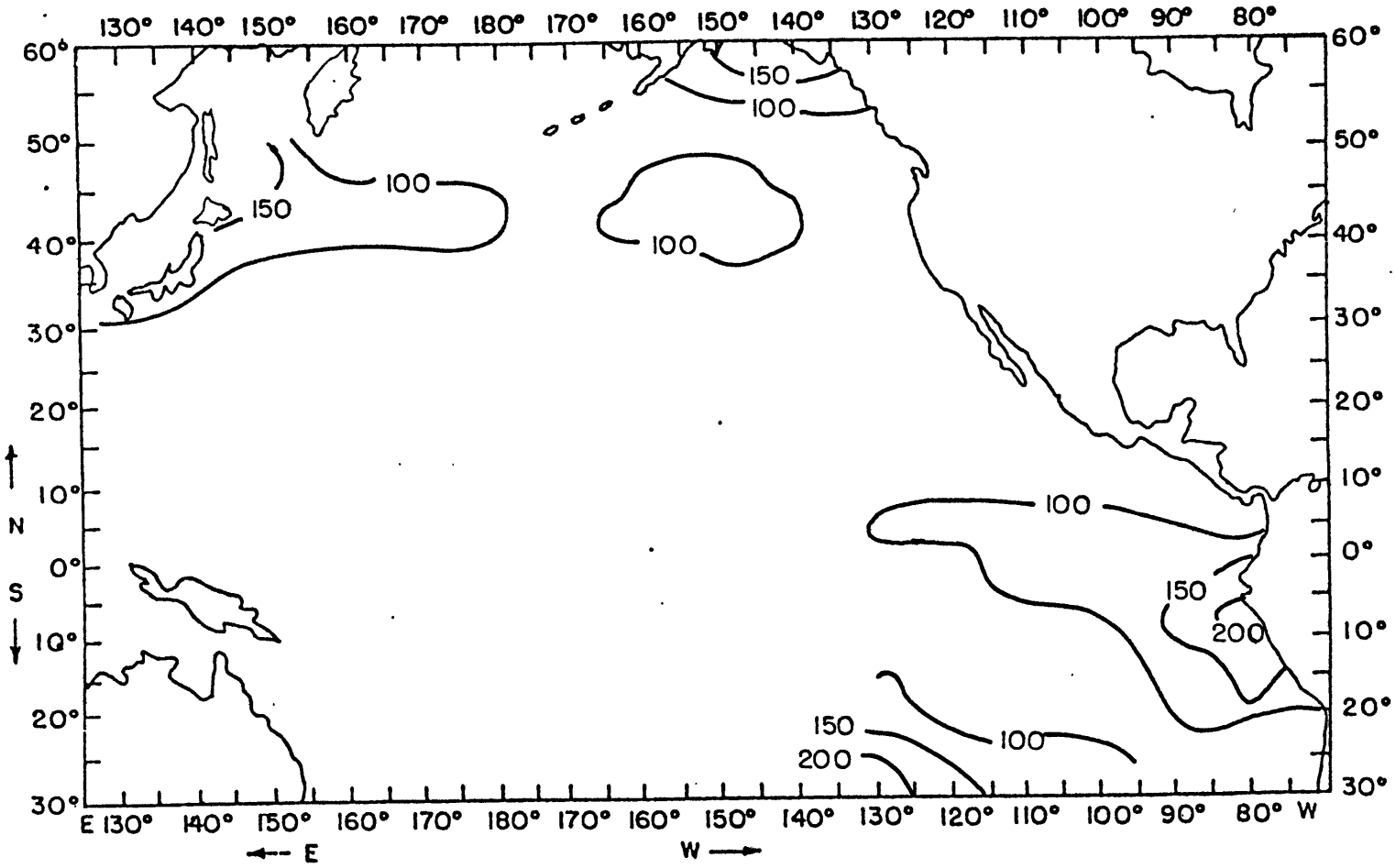
1/10 DEG C

FIGURE 5. Pacific SST Long Term Annual Means



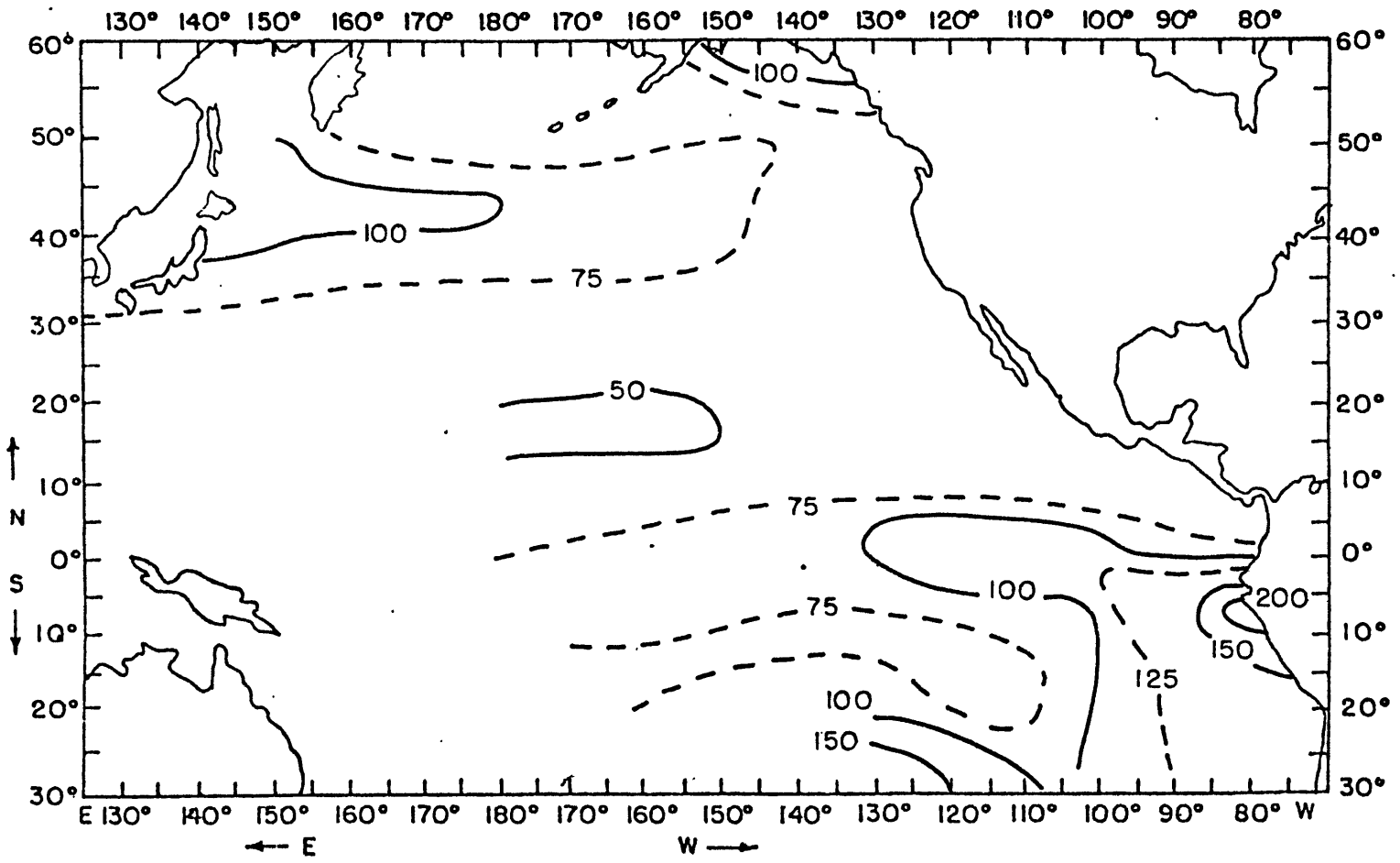
1/100 DEG C

FIGURE 6. Pacific SST Long Term Standard Deviations for January



1/100 DEG C

FIGURE 7. Pacific SST Long Term Standard Deviations for July



1/100 DEG C

FIGURE 8. Pacific SST Long Term Anomaly Standard Deviations

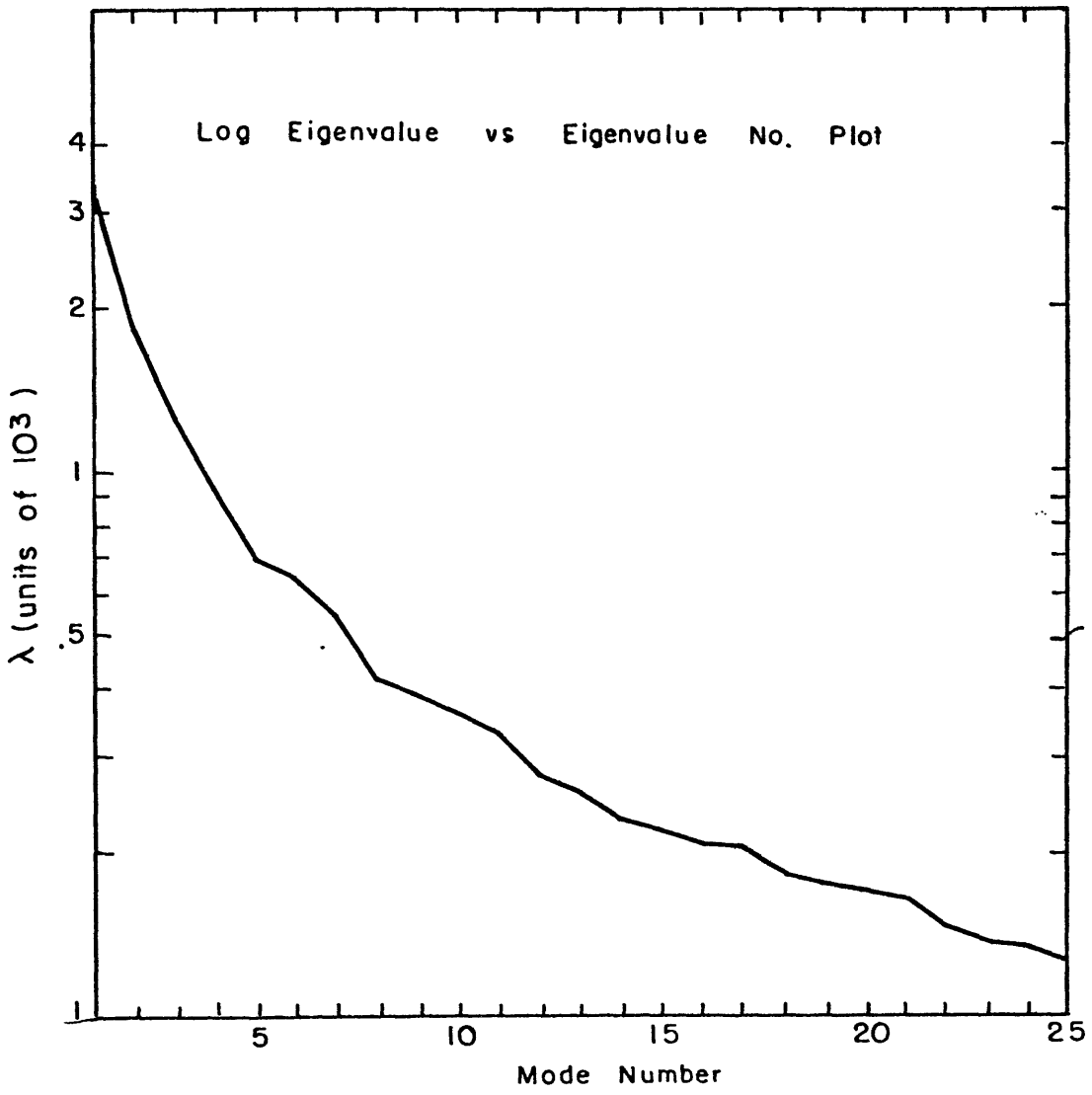

Log Eigenvalue vs Eigenvalue No. Plot
λ (units of 103)
4
3
2
1
.5
1
5
10
15
20
25
Mode Number

