Predictability of Sea Surface Temperature and Sea Level Pressure Anomalies over the North Pacific Ocean

RUSS E. DAVIS

Scripps Institution of Oceanography, University of California, San Diego 92903

(Manuscript received 6 November 1975, in revised form 24 November 1975)

ABSTRACT

Nonseasonal variability of sea level pressure (SLP) and sea surface temperature (SST) in the mid-latitude North Pacific Ocean is examined. The objective is examination of the basic scales of the variability and determination of possible causal connections which might allow prediction of short-term climatic (time scales between a month and a year) variability.

Using empirical orthogonal function descriptions of the spatial structure, it is found that SLP variability is concentrated in a few large-scale modes but has a nearly white frequency spectrum. SST variability is spatially complex (being spread over many spatial modes, some of which have small-scale changes) but is dominated by low-frequency changes.

The use of linear statistical estimators to examine predictability is discussed and the importance of limiting the number of candidate data used in a correlation search is underscored. Using linear statistical predictors, it is found that (A) SST anomalies can be predicted from SST observations several months in advance with measurable skill, (B) the anomalous SLP variability can be specified from simultaneous SST data with significant skill, thus showing the fields are related, and (C) future SLP anomalous variability cannot be predicted from SST data although previous SLP can be specified. The fact that previous SLP variability is better specified by SST data than is simultaneous SLP variability, coupled with a complete inability to predict future SLP anomalies, suggests that, in the region studied and on the time scales of a month to a year, the observed connection between SST and SLP variabilities is the result of the atmosphere driving the ocean.

1. Introduction

Consideration of the quantity of heat exchanged between ocean and atmosphere during the seasonal cycle suggests that on such time scales the ocean may play an important role in establishing the atmospheric climate. The practical importance of short-term climatic changes has provided impetus for a growing interest in such variability and the role played by the ocean in causing or modifying it. Because of the complexity of possible interactions and the large quantities of data required to examine them, most studies have concentrated on selected geographical regions and time periods. From such studies has grown the belief that available data disclose definite connections between ocean and atmosphere on climatic time scales and the hope that understanding these connections can lead to an improved ability to forecast atmospheric climate. A persuasive argument toward this point of view is given by Namias (1972a) along with a review of some of the evidence supporting it.

The present study is concerned with using statistical predictors to examine connections between thermal variability in the North Pacific Ocean and variations in the state of the overlying atmosphere. Attention is focused on time scales greater than one month and (because of limitations in the data) less than a year. Space scales comparable with the ocean basin are of interest and an emphasis is placed on the spatial patterns of variability as well as the patterns of ocean/atmosphere coupling.

The particular data considered here are 28-year records of sea surface temperature (SST) and sea level pressure (SLP) anomalies in the central North Pacific. An anomaly is here taken to be the difference between the observed value of a variable and the expected or average value found at the same location and time of year as the observation. The analytical techniques used are purely statistical ones aimed at determining scales of variability, establishing the existence of a connection between the two fields, and determining the cause of the connection. Specifically, three hypotheses are tested:

(A) Oceanic thermal development is sufficiently regular that SST anomalies can be predicted from prior SST observations.

(B) Atmosphere and ocean are sufficiently connected that SLP anomalies can be specified from contemporaneous SST observations.

(C) These relations can be combined to allow a prediction of SLP from prior SST observations.

The third hypothesis does not follow from the first two. It could easily be the case that oceanic thermal state is the result of both prior thermal state (through advection, perhaps) and atmospheric state (through a relatively rapid forcing such as wind-driven advection); then (A) and (B) would be true but (C) would not follow since the flow of information is from atmosphere to ocean, not the reverse.

The statistical techniques employed in testing the three hypotheses are based on the Gauss-Markov theorem which allows construction of that linear predictor with the minimum error variance relating a given set of data parameters to the quantity to be predicted. Utilization of the method depends on knowledge of the covariances or correlations of the various data parameters and the predictand. While the method is not restricted to linear systems, no other predictor, dynamical or statistical, can produce a lower error variance prediction from the same data parameters when the system is linear. The analytical techniques are similar to those used by Roden and Groves (1960) in their examination of North Pacific SST variability. The primary difference is that in the present study basin-wide patterns are considered, rather than values at particular locations; in view of the large-scale coherences of SST anomalies found by Favorite and McLain (1973) and Namias (1972a), it could be hoped that considering geographical patterns rather than time series from individual locations might lead to different conclusions concerning predictability and disclose different mechanisms.

The primary difficulty of establishing connections and cause/effect relationships from finite data sets using statistical methods is what might be called artificial predictability. This phenomenon is discussed in quantitative terms in the following section. Simply put, the problem is that given a finite number N of realizations of a predictand and a set of data parameters possibly influencing it, there will always appear to be a correlation between the predictand and some of the data variables, even if no true correlation exists. As the number M of data variables is increased the amount of apparent correlation between the data set and the predictand increases roughly as $(M/N)^{\frac{1}{2}}$. One is forced, then, to restrict by some *a priori* criterion the number of variables which are to be considered as data; if this is not done it will inevitably occur that an apparent connection between predictand and data will be found, even if none really exists. The technique of "screening," in which data poorly correlated with the predictand are excluded, does little to minimize artificial predictability since it is based on an *a posteriori* criterion.

In this study empirical orthogonal functions are used to describe the spatial structure of anomalies. This accomplishes two useful tasks. First, it provides a concise description of the distribution of variance between various spatial scales and determines how many different variables are required to describe anomaly patterns to a given accuracy. For example, it is found that one spatial pattern accounts for nearly half the variance of SLP anomalies. The second use of the empirical orthogonal function representation is that it provides an objective *a priori* criterion for limiting the number of data variables used in examining the predictability of SST and SLP. This follows from the fact that empirical orthogonal functions are the most efficient possible data representation in the sense that the amplitudes of the M dominant functions describe more variance of the field than any other possible M parameter description. A natural way of reducing the number of candidate data variables then, is to neglect the amplitudes of patterns which contribute little to the overall variance.

In addition to providing measures of predictability and the ability to make useful predictions, statistical predictors can be used to examine the mechanism. This is so because once the statistical model relating subsequent anomalies to various initial data has been found, it is possible to carry out experiments with the model. In this way it is possible, for example, to examine the role of advection in SST anomaly development by posing initial data describing an anomaly in the West Wind Drift region and seeing if the model "predicts" the anomaly to propagate as expected on the basis of the currents existent there.

The first topic considered below is the theory of Gauss-Markov predictors and the problem of artificial predictability. Then a basic description of the time and space scales of SST and SLP anomalies is given. Attention is then turned to examining the three predictability hypotheses above. Finally, some results of model experiments are given and their implications with regard to mechanism are discussed.

2. Linear statistical predictors

The questions of predictability posed by the three fundamental hypotheses stated in the Introduction all involve the ability to use the values of a set of data variables to predict (or specify) the value of some variable called the predictand. The technique employed here is to construct an efficient linear statistical predictor relating the data and predictand and then to test the hypotheses by the skill of these predictors. It is assumed that the data/predictand relationship is quasi-linear.

To examine the properties of the statistical techniques employed, the value of the predictand is taken as p, the estimate of p (the prediction) is taken as $\hat{p}$, and the values of the M data parameters are taken as d_n. Then

a linear predictor will take the form

$$\hat{p}=\sum_{n=1}^{M}\alpha_n d_n.$$

The measure of the quality of the prediction must be statistical. If the quality is measured by the mean square error then the optimal predictor should minimize

$$\langle(p-\hat{p})^2\rangle=\sum_n\sum_m\alpha_n\alpha_m\langle d_n d_m\rangle-2\sum_n\alpha_n\langle d_n p\rangle+\langle p^2\rangle,$$

where the angle brackets denote the expected or mean value. The optimum is achieved by taking

$$\alpha_n=\sum_m\langle d_n d_m\rangle^{-1}\langle d_m p\rangle, \tag{1}$$

where $\langle d_n d_m\rangle^{-1}$ is the inverse of the matrix $\langle d_n d_m\rangle$. The skill of the resultant prediction might be taken as

$$S=1-\frac{\langle(p-\hat{p}^2)\rangle}{\langle p^2\rangle}=\langle p^2\rangle^{-1}\sum_m\sum_n\langle d_n p\rangle\langle d_n d_m\rangle^{-1}\langle d_m p\rangle, \tag{2}$$

which represents the fraction of $\langle p^2\rangle$ which is predictable. It is a standard result of estimation theory, the Gauss-Markov theorem (Liebelt, 1967), that when the mean $\langle p\rangle$ of the predictand vanishes the above predictor produces the minimum error variance and, for the case of joint normally distributed data and predictand, the most probable value of p.