FIGURE 9

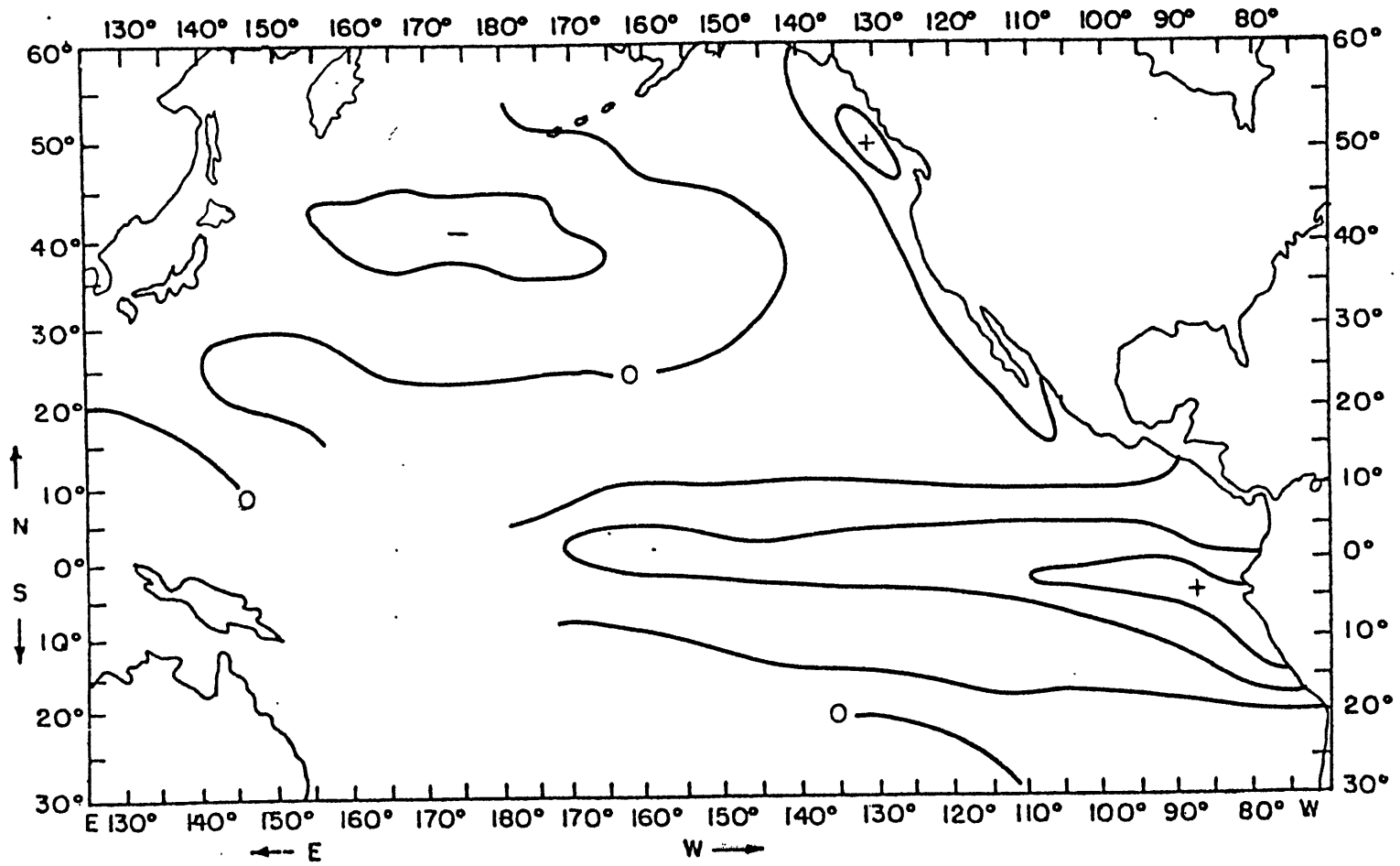


FIGURE 10. Pacific SST Anomalies EOF Mode 1

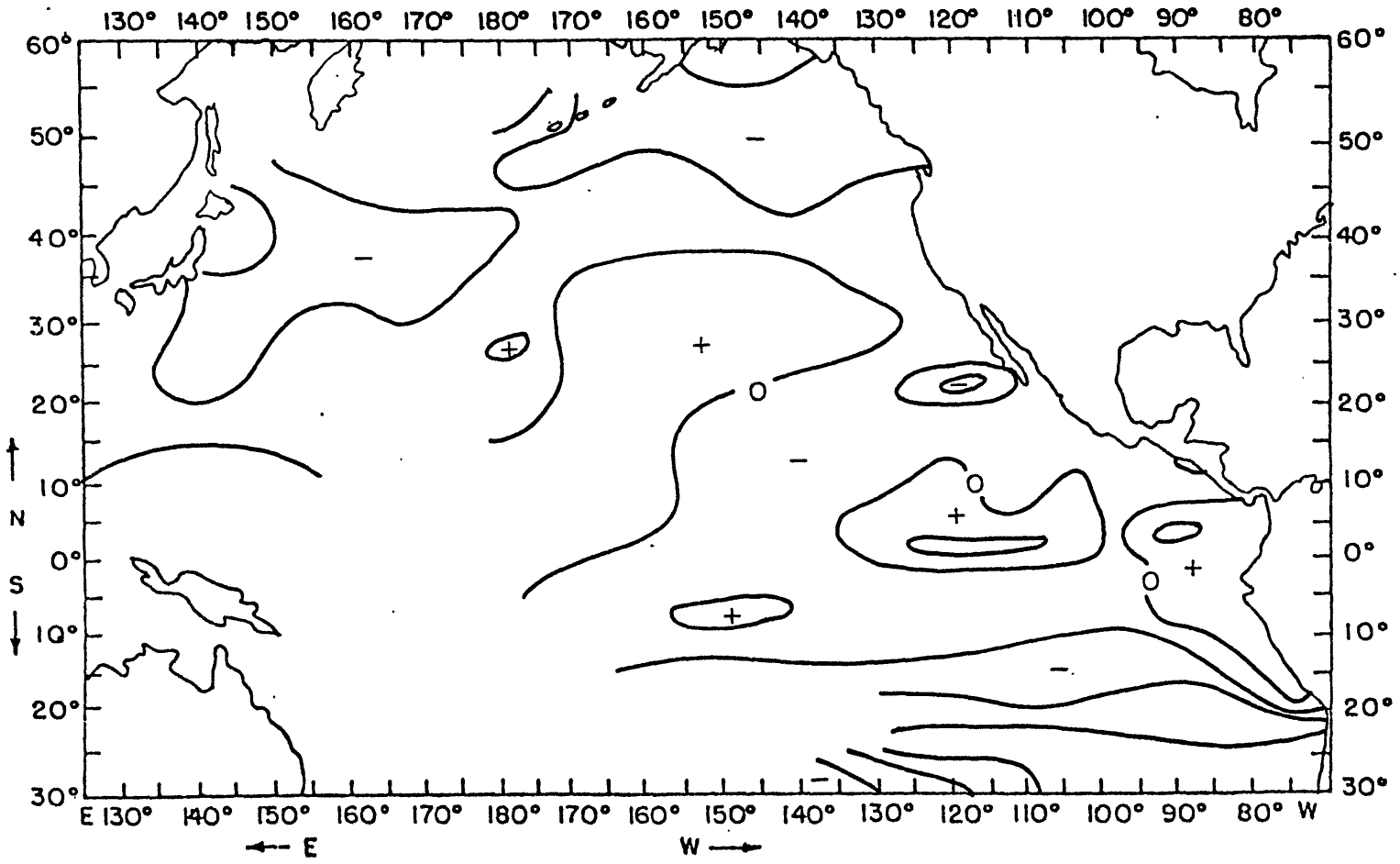


FIGURE 11. Pacific SST Anomalies EOF Mode 2