The difficulty in direct application of statistical predictors to finite data sets is that the true values of the covariances involved cannot be established owing to the limited number of realizations contained in the data set. A reasonable approach when the sample interval is fairly uniform is to replace the mean value by the sample mean, i.e., by

$$\{q\}=\frac{1}{N}\sum_{i=1}^{N}q(i),$$

where N is the number of realizations available in the data set. This causes no fundamental change in the derivation of the "optimal" predictor which minimizes the sample error variance $\{(p-\hat{p})^2\}$. If the data are redefined so that simultaneous data are uncorrelated over the data set (this can always be done if $N>M$, and is true of the empirical orthogonal functions used here) then because $\{d_n d_m\}=\delta_{nm}\{d_n^2\}$ the predictor is

$$\hat{p}=\sum_{m=1}^{M}\{d_m^2\}^{-1}\{p d_m\}d_m. \tag{3}$$

There are two important measures of the quality of this predictor, namely, the skill with which it hindcasts $\hat{p}$ over the data set from which the sample covariances were estimated, and the skill with which it predicts realizations which were not used in establishing the statistical estimates.

In the case of hindcasts we will find that the expected skill is enhanced by artificial predictability. For any given data set the hindcast skill is, similar to (2), given by

$$S_H=1-\frac{\{(p-\hat{p})^2\}}{\{p^2\}}=\sum_{n=1}^{M}\frac{\{p d_n\}^2}{\{p^2\}\{d_n^2\}}. \tag{4a}$$

The quantity of interest is the expected value of S_H or, equivalently, the mean value when averaged over all possible data sets of the same structure (size, sampling interval, etc.). Since the skill S_H depends on squares of apparent correlations it is evident that, even if the true correlations all vanish, any error in the sample correlations will, on average, lead to an increase of skill above the true value. As the number of realizations increases, the sample correlations will approach the true correlations and the expected skill will decrease until it reaches the value

$$S=\sum_{n=1}^{M}\frac{\langle p d_n\rangle^2}{\langle p^2\rangle\langle d_n^2\rangle}, \tag{4b}$$

which can be considered the true predictability of the process.

Accurate estimates of the artificial predictability, defined as $\langle S_H\rangle-S$, are in general difficult to make. Reasonable quantitative estimates, however, can be made by assuming that the various quantities have approximately joint normal distributions and that N is large enough that the sample statistics are approximately correct. Defining

$$\left.\begin{array}{lll}\langle p d_n\rangle=R_n, & \langle d_n^2\rangle=Q_n, & \langle p^2\rangle=Q_0\\ \{p d_n\}=R_n+r_n, & \{d_n^2\}=Q_n+q_n, & \{p^2\}=Q_0+q_0\end{array}\right\},$$

the expected hindcast skill is approximately

$$\langle S_H\rangle\approx\sum_{n=1}^{M}\left\langle(R_n^2+2r_nR_n+r_n^2)\times\frac{Q_n^2-Q_nq_n+q_n^2}{Q_n^3}\,\frac{Q_0^2-Q_0q_0+q_0^2}{Q_0^3}\right\rangle.$$

When N is large the dominant terms will be quadratic in the errors. Under the assumption of approximately joint normal distribution, these are easily computed. For example, by definition,

$$r_n^2=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}p(i)d_n(i)p(j)d_n(j)-\langle p d_n\rangle^2,$$

and the average of this is easily computed using the fact that the mean of the product of four joint normally distributed variables with zero mean is the sum of all

combinations of covariances so that

$$\langle r_n^2\rangle=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}[\langle p(i)p(j)\rangle\langle d_n(i)d_n(j)\rangle + \langle p(i)d_n(j)\rangle\langle p(j)d_n(i)\rangle].$$

It is usually the case that the correlations of p and d_n are small (except when the predictability is very high) so that it is permissible to neglect the second term as smaller than the first. Relating the covariances involved in the first term to the correlation functions defined by

$$\langle p(t)p(t+t')\rangle=Q_0C_0(t'),\quad \langle d_n(t)d_n(t+t')\rangle=Q_nC_n(t'),$$

produces a relatively compact representation for $\langle r_n^2\rangle$ when the sample means are taken from realizations which are equally spaced by the time interval Δt. If the record length $N\Delta t$ is long compared with the time scales of these autocorrelations, then

$$\langle r_n^2\rangle\approx Q_0Q_n\frac{1}{N}\sum_{i=-\infty}^{\infty}C_0(i\Delta t)C_n(i\Delta t)=Q_0Q_n\frac{\tau_n}{N\Delta t},$$

where

$$\tau_n=\sum_{i=-\infty}^{\infty}C_0(i\Delta t)C_n(i\Delta t)\Delta t$$

is an integral time scale determining the time period required to gain a new "degree of freedom" in the estimation of R_n. It is seen that the degrees of freedom from a record of N data is not N but rather $N\Delta t/\tau_n$.

The procedure outlined above can be used to estimate the size of all the error terms involved in the expression for the expected hindcast skill $\langle S_H\rangle$. When this is done it is found that the dominant term is the one involving r_n^2. Neglecting the smaller terms leads to the result

$$\langle S_H\rangle=S+\frac{1}{N}\sum_{n=1}^{M}\frac{\tau_n}{\Delta t}\approx S+\frac{M}{N}\frac{\tau}{\Delta t},\qquad(5)$$

where all the time scales have been approximated by τ. Even if the true predictability S is zero a finite skill of $O(M/N)$ will be found in hindcasting the data from which the predictor is constructed. It is this dependence on M which requires an *a priori* reduction in the number of data variables; if this is not done the artificial predictability will outweigh the true predictability S and will lead to erroneous conclusions about the degree of connection between the predictand and the data.

A second question bearing on the utility of statistical predictors concerns the skill expected when a predictor constructed from a certain finite number of realizations is used to predict a new, independent realization. It is interesting to consider, for this case, a slight generalization of the predictor (3), namely,

$$\hat{p}=K\sum_{m=1}^{M}\{d_m^2\}^{-1}\{pd_m\}d_m.\qquad(6)$$

It should be recalled that the original predictor, which is (6) with $K=1$, was chosen to minimize $\{(p-\hat{p})^2\}$, the mean square error over the same realizations used to determine the predictor. This does not necessarily optimize the prediction $\hat{p}^*$ of a realization p^* which is not a member of the set used to evaluate $\{pd_m\}$. The optimum predictor is the one which minimizes the expected forecast (as opposed to hindcast) error

$$\langle(p^*-\hat{p}^*)^2\rangle=\langle p^{*2}\rangle-2K\sum_{m=1}^{M}\left\langle\frac{\{pd_m\}}{\{d_m^2\}}p^*d_m^*\right\rangle + K^2\sum_{m=1}^{M}\left\langle\frac{\{pd_n\}^2}{\{d_n^2\}^2}d_n^*d_n^*\right\rangle,$$

where some terms have been dropped on the assumption that because $\{d_nd_m\}=0$ for $n\neq m$ that $\langle d_n^*d_m^*\rangle$ will be much smaller than $[\langle d_n^2\rangle\langle d_m^2\rangle]^{\frac{1}{2}}$ for $n\neq m$. Since the realizations from which the statistics were estimated (without asterisks) are independent of the forecast realizations (with asterisks) the averages involving both kinds of variables factor, leaving the expected forecast skill

$$\langle S_F\rangle=1-\frac{\langle(p^*-\hat{p}^*)^2\rangle}{\langle p^2\rangle}=2K\sum_{m=1}^{M}\frac{\langle pd_m\rangle^2}{\langle p^2\rangle\langle d_m^2\rangle} - K^2\sum_{m=1}^{M}\left\langle\frac{\{pd_m\}^2}{\{d_m^2\}^2}\right\rangle\frac{\langle d_m^2\rangle}{\langle p^2\rangle}.$$

Following along lines identical to those used to estimate the expected hindcast skill produces the result

$$\langle S_F\rangle=(2K-K^2)S-K^2\frac{M}{N}\frac{\tau}{\Delta t},\qquad(7)$$

where the variables S, τ and Δt are defined as in (5), which determined the expected hindcast skill. If forecasts were the point of interest, $\langle S_F\rangle$ could be improved by using the observed S_H and the resultant estimate of S from (5) to optimize the choice of K, which will then be smaller than unity.