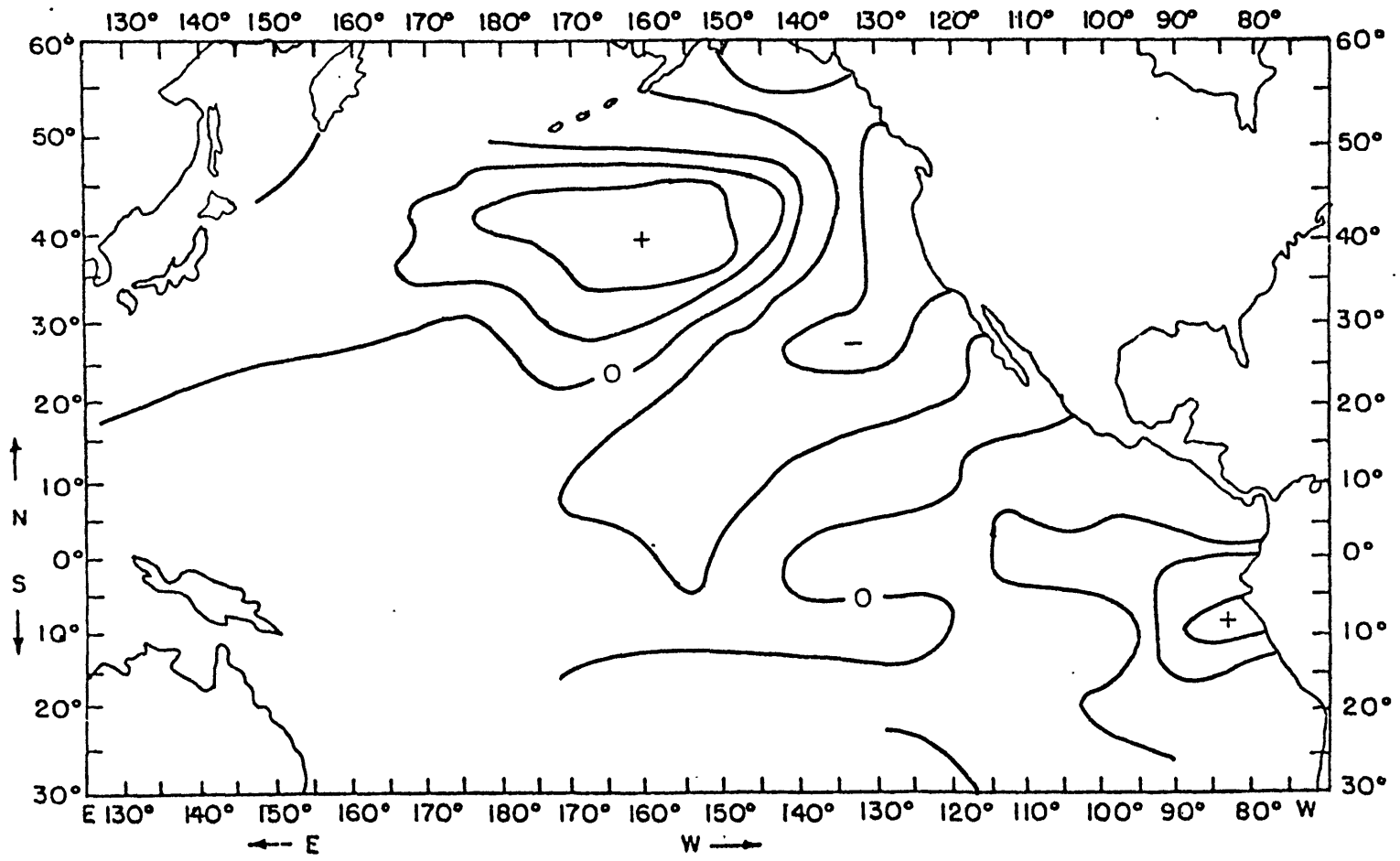


FIGURE 12. Pacific SST Anomalies EOF Mode 3

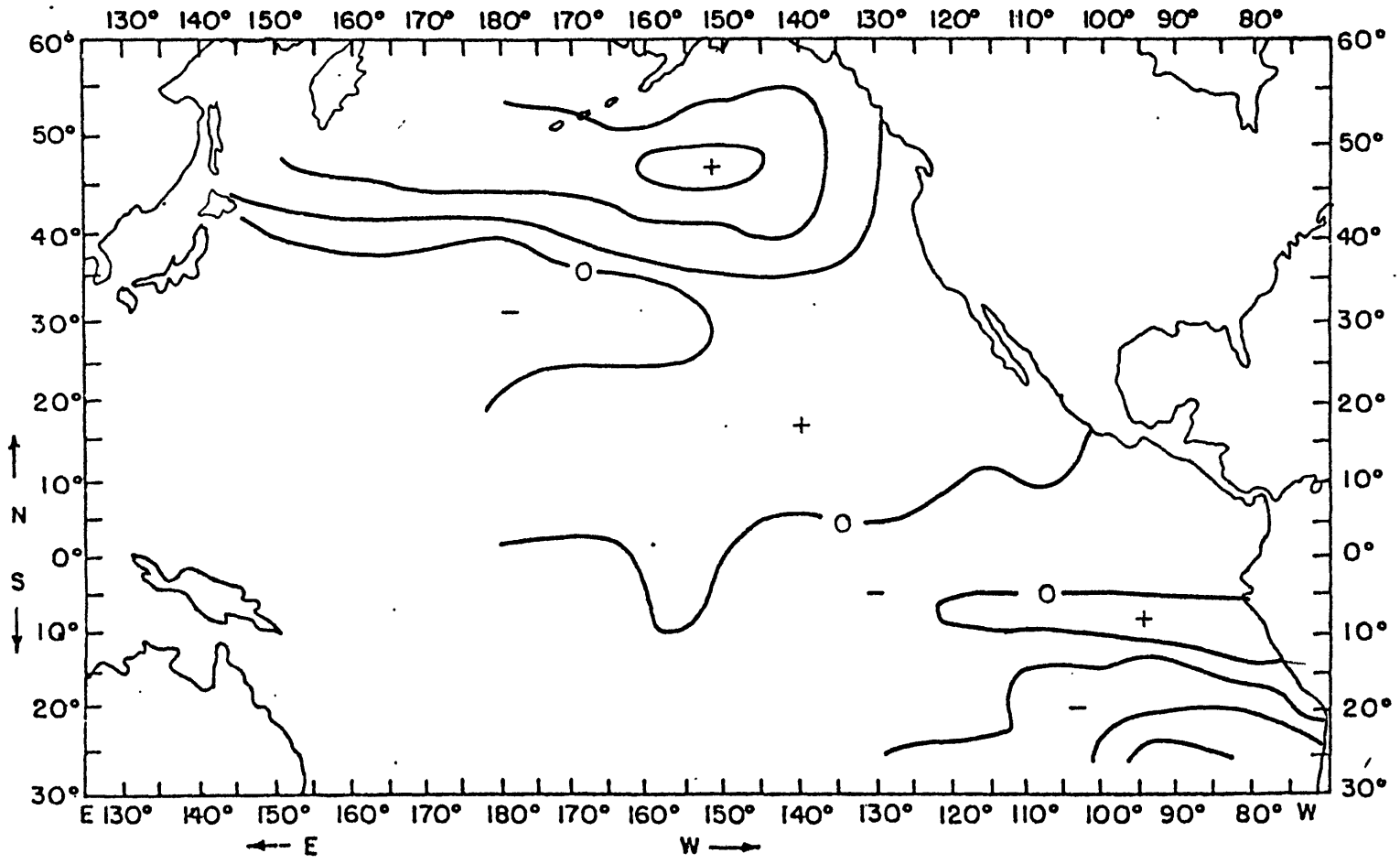


FIGURE 13. Pacific SST Anomalies EOF Mode 4

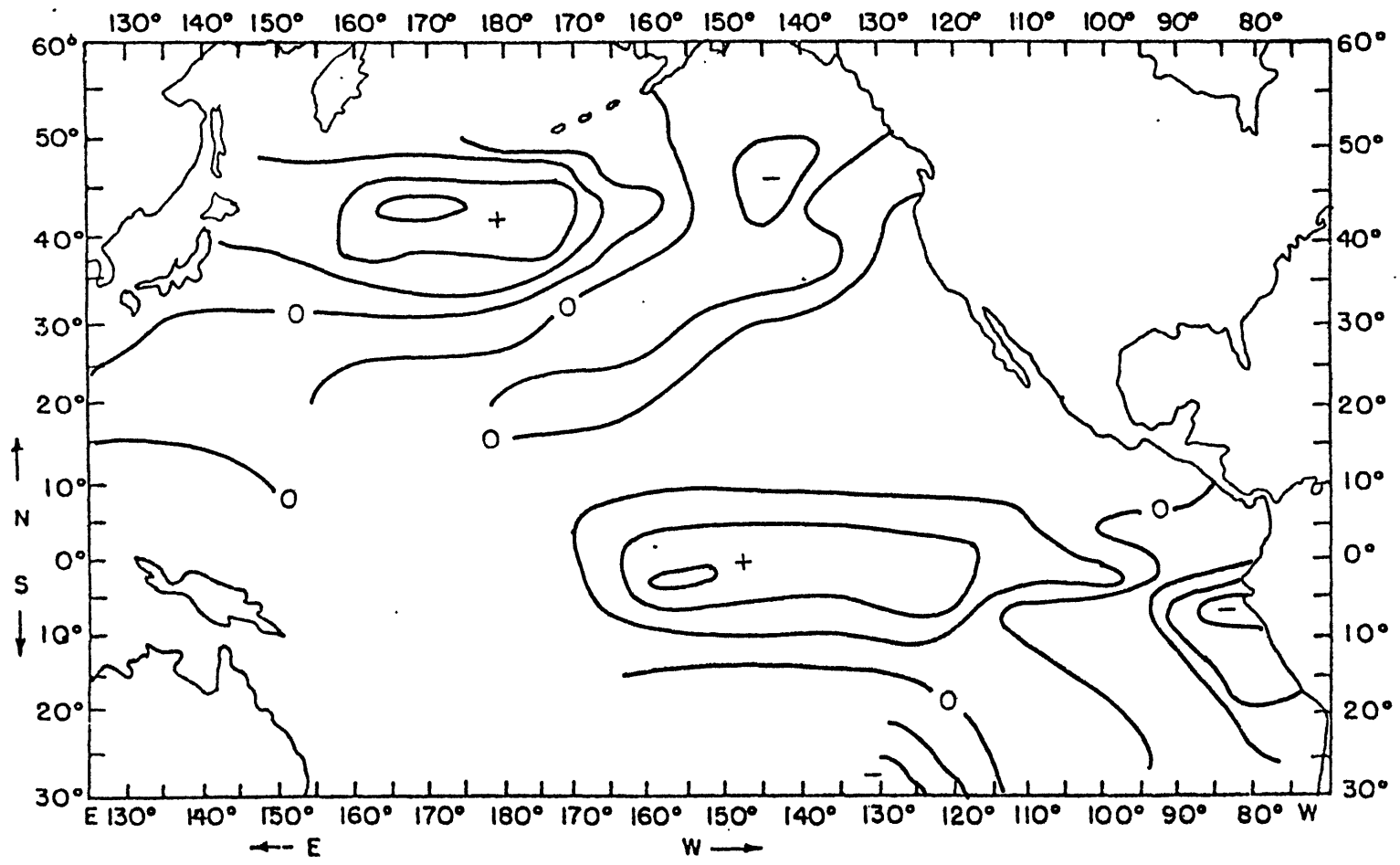