Eqs. (5) and (7) give quantitative estimates of the prediction skill of realizable predictors in terms of the true predictability S which could be achieved only if precise statistics were known. The conclusion is the obvious one that the predictor performs artificially well in hindcasting the realizations from which it was constructed (by an amount proportional to M/N) and that the same predictor when used on independent data does not perform as well as is theoretically possible by a similar amount. This brings out clearly the balances involved in using statistical predictors both to verify predictability of a system and to make forecasts from

new data:

1) Since S is the sum of positive quantities the true predictability is always increased by using more prediction parameters, i.e., increasing M.
2) Increasing M always improves the ability to hindcast, if only because of increased artificial predictability. Verification of true predictability requires achieving a hindcast skill considerably in excess of $M\tau/N\Delta t$.
3) Although increasing M improves hindcasts it can easily harm forecasting skill if the increase in $M\tau/N\Delta t$ exceeds the increase in S. It is quite possible, for this reason, to make statistical predictors which are worse than predicting the mean value, $\hat{p}=\langle p\rangle$, since $\langle S_F\rangle$ can be driven negative by the M/N term.
4) Given the hindcast skill it is possible to estimate the true predictability S and, using this estimate, to optimize the forecast predictor (6) by choosing K to maximize $\langle S_F\rangle$.

It must be emphasized that these considerations pertain only if the data variables d_n are chosen by an *a priori* criterion. The origin of the M/N terms is the fact that any variables will appear correlated in a finite data set, even if they are not truly correlated. Choosing as data those variables with large apparent correlations may simply involve finding the large artificial correlations; given a large number M of uncorrelated data the m data with highest apparent correlation will produce an artificial predictability closer to M/N than to m/N.

It is worth pointing out that the predictability skill S as given by (2) or, for the case of uncorrelated data, by (4) provides a convenient measure of the total correlation between the variable p and the set of data d_n. It can easily be computed once the individual covariances have been estimated. In this way determination of the predictability is easily accomplished without ever carrying out a prediction. It must be recalled that any such estimate is actually an estimate of $\langle S_H\rangle$ and will, on average, be larger than the true value. Further, it must be noted that S is a weighted sum of squares of correlations so that $S^{\frac{1}{2}}$ is most nearly the total correlation.

3. Scale description

The data used in this study were 28-year records (1947–74) of monthly mean values of SST and SLP in the central North Pacific (20°N to 55°N) on 5° latitude by 10° longitude grids. The construction of these basic monthly maps, which were made available to me through the generosity of Jerome Namias of Scripps, and the data grids are described more fully in Appendix A.

The first analysis step was the preparation of anomaly maps. A mean value was found for each grid point and month of the year by averaging over the 28 years. The resulting maps of monthly normal states showed no significant departure from the seasonal norms of SLP found by O'Connor (1961) and of SST found by Robinson (1975). The analysis here concerns anomalies defined as the departures of the observations from these monthly normal fields.

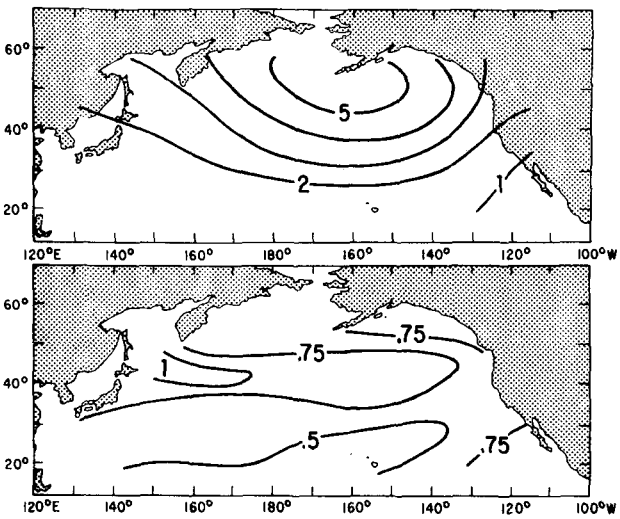

FIG. 1. Standard deviation of SLP anomalies (mb), upper, and SST anomalies (°C), lower. Anomalies are departures from monthly normal values. The anomaly variances are averaged over all months of the 28-year record.

The mean square anomaly distribution shows interesting differences between the character of SST and SLP variability. The maps of mean square anomaly in each month of the year may be characterized qualitatively as follows[1]:

(a) In each month the region of maximum SST variability is a band around 40–45°N extending eastward from Japan with variability decreasing to the east. The most variable locations are the four grid points with latitudes 40 or 45°N and longitudes between 150 and 165°E (south of Kamchatka) and this variability shows a broad maximum through the cooling season centered around November. The variability in the eastern part of the active band (around 160°W) is greatest in the summer with a secondary maximum in January. The centers of activity are shown in Fig. 1 in which the standard deviation of SST anomalies (averaged over the entire year) are plotted.

(b) The geographical pattern of SLP variance shows, in every season, a large-scale maximum centered near the Aleutians. While the precise center of activity moves relatively little through the year the magnitude of activity shows a pronounced broad maximum in the cooling season centered around November; during this season the center of variability is generally to the

[1] General conclusions are labeled by lower case letters which will be used in subsequent reference to them.

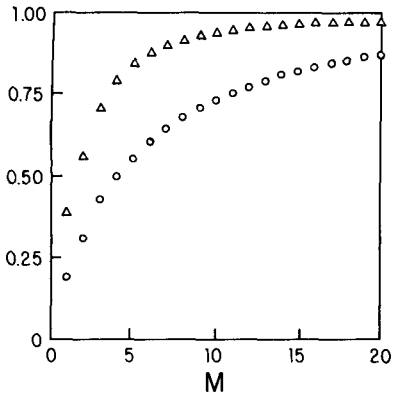

FIG. 2. The fraction of total SST (circles) and SLP (triangles) anomaly variance accounted for by the first M empirical orthogonal functions.

south of the minimum in the Aleutian low. The yearly mean of pressure variability is shown in Fig. 1.

The second step of analysis involved representing the monthly anomaly maps of SST and SLP in terms of the dominant empirical orthogonal functions of these fields. The properties and details of this kind of representation are discussed in detail by Lorenz (1959) and in Appendix B. The full power of this technique can be appreciated only from such discussions but the major virtue in the present context is that it allows an efficient approximation of the data using less parameters than contained in the full set of complete maps. For example, the maps $T(x,t)$ of SST for month t can be approximated by

$$\tilde{T}(x,t)=\sum_{m=1}^{M}\theta_m(t)T_m(x), \tag{8}$$

where M is a number less than the number of grid points x. Choosing the functions of grid position $T_m(x)$ as the dominant M empirical orthogonal functions leads to the smallest sample mean square error

$$\sum_x\{(T-\tilde{T})^2\} \tag{9}$$

which can be obtained from any set of M functions. The additional properties of the empirical orthogonal functions $T_m(x)$ [and the analogous functions $P_m(x)$ for the SLP field] are that the functions are orthogonal according to

$$\sum_x T_n(x)T_m(x)=\delta_{nm},$$

and that the amplitudes are uncorrelated over the data set, that is

$$\{\theta_n\theta_m\}=\delta_{nm}\{\theta_m{}^2\},$$

with an equivalent relation holding for the SLP amplitudes $\pi_n(t)$.

The criterion used to define the empirical functions is admittedly *ad hoc*. Equal weighting of the square error from each grid point in (9) could be replaced by any other weighting, such as a weighting proportional to the physical area of the associated 5° by 10° box, and slightly different empirical functions would result. It is a basic hypothesis of this analysis that the appropriate measure of variance is the sum of squares over all grid points and from this assertion the particular error criterion (9) follows, as does the hypothesis that the least "important" modes of variability are those associated with the smallest mean square amplitude. It will be seen later that the latter hypothesis is not strictly correct and that the development of the dominant SST patterns can be influenced by empirical functions which contribute little to the overall SST variance.

Computation of the empirical orthogonal functions involves finding the eigenvectors (the empirical functions) and the eigenvalues (the mean square amplitudes) of the covariance matrix of simultaneous data from different locations. The matrix used was the sample covariance obtained by averaging all months together into an overall mean. The number of pairs averaged was slightly less than 336 for certain elements of the SST covariance matrix because some data were missing.