FIGURE 14. Pacific SST Anomalies EOF Mode 5

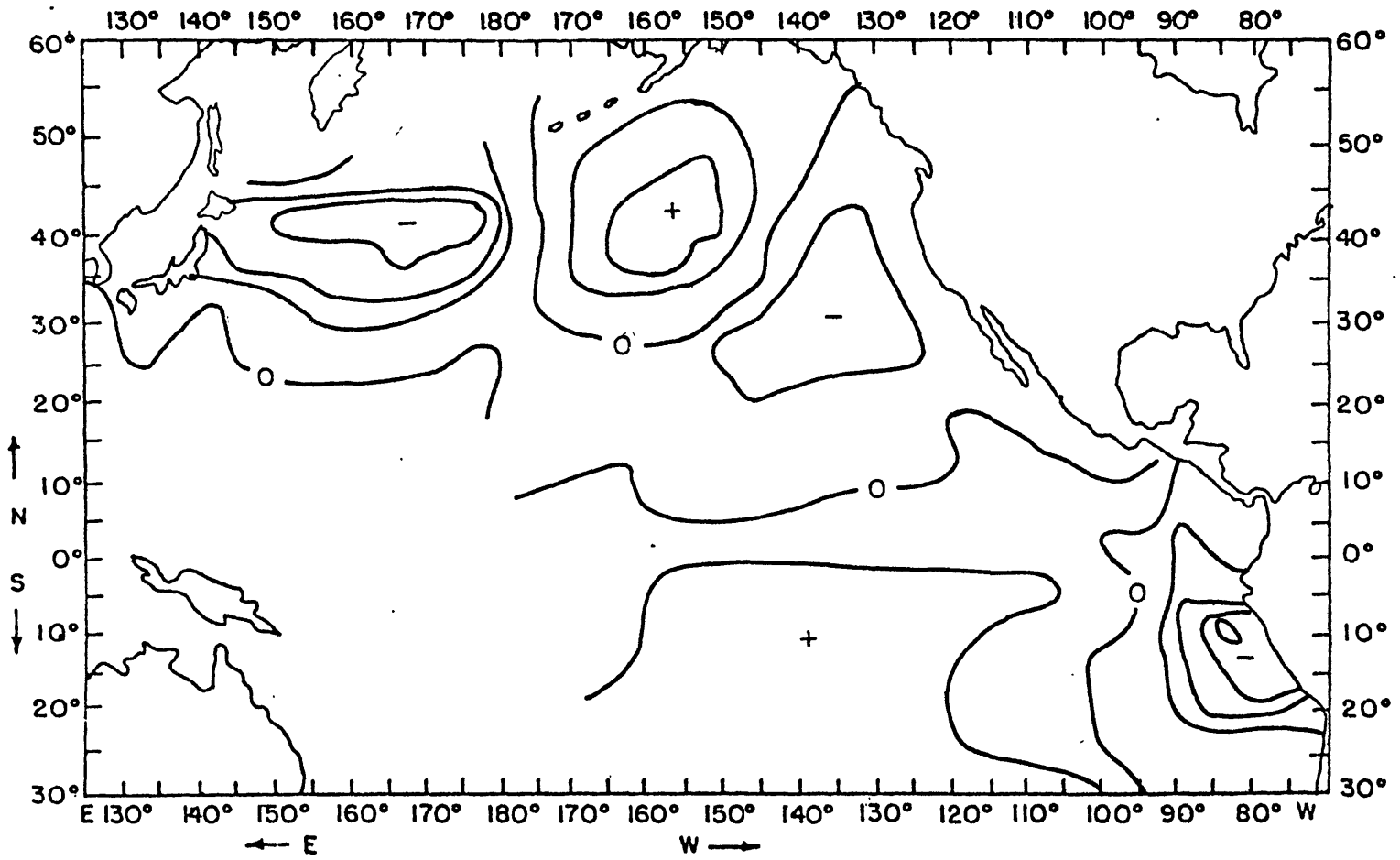


FIGURE 15. Pacific SST Anomalies EOF Mode 6

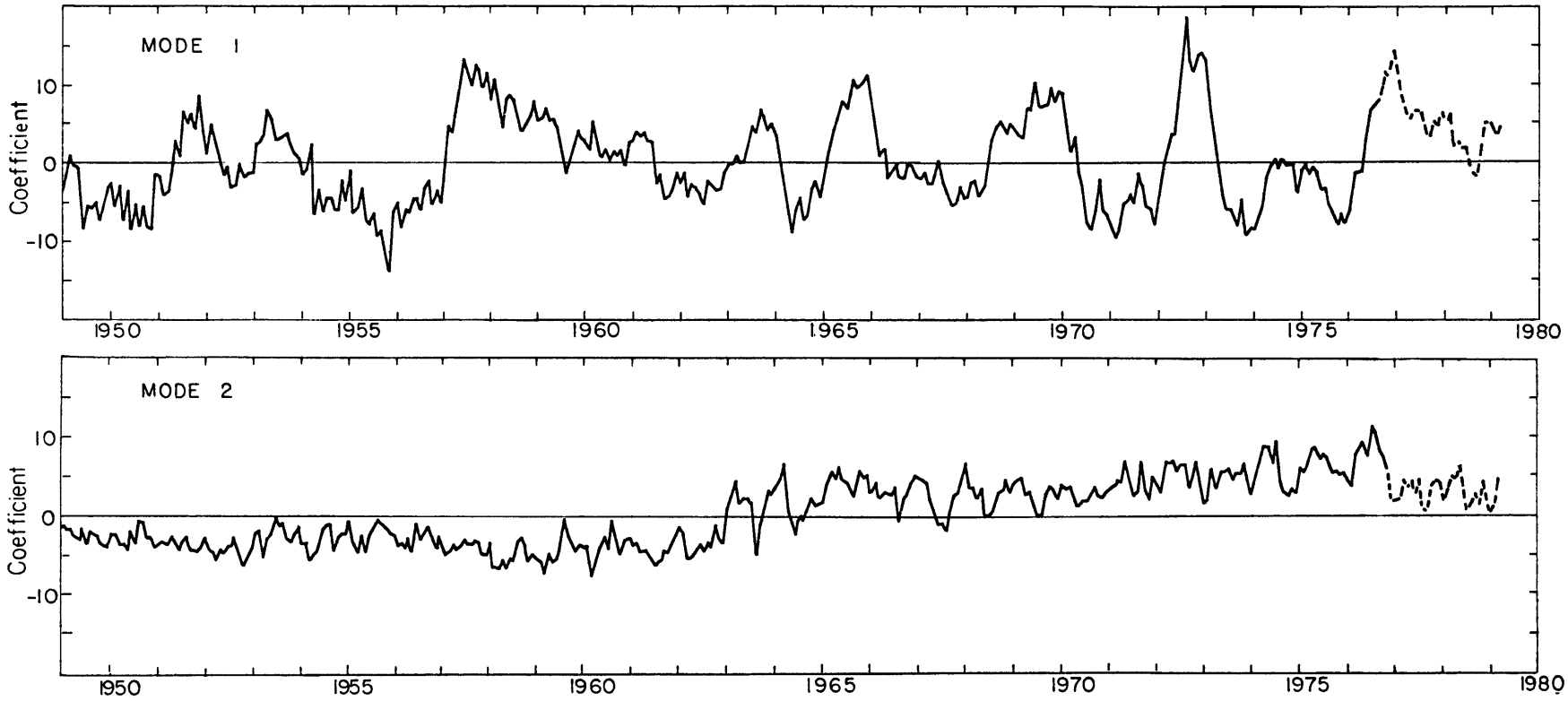


FIGURE 16. Time Series Coefficients for EOF Modes 1 & 2