Orthogonality of the empirical functions makes computation of the amplitudes $\theta_n(t)$ and $\pi_n(t)$ straightforward except in the case of missing data. In this case an objective amplitude estimate was used. The amplitude was taken as

$$\tilde{\theta}_n(t)=\beta_n\sum_x T(x,t)T_n(x),$$

where the sum over x includes only those points where T is known; when there are no missing values the choice $\beta_n=1$ produces the exact amplitude. The mean square error of this amplitude estimate is

$$\langle(\tilde{\theta}_n-\theta_n)^2\rangle=\beta^2\sum_{m\neq n}\langle\theta_m{}^2\rangle\gamma_{nm}{}^2+[1+\beta(\gamma_{nn}-1)]^2\langle\theta_n{}^2\rangle,$$

where

$$\gamma_{nm}=\sum_{x^*}T_n(x)T_m(x)$$

with the sum over x^* being taken over the missing data points. Minimization of this expected square error leads to

$$\beta_n=\frac{(1-\gamma_{nn})\langle\theta_n{}^2\rangle}{(1-\gamma_{nn})^2\langle\theta_n{}^2\rangle+\sum_{m\neq n}\langle\theta_m{}^2\rangle\gamma_{nm}{}^2}.$$

The number of missing points was small enough that the expected error of the dominant amplitudes was small. In no case did the relative expected error in estimating any of first 15 amplitudes exceed 30% and in only three months did it exceed 10%.

Some of the fundamental difference between SST and SLP variability is disclosed through examination of the basic empirical function description. The most obvious

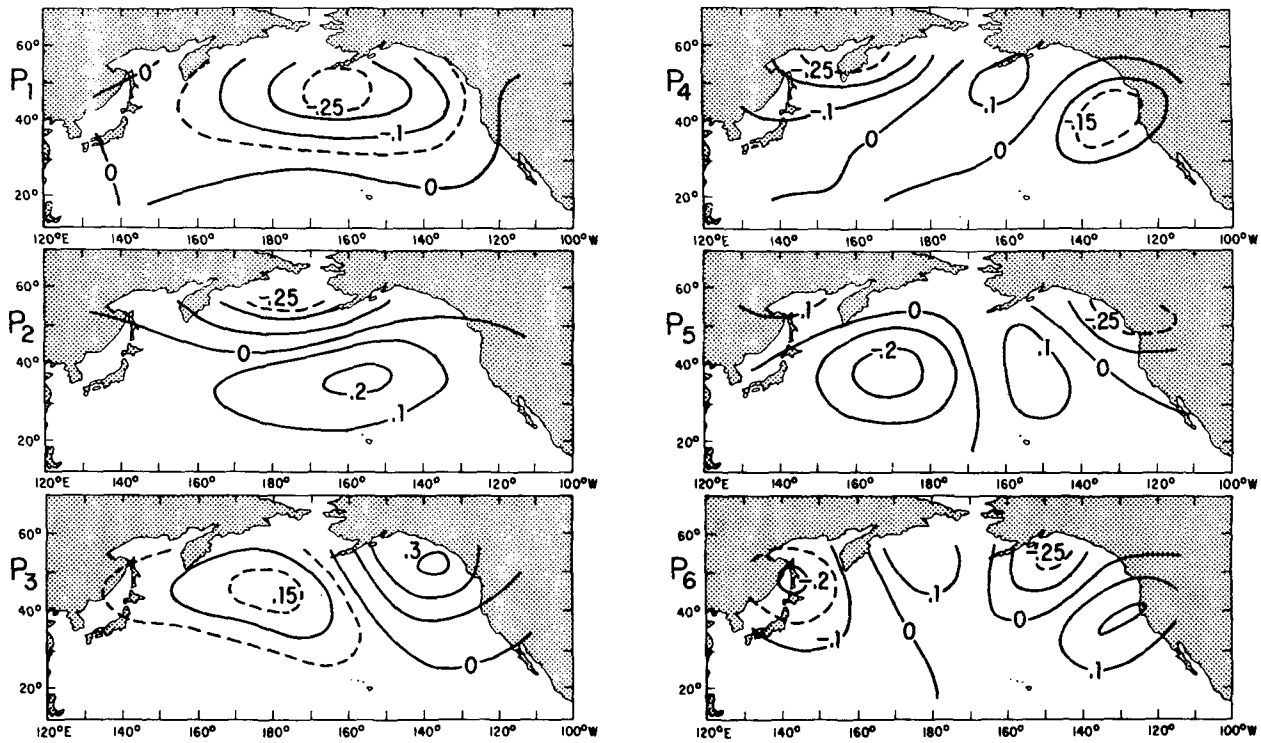

FIG. 3. The six principal empirical orthogonal functions, P_1–P_6, describing SLP anomalies.

differences are as follows:

(c) A relatively few large-scale patterns (empirical functions) describe most of the month to month SLP variability whereas many more patterns, some of which contain rapid variations over small scales, are required to describe SST variability to the same accuracy. This is shown by Fig. 2, a plot of the fraction of variance associated with the first few empirical functions, and Figs. 3 and 4 which show, respectively, the dominant SLP and SST functions.

(d) As the spatial scales of SLP variability are long compared with those of SST so are the time scales of SST anomalies significantly longer than those of SLP. This anticipated result is demonstrated in Fig. 5 which shows the frequency spectra of θ_1 and π_1, the amplitudes of the dominant SST and SLP patterns. The spectra of the dominant six SLP modes are similar to that of P_1 and those of the dominant 10 SST modes are similar to that of T_1 except for a general increase in the relative importance of higher frequencies in the higher modes.

The spectra presented in Fig. 5 were computed from Fourier transforms of four overlapping blocks of 8-year length taken from the 28-year records and the spectral estimates for frequencies above one cycle per year represent a block average over three adjacent frequency bands. The composite spectrum obtained by averaging the spectra of the dominant 10 SST modes shows no significant finestructure and discloses no tendency toward a decrease of spectral intensity as frequency approaches zero; the lower spectral density in the zero frequency band shown in Fig. 5 is neither statistically significant nor a general feature. Similarly, the composite spectrum of the six dominant pressure patterns discloses neither significant finestructure nor evidence of any departure from a smooth spectrum, decreasing very slowly with increasing frequency.

Some final scale-descriptive observations can be made by comparison of the spatial distributions of variance and the empirical functions. The first pressure pattern P_1 so dominates the SLP variability that it rather strongly resembles the distribution of variance itself. In contrast, the dominant SST pattern bears little resemblance to the distribution of variance; while the greatest SST variability is concentrated south of Kamchatka the two dominant patterns seem to be descriptive of the central portion of the ocean and the subarctic gyre, respectively. The large variance associated with the active region south of Kamchatka seems to be divided among several different modes (T_3, T_6, T_7, T_8, for example) indicating that the energetic variability in this region does not involve uniform SST changes over the entire area but is of a smaller scale, probably not adequately resolved by the present grid. The conclusion that the central ocean variability is of a different character than that in the active western ocean is supported by the difference in the annual cycles of anomaly variance noted in (a).

In short, SST and SLP anomalies are characterized by a broad distribution of scales which are not well

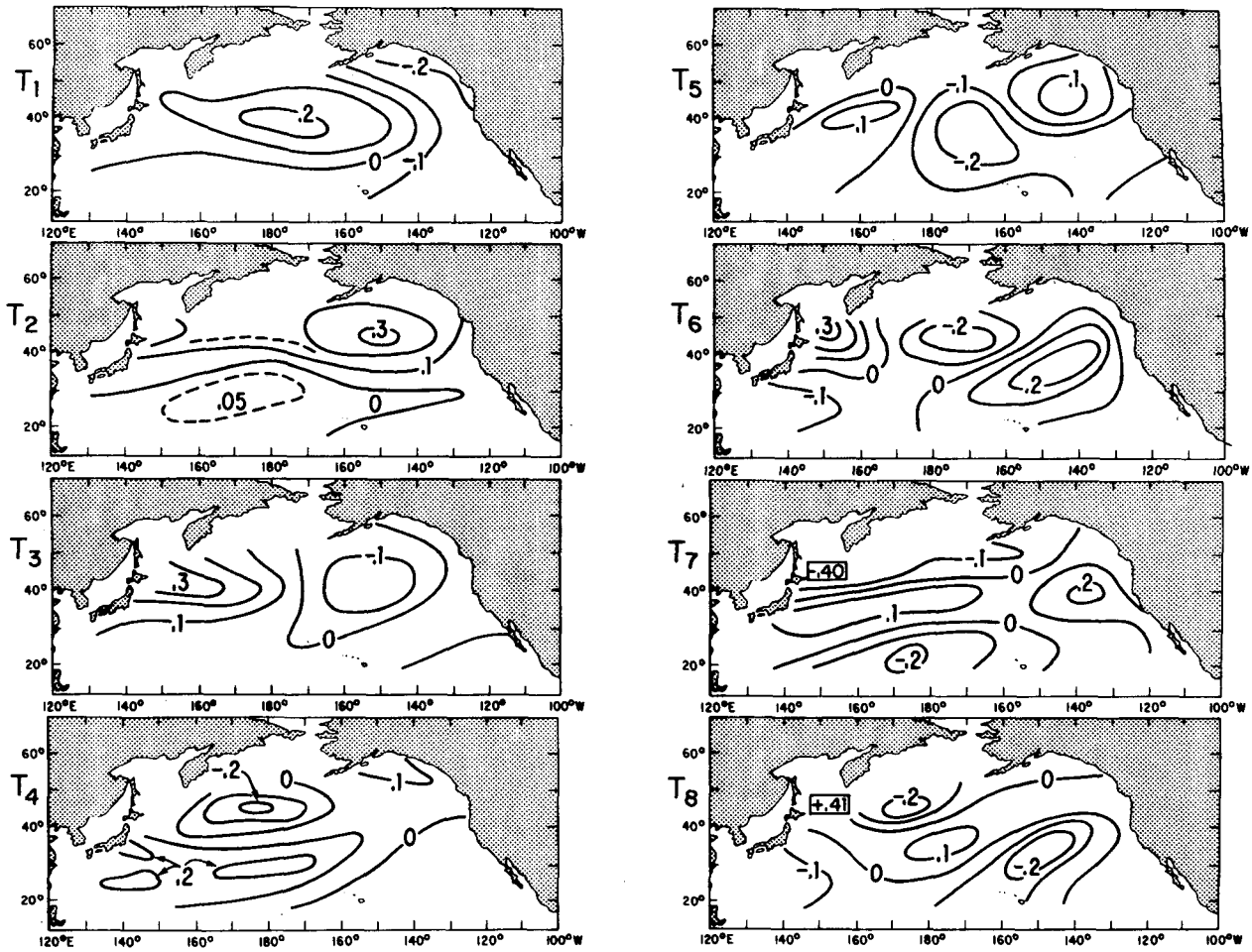

FIG. 4. The eight principal empirical orthogonal functions, T_1–T_8, describing SST anomalies.

matched. SST variability appears to have a continuous spectrum dominated by low frequencies (lower than can be resolved in a 28-year record) with a rich distribution of space scales whose statistics are strongly dependent on position. SLP variability is also inhomogeneous but is dominated by large spatial scales and a nearly uniform distribution of time scales.

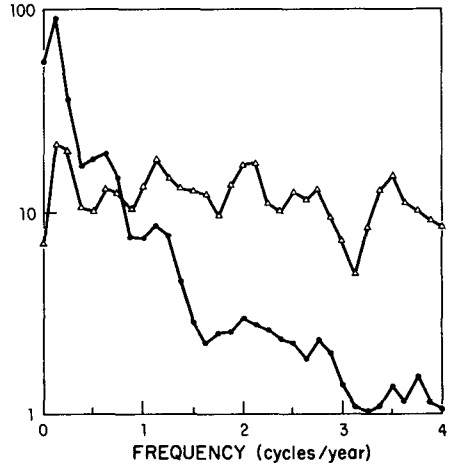

FIG. 5. Frequency spectra of θ_1, the amplitude of the dominant SST mode (circles), and π_1, the amplitude of the dominant SLP mode. Spectral units are arbitrary.

4. Predictability

As discussed in the Introduction, the objective of this study was testing the three hypotheses (A) SST anomalies can be predicted from prior SST data, (B) SLP anomalies can be specified from SST observations, and (C) SLP anomalies can be predicted from prior SST observations. It was pointed out that (C) need not follow from (A) and (B) since the present SST variability could be the result of both predictable oceanic processes (such as advection) and atmospheric forcing [thus explaining the connection between SST and SLP which leads to (B)], but this would not allow prediction of SLP, since it would be a cause, rather than an effect, of SST variability.

Because of the problem of artificial predictability, discussed in Section 2, it is necessary to limit the number of parameters of the SST field which can be

considered as data for prediction or specification, and this limitation must be imposed by an *a priori* criterion not related to data/predictand correlation. It is also necessary to maximize the number of realizations used to construct the sample covariances upon which linear predictors are based. Reduction of the number of data parameters was accomplished by including as data only the amplitudes of the ten dominant empirical functions. This reduced the number of degrees of freedom in the basic SST anomaly maps from 73 to 10. The need to maximize the number of realizations used in examining predictability and correlation prevents treatment of individual seasons. This is so because the time scale of SST anomalies is comparable with a season and consequently use of data from only one season would lead to an expected artificial predictability approximately equal to

$$\frac{M}{N}\frac{\tau}{\Delta t}\approx\frac{10}{28}$$

from the present 28-year record. Since the true predictability or correlation is of this order, any conclusions based on seasonally stratified correlations must be regarded as quite uncertain owing to lack of statistical reliability. A similar effect prevents examination of time scales in excess of a few months since the correlation time scale τ of these components will exceed significantly the shortest time scale admitted. Thus if a 6-month filter were applied to the data the correlation scale of SST could be expected to approach one year and any less than perfect correlation could not be regarded as significant. As a consequence of these considerations it appears that:

(e) The available 28-year record does not allow statistically significant conclusions to be made about more than a few SST parameters, time scales in excess of a season, or seasonally dependent correlations.

Before turning to the results of testing the hypotheses, it is important to point out the restrictions on the kinds of SST/SLP connections which can be examined by the techniques employed here. The analysis is based on correlation of temporally and spatially smoothed data because the data are monthly averages and only a few of the empirical spatial functions are admitted. Some processes have, therefore, been removed from consideration, most noteworthy being synoptic scales in the SLP. Because of this filtering, and because the analysis is based on correlations rather than higher order statistics, the relationship between SST variability and the intensity of synoptic-scale activity, for example, cannot be examined. Most other plausible effects should, however, be disclosed. Persistence of oceanic heat anomalies and their advection by major currents should appear, respectively, in the time-lagged autocorrelations and cross correlations of the amplitudes of SST patterns. SST variations resulting from advection by wind-driven fluctuating currents will be disclosed by cross correlations between SST patterns and SLP. This effect will lead to cross correlations not totally dissimilar to those associated with atmospheric circulation changes resulting from anomalous heat fluxes correlated with SST. Any other similar quasi-linear effect will also appear, whether the mechanism involves correlations between local values of SST and SLP or between local derivatives (in time or space) of these fields. Use of the empirical function representations allows discovery of processes which are spatially inhomogeneous or occur only at selected locations. The difficult problem of inferring mechanisms from the various correlations is discussed in a subsequent section.

The procedure for testing the three fundamental hypotheses is to examine the skill with which the optimal linear statistical predictor hindcasts the 28 years of available data. The predictand considered representative of SST is $\tilde{T}$, the field described by the dominant ten empirical functions T_1 through T_{10}. The SLP predictand is $\tilde{P}$, the portion of SLP anomaly described by the five dominant modes P_1 through P_5. These spatially filtered fields account for 73 and 84% of the respective total fields of variability. In either case the data were the amplitudes of the first ten SST empirical functions from a single month, so that M of Section 2 has the value 10. In order to investigate the possibility that different time scales of variability have different predictability, separate examinations were made of the basic records of one month averages and of records smoothed by a three month running mean filter with equal $\frac{1}{3}$, $\frac{1}{3}$, $\frac{1}{3}$ weights. As discussed in arriving at (e), examination of longer time scales is prevented by consideration of artificial predictability.

The quantitative measure of predictability is based on the mean square error between the estimates, $\hat{P}$ or $\hat{T}$, and the corresponding observations, $\tilde{P}$ or $\tilde{T}$. For each observed month the square errors are summed over the spatial grids and the skill index Z is computed from the mean value of this square error averaged over the 28 years of data. If we let Q be the variable being estimated (either $\tilde{T}$ or $\tilde{P}$) and q_n be the corresponding empirical function amplitude, then the skill index, measured as the fraction of variance predicted, is

$$Z=1-\sum_x\{(Q-\hat{Q})^2\}/\sum_x\{Q^2\}=1-\sum_x\{(q_n-\hat{q}_n)^2\}/\{q^2\}$$

$$=\sum_n\sum_m\{q_n d_m\}^2/[\{d_m^2\}\cdot\{q^2\}]$$

$$=\sum_n\{q_n^2\}S_H(n)/\{q^2\}, \quad (10)$$

where $\{q^2\}=\sum_n\{q_n^2\}$, and $S_H(n)$ is the skill of estimating q_n as defined in (4a) with the predictand p being q_n and, for both SLP and SST estimation, the data d_m being the amplitudes θ_m of the ten dominant SST