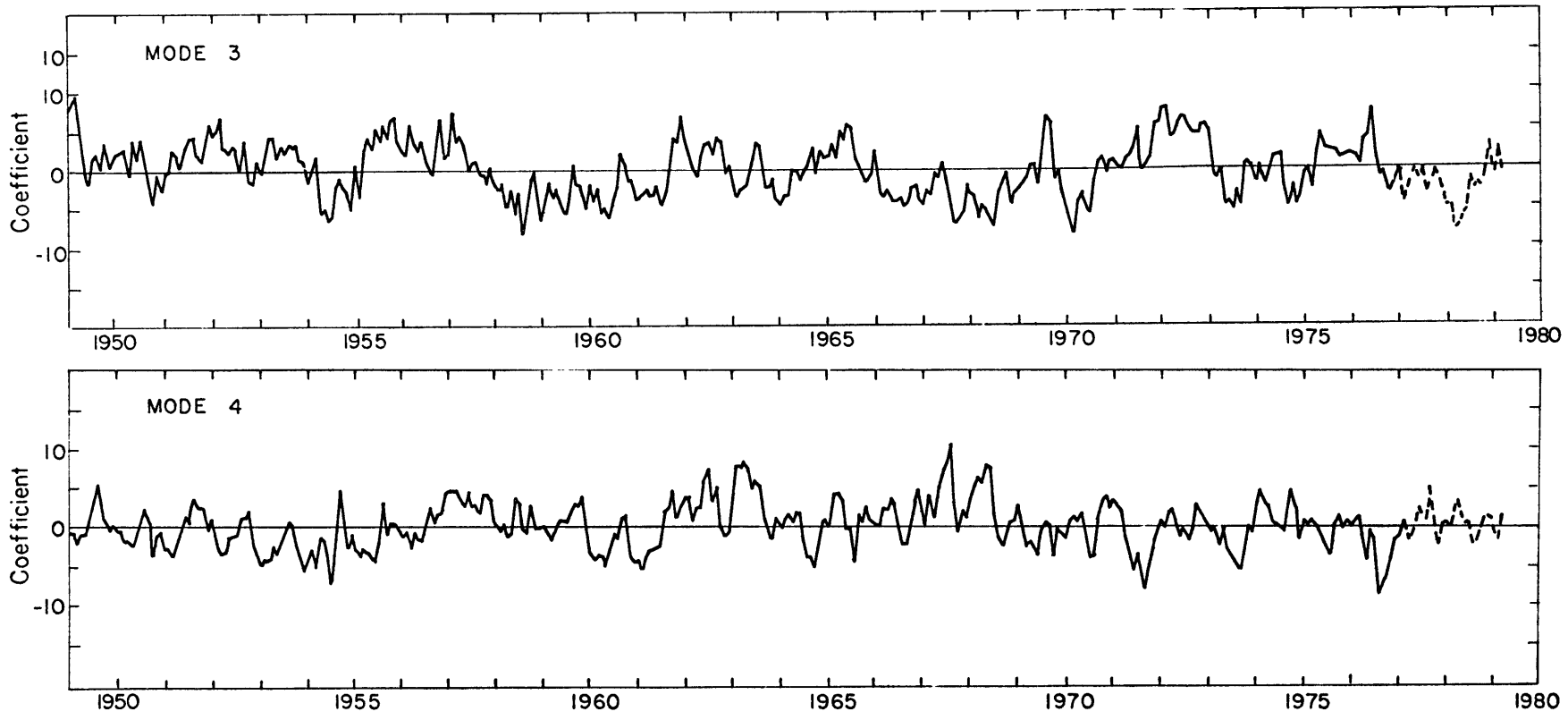


FIGURE 17. Time Series Coefficients for EOF Modes 3 & 4

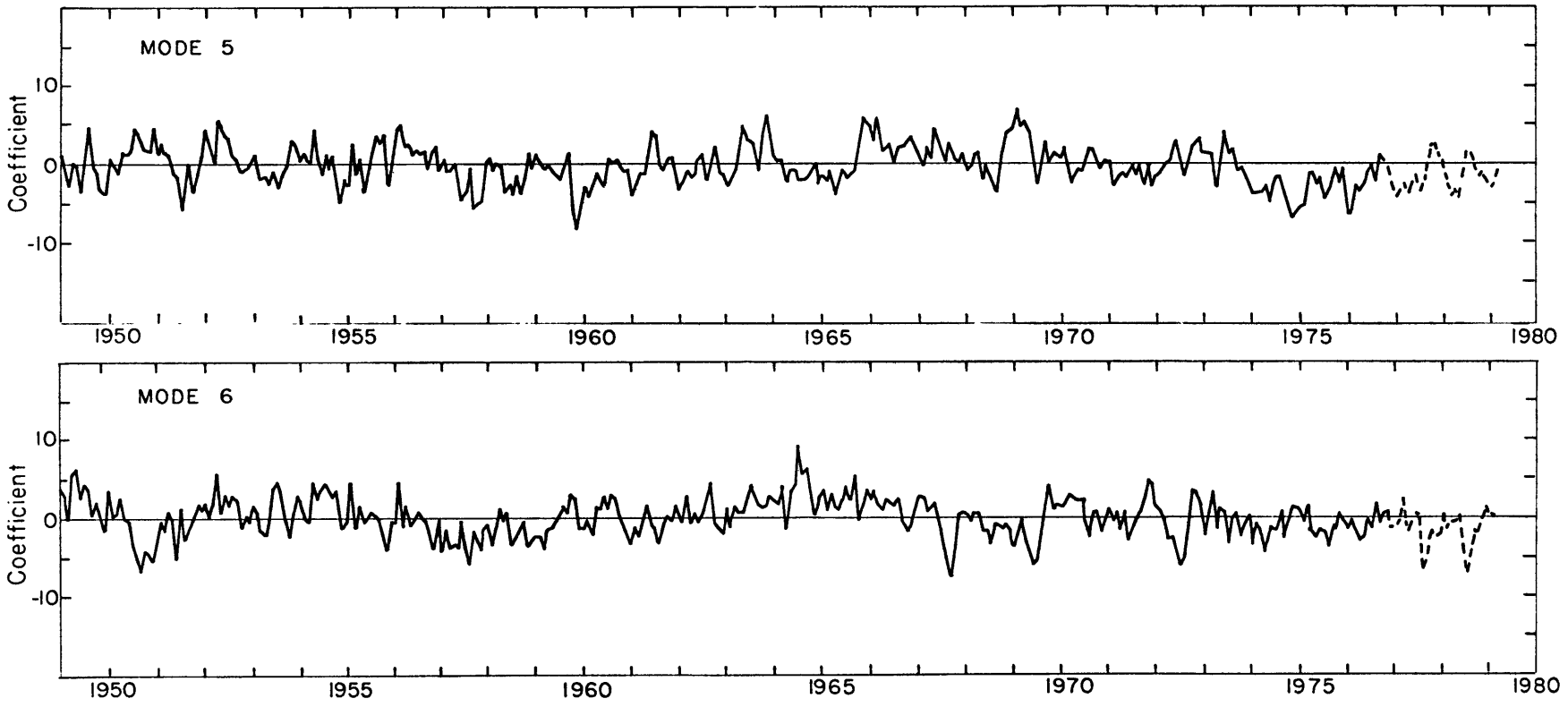


FIGURE 18. Time Series Coefficients for EOF Modes 5 & 6

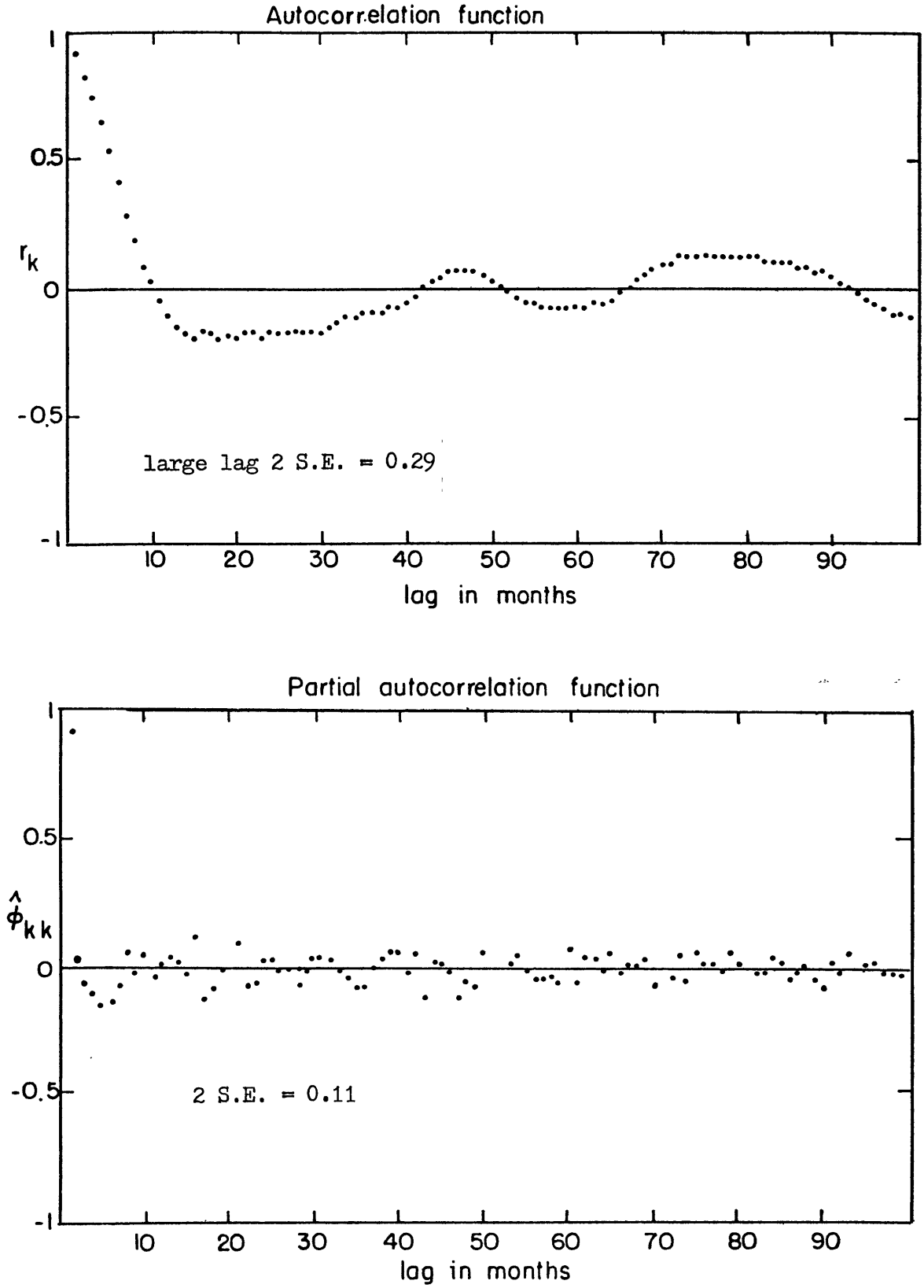


FIGURE 19. For EOF Time Series 1

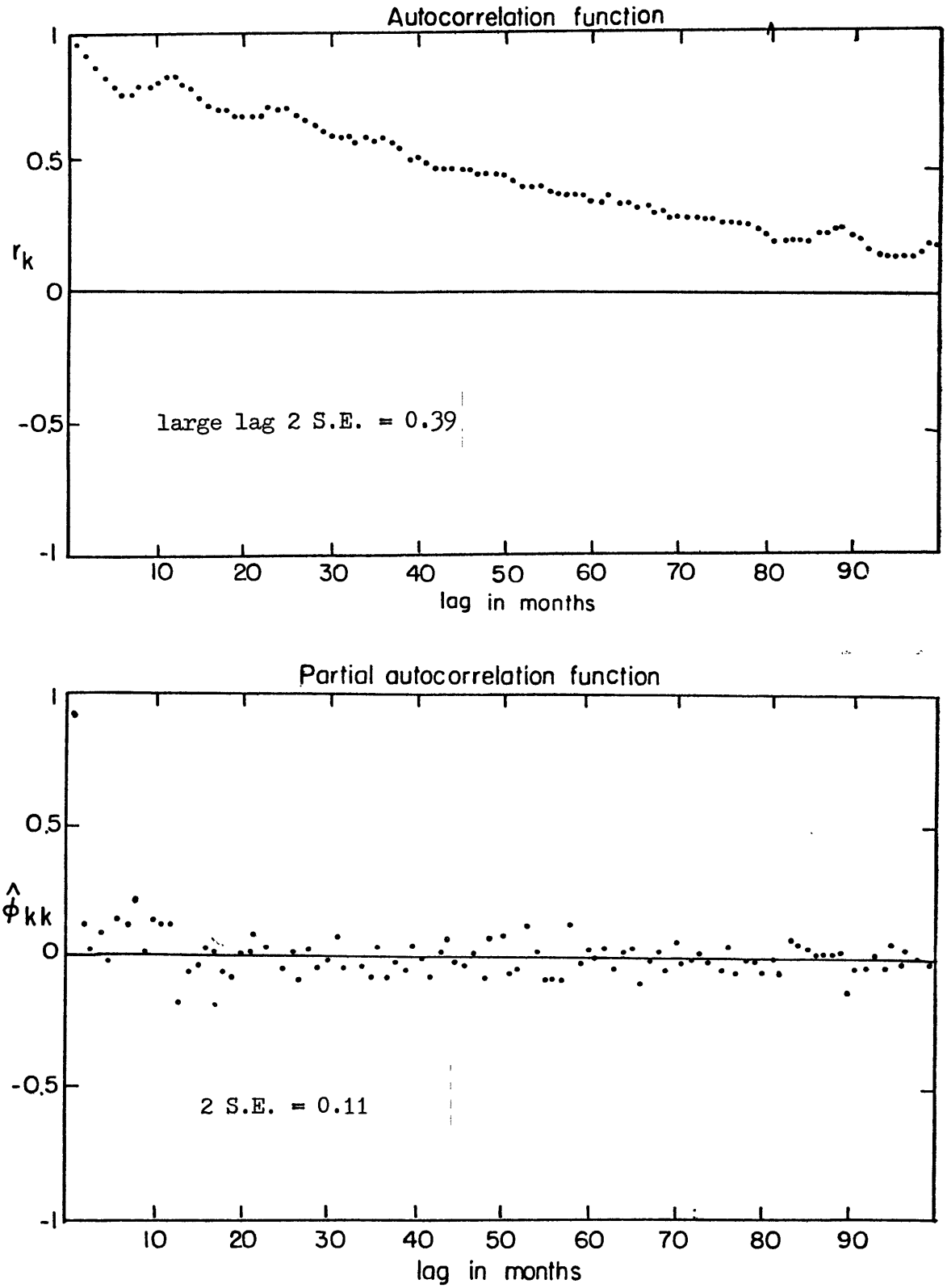


FIGURE 20. For EOF Time Series 2

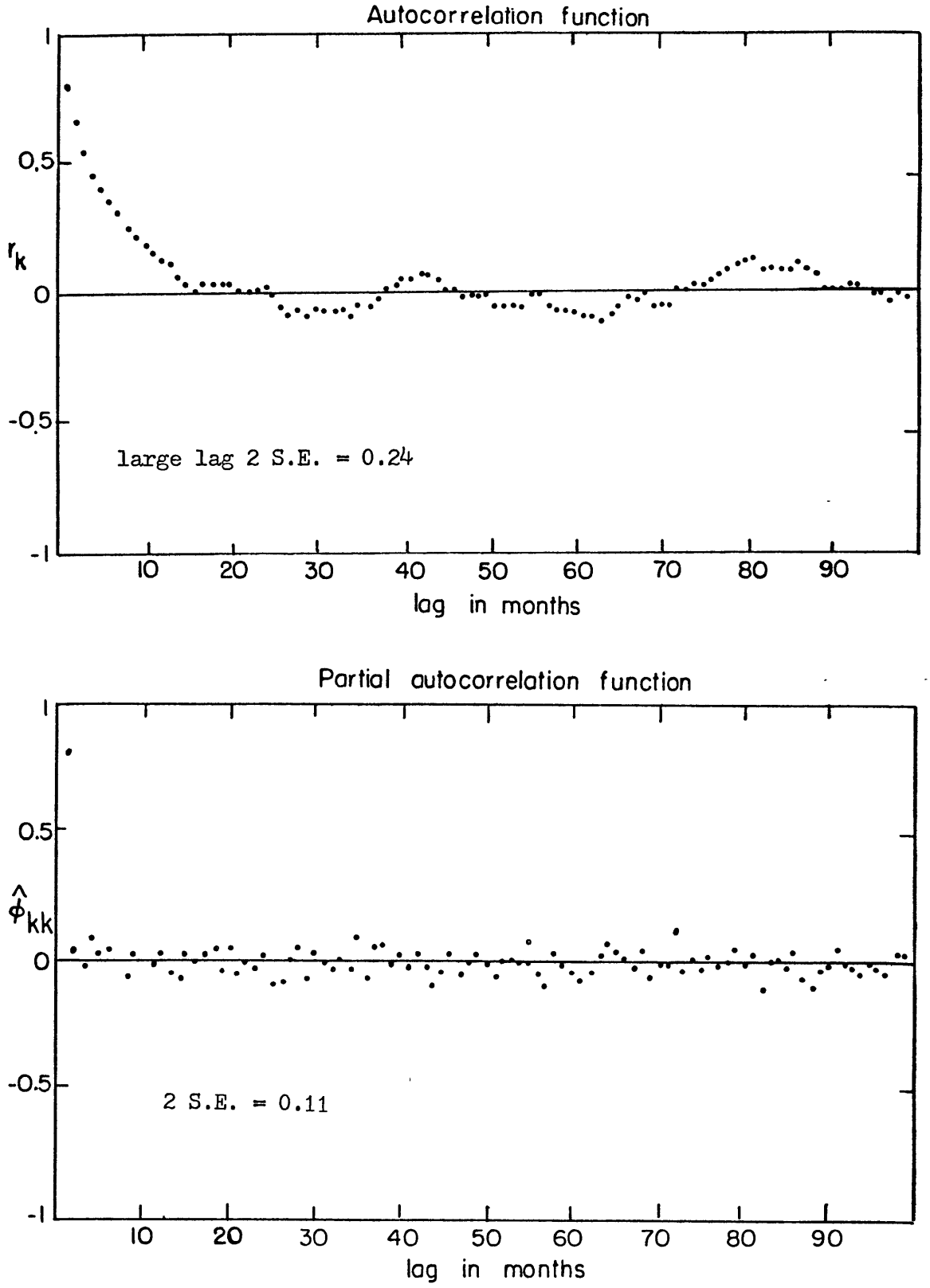