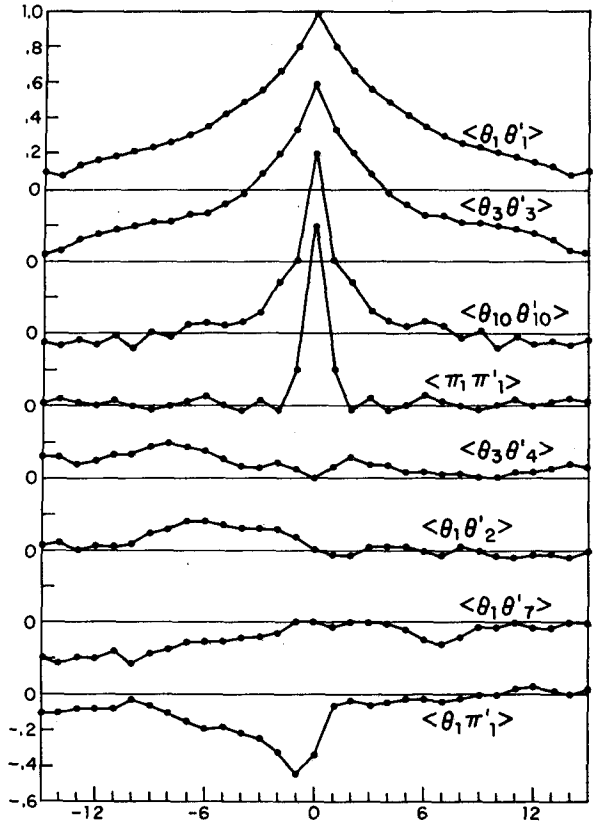

FIG. 6. Correlations between amplitudes of various SST and SLP modes. The amplitudes of SST mode n and SLP mode m are θ_n and π_m, respectively. The abscissa is the time lag t' so that the curve labeled $\langle\theta_3\theta_4'\rangle$ is the correlation of $\theta_3(t)$ and $\theta_4(t+t')$. Thus, significant correlation of π_1 and θ_1 occurs only when π_1 leads.

patterns. The covariance $\{pd_m\}$ appearing in (4a) is then the covariance of $\theta_m(t')$ and $q_n(t'+t)$ where t is zero for a specification and is, when positive, the range of the forecast. In the case of the three month averages the only change in the skill index is that the quantities θ and π appearing in (4a) and (10) are the smoothed values rather than one month averages.

In assessing the utility of a prediction associated with a particular skill index it is important to note that Z is the square of the correlation between the predicted and observed fields, i.e.,

$$\sum_x \{Q\hat{Q}\}/[\sum_x \{Q^2\}\cdot\sum_x \{\hat{Q}^2\}]^{\frac{1}{2}}$$

$$=\sum_n \{q_n\hat{q}_n\}/[\{q^2\}\sum_n \{q_n^2\}]^{\frac{1}{2}}$$

$$=\sum_n\sum_m \{q_nd_m\}^2\{d_m^2\}^{-1}[\{q^2\}\sum_n\sum_m \{q_nd_m\}^2\{d_m^2\}^{-1}]^{\frac{1}{2}}$$

$$=\sum_n [\{q_n^2\}S_H(n)/\{q^2\}]^{\frac{1}{2}}=Z^{\frac{1}{2}}.$$

Thus a skill index Z equal to 0.1 corresponds to a correlation between prediction $\hat{Q}$ and observation Q, summed over the entire grid, of 0.3; this would probably be a useful prediction.

The long time scale of the SST variability, as seen in the spectrum of Fig. 5, suggests that SST is reasonably predictable, even if only as the result of persistence of individual patterns. That such persistence is not the complete story of SST anomaly development can be seen from the cross correlations between amplitudes of different empirical patterns. Although simultaneous amplitudes are uncorrelated, many amplitude pairs show increasing cross correlation with increasing time separation up to a year. This is shown in Fig. 6 where various selected SST and SLP correlations are presented. The occurrence of time-lagged cross correlations between different SST amplitudes shows that thermal anomaly development involves regular pattern transformations such as might result from advection by permanent currents. The patterns of anomaly development are discussed later; the point of interest here is that predictability does not depend on persistence alone.

The skill of hindcasting SST in month $t'+t$ using SST data from month t' is shown in Fig. 7, for both one and three month averaged records as a function of time lag t. The skill of estimating SST in months prior to the data month (negative lags) are shown along with the skill of forecasts. Estimates of artificial predictability, made from the unapproximated form of (5) and the observed sample correlations, indicate that the expected hindcast skills in the case of no true predictability are 0.12 and 0.18 for the one and three month averages, respectively. The skill exceeds significantly the artificial predictability for all lags from -12 months to $+12$

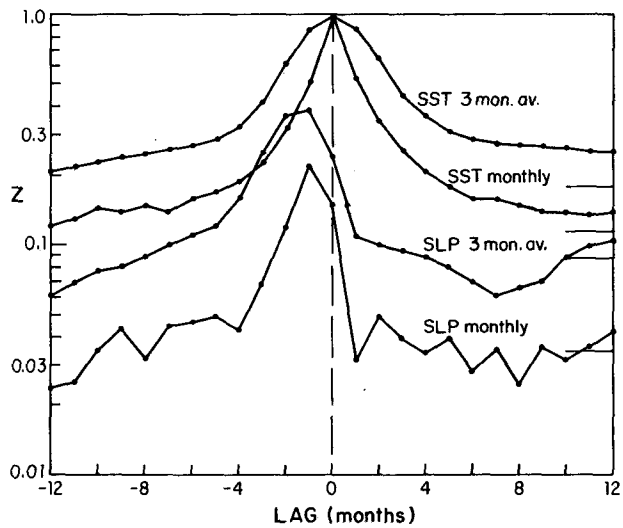

FIG. 7. Skill of estimating SST and SLP in month t+lag using as data ten SST modes from month t. The skill index Z, described in the text, measures the fraction of the entire spatial field correctly estimated. The top two curves refer to estimating one and three month averages of SST; the lower two curves are SLP estimation skill. The horizontal lines at the right are the artificial skill expected when there is no true skill.

months. Although the error rises rapidly with lag, it is likely that the predictions of one month averages are useful for time separations up to three or four months while prediction of three month averages may be useful for time lags of six months or more. Again it must be noted that Z is the square of the correlation between observation and estimate. The degree of predictability of SST from SST data leads to the conclusion that:

(f) Hypothesis A is correct and SST is reasonably predictable from observations of SST. The predictability is primarily the result of persistence of individual patterns but there are significant cross correlations between modes, as there would be if SST anomalies undergo advective displacements or scale conversion.

Before turning to the question of the degree to which SLP anomalies are determined by SST variability it is instructive to examine how well SLP can be predicted from prior SLP data alone. From Fig. 7 it is clear that there is some correlation of π_1 at one month separations. There are other weak correlations at one month separation which lead to a total skill of 0.041 for a one month prediction of SLP from SLP data. This significantly exceeds the expected artificial predictability (which is 0.015) but is certainly too small to be useful; longer range predictions show no evidence of any true skill. It must be remembered that correlation of one month averages separated by one month does not imply that the fields themselves are correlated at one month separations. The correlation could result entirely from averaging and the fact that a short time scale disturbance at the boundary between two adjacent months will influence both monthly averages.

Because the question of predictability at ranges exceeding a few days is of considerable interest it is worth examining the significance of the autocorrelation of π_1 averages at one month separation. The short time correlation (for time separations much less than a month) of pressure patterns is often approximated by the form

$$R(\tau)=e^{-\alpha|\tau|},$$

where α is of the order of 0.25–0.3 day^{-1} (cf. Leith, 1973). This correlation implies that pressure development is a Markov process and would then be predictable only through persistence. It is straightforward (Munk, 1960) to show that the corresponding correlation of adjacent one month averages would be approximately

$$0.5[\alpha(30\text{ days})-1]^{-1}\approx 0.06\text{–}0.08.$$

The autocorrelation of π_1 exceeds this value by a statistically significant amount. One must conclude either that pressure development over the Pacific is not a Markov process (there is some intrinsic predictability) or that the time scale of the dominant pressure mode approaches 8.5 days, a value approximately twice as long as usually accepted for mid-latitude pressure patterns. Without additional data this question cannot be resolved. In any case, it does not seem likely that one month statistical predictions of pressure which use as data only previous pressure observations will lead to a useful level of skill.