FIGURE 21. For EOF Time Series 3

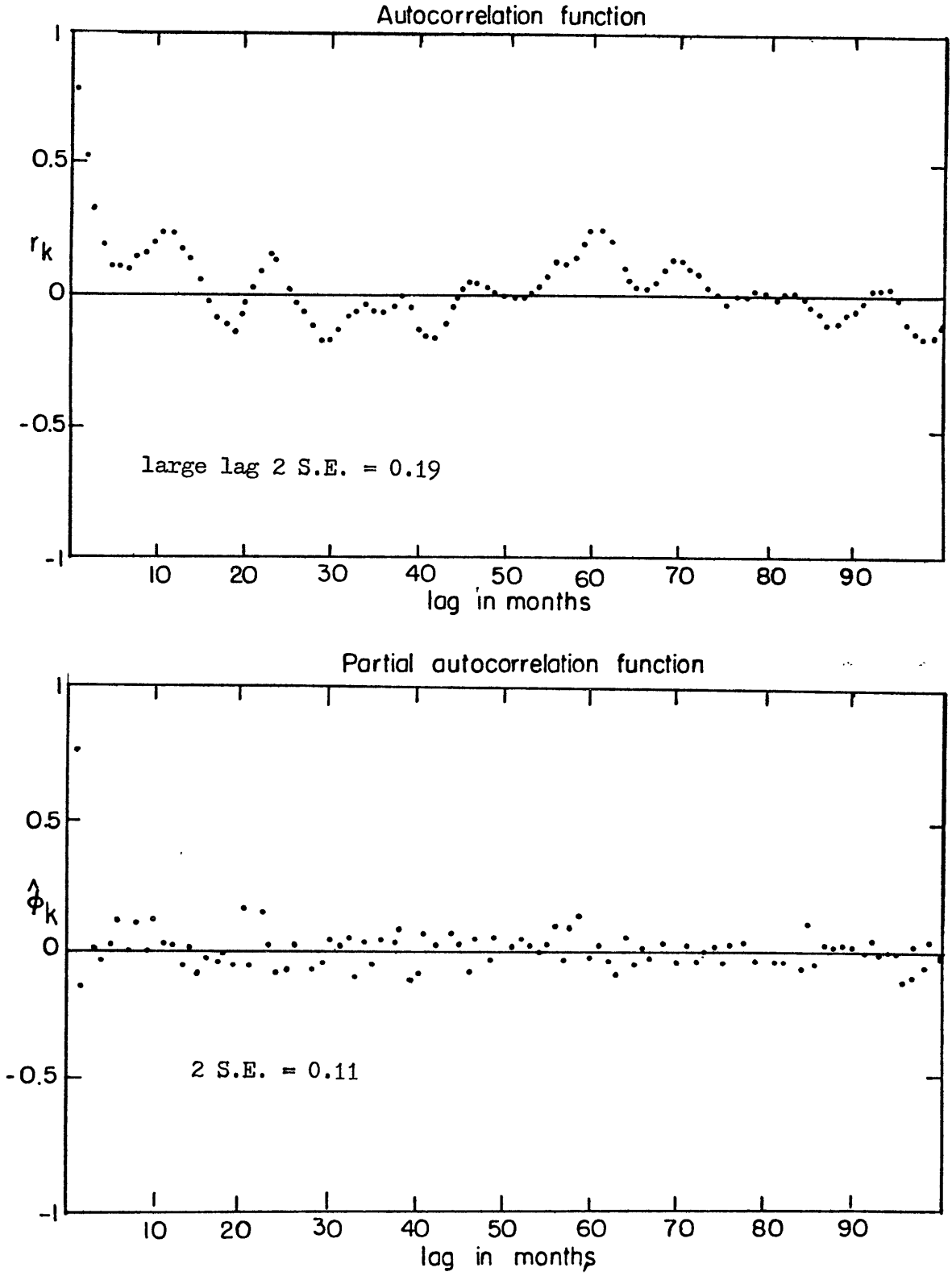


FIGURE 22. For EOF Time Series 4

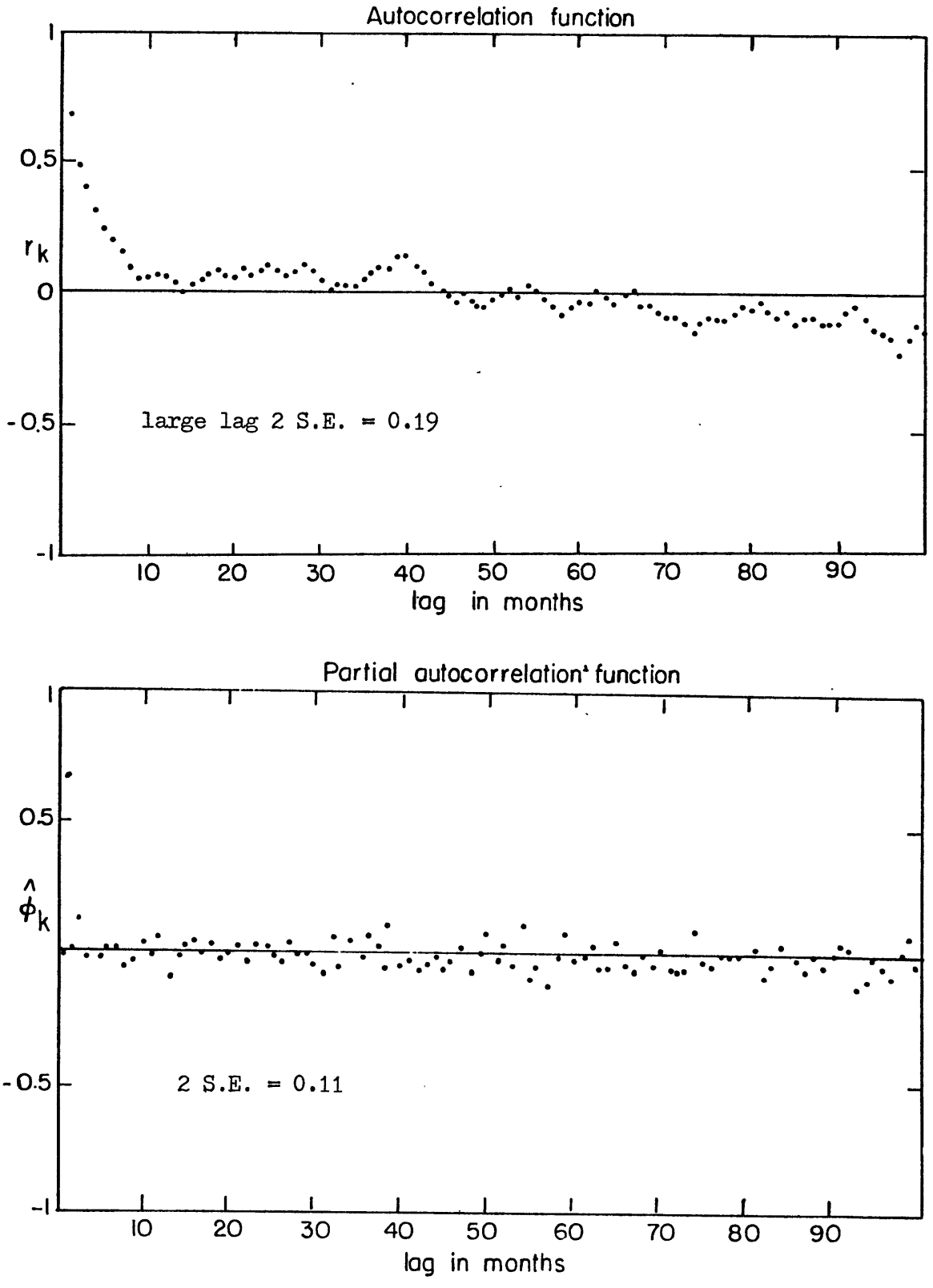
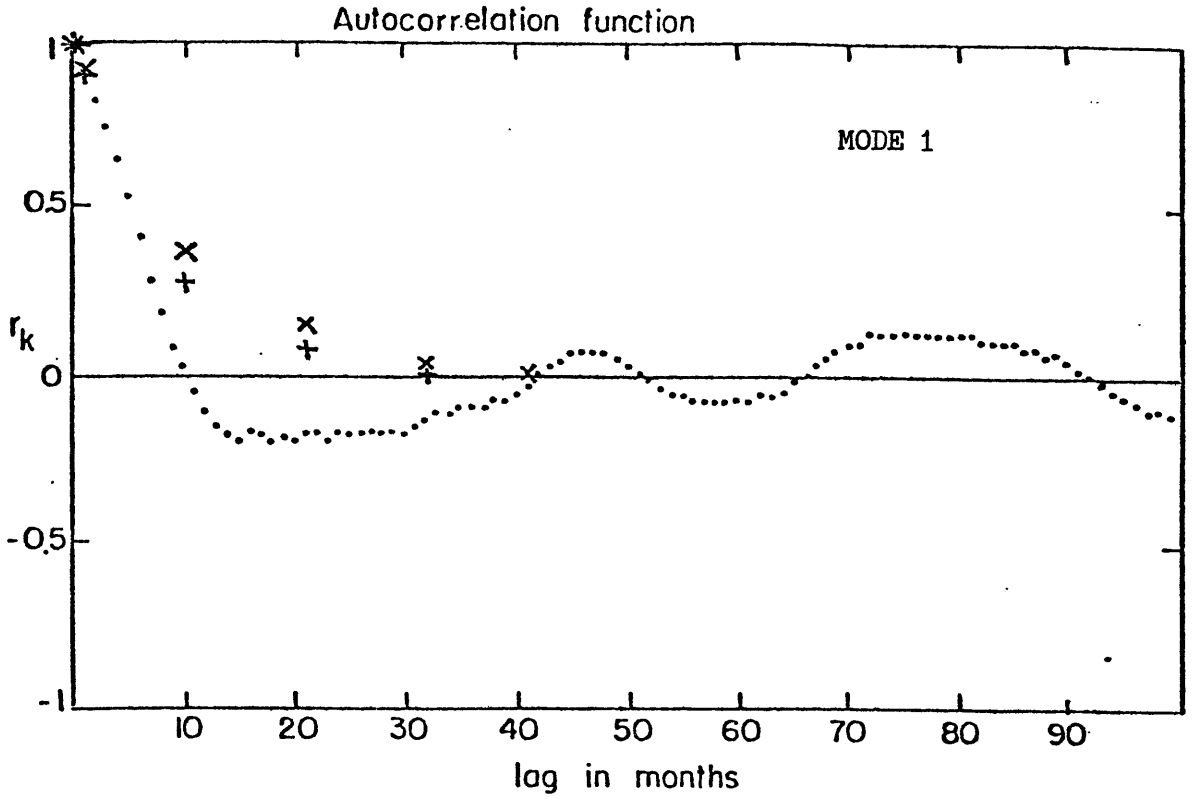


FIGURE 23. For EOF Time Series 5



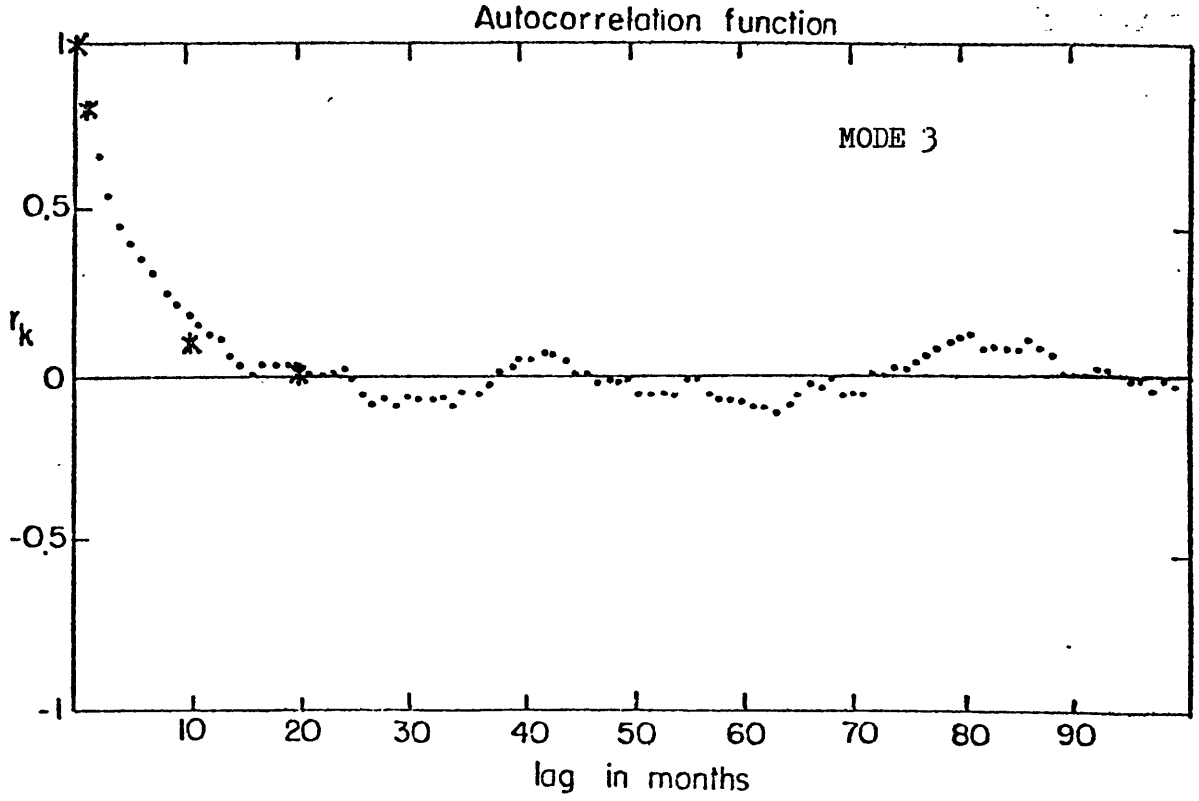


FIGURE 24. Autocorrelation for Fit to First Order Process

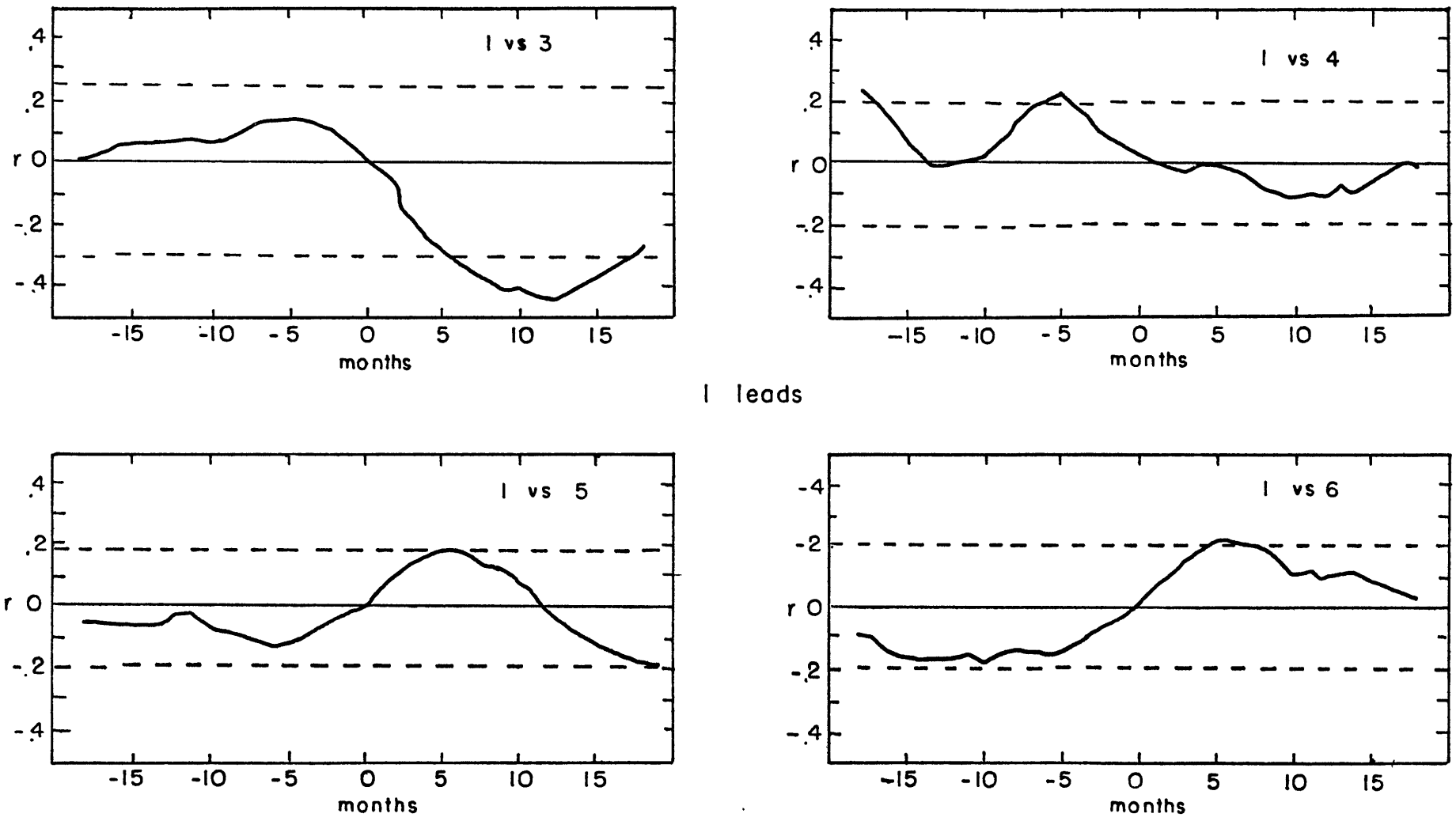


FIGURE 25. Cross correlations of EOF Time Series

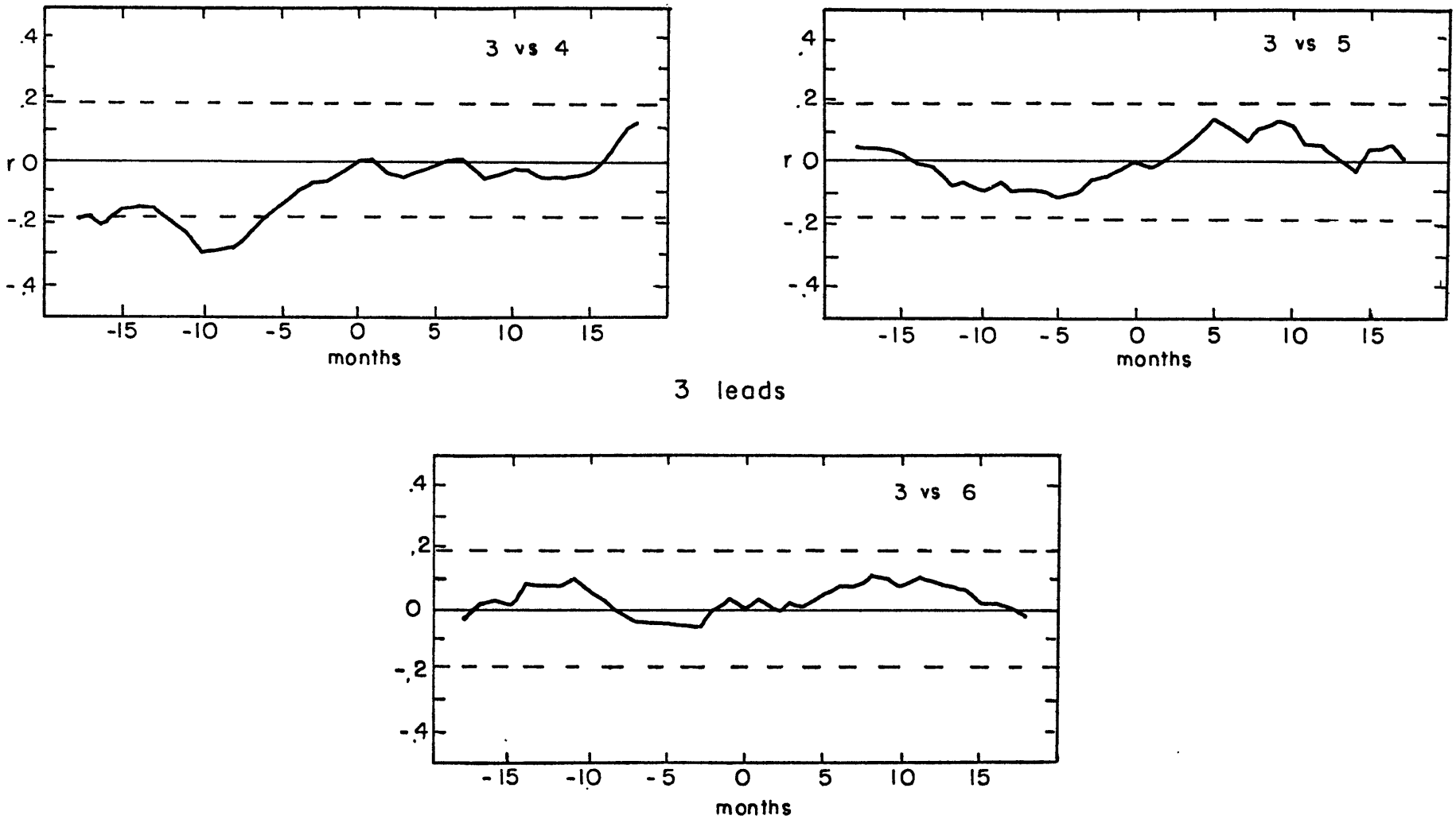


FIGURE 26. Cross correlations of EOF Time Series

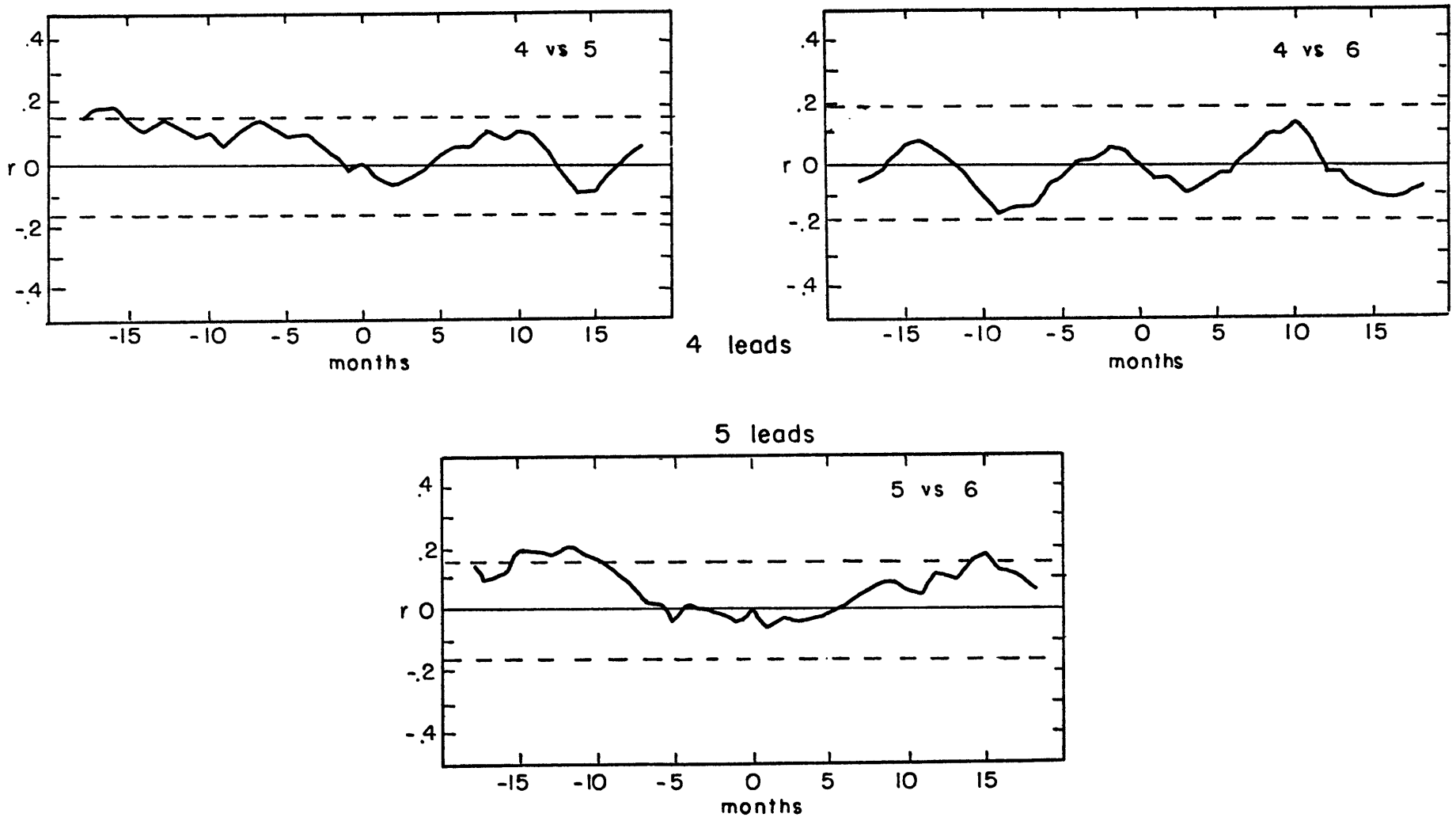


FIGURE 27. Cross correlations of EOF Time Series