The two hypotheses B and C, which are concerned with how well SLP can be specified (estimated at the same time as data is observed) and predicted, can be examined by the same procedure as used for SST. The only changes are that in (10), which defines the skill index Z, Q becomes $\tilde{P}$ and q_n becomes π_n. The resulting skill for estimating $\tilde{P}$ at time $t'+t$ from SST data at time t' are shown in Fig. 7 for positive t lags (prediction), zero lag (specification), and for negative lags (estimating previous SLP from present data). This curve shows quite clearly that SST and SLP anomalies are coupled and suggests that the atmosphere is cause and the ocean is effect; it is clear that the atmospheric response to SST is at least undetectable. It is remarkable, in fact, that the SST data specifies the SLP three months previous with considerably greater accuracy than it predicts SLP in the following month. The skill of a one month prediction of monthly mean SLP anomaly from SST data is slightly less than the expected artificial predictability and is also less than the skill of the equivalent prediction using SLP data; the SLP data are associated with considerably smaller artificial predictability.

SLP anomalies are so well correlated with SST variability that the connection can be seen directly in the time series of the amplitudes of the dominant SLP pattern, P_1, and the dominant SST mode, T_1, shown in Fig. 8. On the basis of the similarity of π_1 and θ_1 one does not hesitate to conclude that a connection exists and the predictability results confirms that:

(g) Hypothesis B is correct and the connection of ocean and atmosphere is strong enough to allow skillful specification of SLP using simultaneous observations of SST.

What is not so clear from Fig. 8 is that SST lags SLP. This was shown clearly for the two amplitudes θ_1 and π_1 by the correlation function plotted in Fig. 6. From the estimation skill Z one concludes that it is true for the entire fields and that:

(h) Hypothesis C is not true. Seasonal time scale anomalies of SLP cannot be predicted from SST, apparently because the primary coupling is atmosphere driving ocean. The process of estimating simultaneous SLP from SST data is evidently specifying cause from observed effect which is possible because the one month averaging interval is comparable with the time scale of the interaction.

Further discussion of the significance of these results is deferred to a later section.

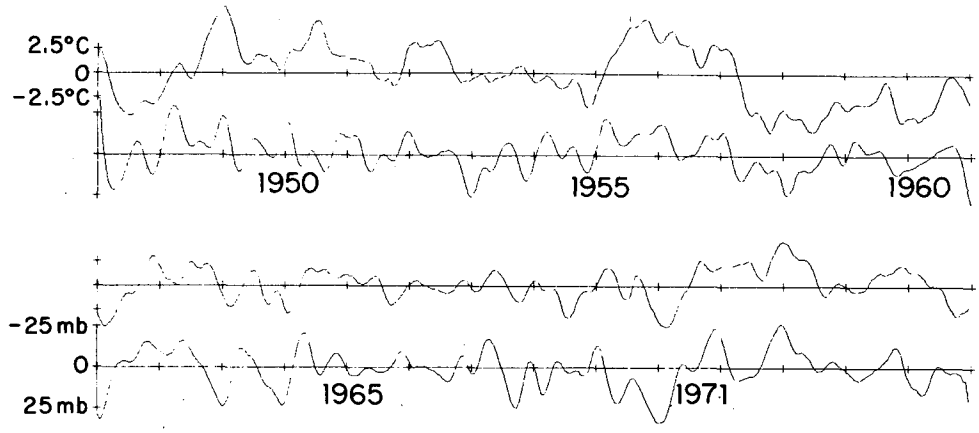

FIG. 8. Time series of θ_1 (upper) and π_1 (lower), the amplitudes of the dominant SST pattern and SLP pattern, respectively. The SST curve is a three-month running mean of monthly values; the π_1 curve has been smoothed twice with a three-month running mean. Note that the ordinate for π_1 is inverted. Tic marks are January and year labels are below January of that year.

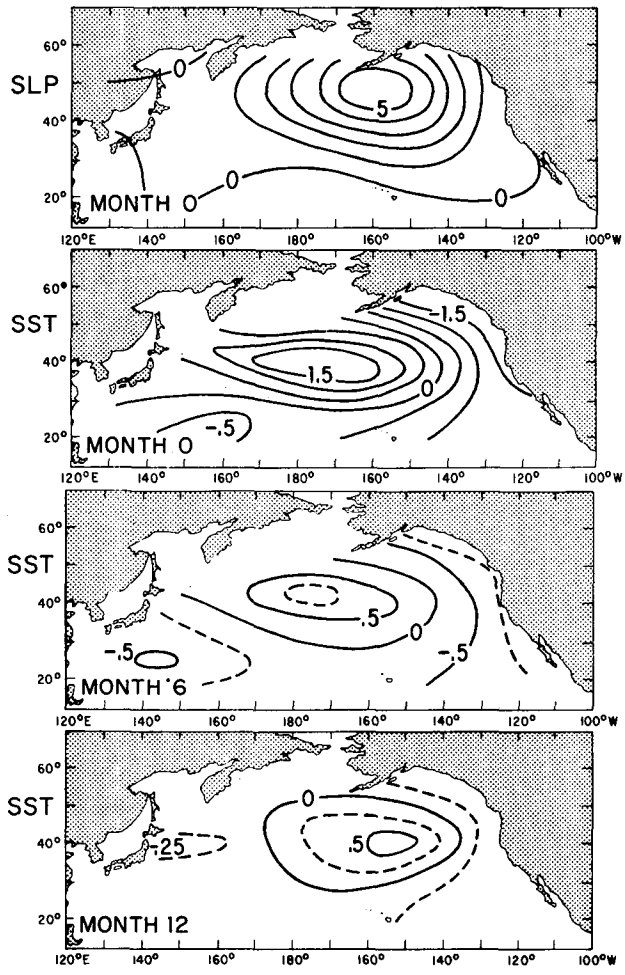

FIG. 9. Simulated three-month average of simultaneous SLP anomaly (upper panel, mb) specified from an initial SST anomaly (second panel, °C) composed of T_1. Lower panels show SST anomaly developed after six months and one year. Contour interval is 1 mb and 0.5°C; dashed curves are medians.

5. Model predictions

In the previous section the hindcast skill of statistical estimators based on SST data was examined. The structure of these estimators, described in Section 2, is that an estimate or prediction is made as a linear combination of input data. As such, they represent a kind of model of the process which, although based on observation rather than "first principles," functions in very much the same way as dynamical models and can be used to carry out model experiments which describe the behavior of the system.

In addition to the conclusions which can be drawn from the prediction skill [see conclusions (e), (f) and (g)] something of the structure of basin-wide, seasonal time-scale, oceanic and atmospheric variability can be learned from the structure of model simulations. A procedure for such examination is to propose an initial SST anomaly as data and examine the model specification of the associated simultaneous SLP anomaly and the model predictions of subsequent SST anomaly development. The procedure is then similar to proposing initial data for a numerical model and observing the simulated development. The only restriction is that the initial data must be made up from the ten SST empirical modes used as data for the model. All simulations described are for three-month average anomalies.

A number of such model simulations have been made and the following qualitative observations hold:

(i) There appears to be a regular relationship between SLP anomalies and the SST data from which they are specified. In most cases the regions of cool SST coincide with regions of anomalous geostrophic wind from the north or northwest, and vice versa for warm anomalies.

(j) The dominant features of SST anomaly predictions are propagation to the east in the region between 30 and 45°N and a general tendency toward conversion

to larger scales. There are, however, initial anomaly patterns whose development is not described by large scale advection by known current systems.

Fig. 9 depicts the model simulations of three month averages associated with an initial SST anomaly described by the dominant mode T_1 with $\theta_1=8°C$. The associated simultaneous SLP anomaly is as large as 4.5 mb and consists almost entirely of the dominant SLP mode P_1. Accepting that the SLP anomaly is cause rather than effect, plausible explanations for the SST/SLP connection are (1) wind-driven oceanic advection, with south winds over the central ocean advecting northward warm water and north winds off the North American coast producing advection southward and possible upwelling; or (2) surface heat flux, with south winds over the central ocean bringing northward warmer and more humid air than normal with a resultant anomalous heat flux to the ocean and much the reverse occurring along the North American coast. Also shown are predicted SST anomalies after six months and one year. The regular SST development is obvious and could be described as eastward advection by the Pacific West Wind Drift; the motion of the small negative temperature anomaly initially at 20°N, 150°E could be ascribed to advection by the western boundary current system of the subtropical gyre.

Fig. 10 further demonstrates the general relation between simultaneous SST and SLP anomaly patterns. The specified SST anomalies are made up of less dominant modes, specifically T_3 and T_6. Despite the fact that these modes are relatively more susceptible to noise, of both statistical and observational origin, clear patterns of associated SLP anomaly emerge and in both cases the general observation (i) holds true. In fact, a search of the SLP anomaly associated with the ten dominant SST modes discloses no contradiction to this statement, although for the higher, less energetic modes the relationship is somewhat clouded by what appears to be noise.

Fig. 11 explores the development of an initial SST anomaly composed of T_4 and T_5 ($\theta_4=6°C$, $\theta_5=2°C$) which are patterns associated with much less energy and much smaller scale than the dominant mode examined in Fig. 9. The development demonstrates the tendency of the higher modes, with their smaller spatial scales, to convert into large-scale patterns, a tendency which was inferred from the asymmetry of the SST prediction skill curves (see Section 4). Also evident in the curve is eastward propagation in the West Wind Drift region. The conversion to larger scales is a general feature of all simulations investigated but eastward propagation, although typical, is not always found.

6. Conclusions

The conclusions of this study fall into three catagories, namely, those concerning analysis methodology, scale description and predictability.

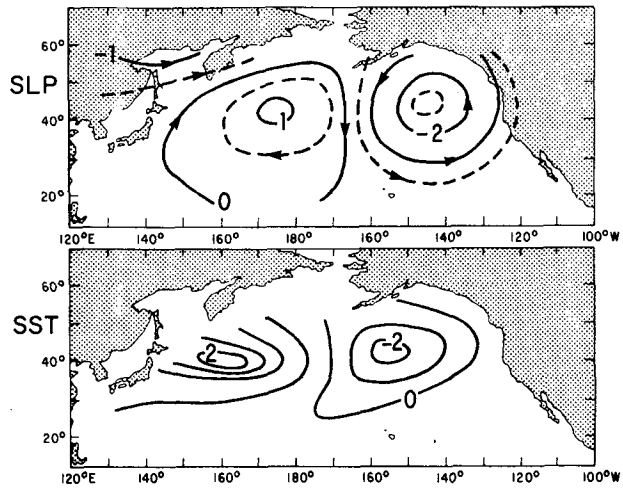

FIG. 10a. Three month average SLP anomaly (upper, mb) specified from three month average SST anomaly (lower, °C) made from T_3.

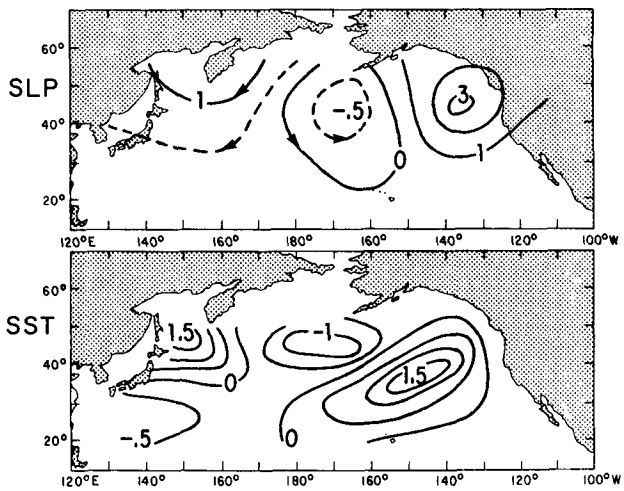

FIG. 10b. As in Fig. 10a except for SST anomaly made from T_6

The discussion of Section 2 underscores the importance of limiting the number of possible data parameters before carrying out a statistical examination of correlation and/or predictability. If the number of realizations in the set of observations is finite then an exhaustive search of all possible data/predictand pairs will almost inevitably show some apparent correlations, even if no true correlation exists. These apparent correlations imply a degree of predictability which is artificial. The results of Section 2 show that the candidate data variables admitted for examination should be limited by an *a priori* criterion and that their number should be reduced until the expected artificial predictability is significantly less than the predictability found; this makes it likely that application of the predictor to data not included in its construction will lead to forecasts not seriously deteriorated by statistical uncertainty in the covariances used in building the predictor.

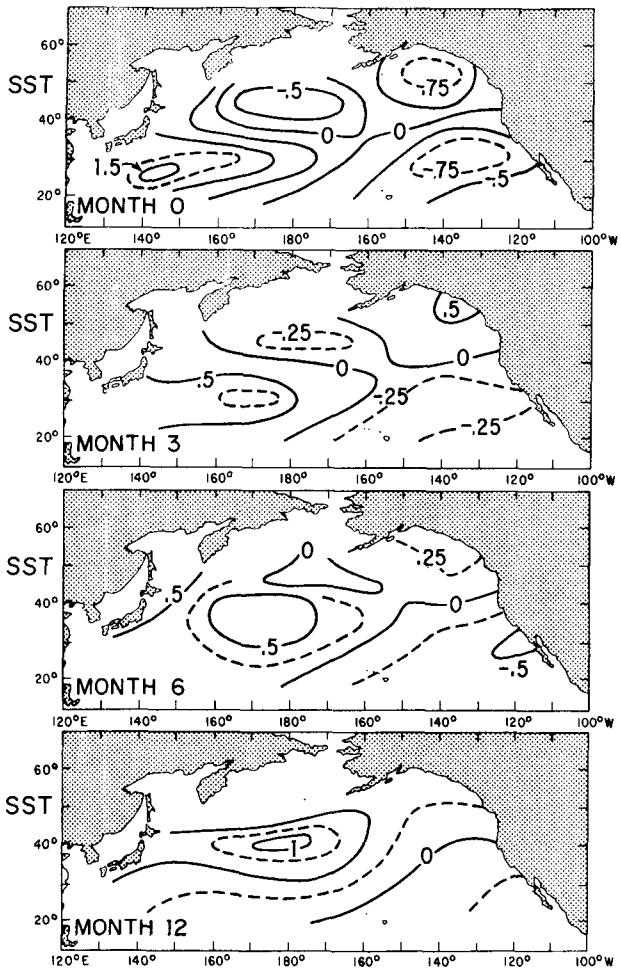

FIG. 11. Simulated development of three-month average SST anomaly initially specified by empirical modes T_4 and T_5. Contour interval between solid curves is 0.5 and dashed curves are medians. The forecast range is up to one year.

The expected artificial predictability, and hence the maximum number of admissable data variables, is determined by the ratio of the length of the series of observations and the integral time scales τ_n (defined just prior to Eq. (5)] which measure the time required to obtain a new "degree of freedom" in estimating the data/predictand correlations. Representation of the data in terms of empirical orthogonal functions provides a seemingly plausible way of eliminating data variables, namely deleting those modes which contribute little to the variance of the data field.

The time scale of SST and SLP anomalies are characterized by the frequency spectra of the amplitudes of the associated empirical modes. The spatial scales are characterized by the fraction of variance associated with different empirical functions in much the way that Fourier spectra associate variance with trigonomeric functions with different wavenumbers. In this investigation it was found that[2] (a) the region of maximum SST variability is in the western part of the Pacific West Wind Drift while (b) SLP variability is greatest just south of the Aleutian Islands. SLP variability is concentrated in a few spatial patterns with large spatial scales while SST variability has a rich spatial distribution encompassing many modes (c). SLP variability is described by a nearly uniform distribution of variance by frequency whereas SST anomalies appear to be characterized by a "red" frequency spectrum without any apparent peak at periods less than 8 years (d).

While the above conclusions are fairly well defined, any estimate of predictability found by a statistical examination must be interpreted with care. On the one hand, it is always possible that the prediction skill evidenced is artificial predictability resulting from fortuitous coherence of variables which are actually uncorrelated, it must be remembered that only the *expected* artificial predictability can be estimated. On the other hand, once artificial predictability is accounted for, the skill of a statistical predictor must be considered as a lower bound for the predictability of the system. This is so because the statistical predictor must, in order to obtain statistical reliability, be based on a limited numer of data parameters, or potential "causes." It is always possible that the most relevant data variables have been excluded, thus giving a low estimate of predictability. It is inevitable, then, that a statistical predictability study is subject to criticism as to choice of data. Unfortunately, rectification of this limitation cannot easily be achieved without some *a priori* criterion for choosing the data. It is here that models, theoretical or numerical, play a valuable role in isolating potential causes and effects.

There are three specific ways in which the predictability results of the present study may be misleading. First, the choice of SLP as the description of atmospheric circulation may be unfortunate in that other atmospheric variables could show a greater influence of SST. It is to be recalled that Namias, for example, usually employs 700 mb height data as the indicator of circulation. Equally it is plausible that variables like cloud cover, precipitation, stability or storminess show a greater response to SST than does pressure. A second possible criticism is that the choice of the ten dominant empirical SST modes is inappropriate and that elimination of the less energetic modes leads to loss of predictability. A third criticism is failure to account for seasonal differences in the processes which lead to predictability. To do so requires construction of different estimation models, based on seasonally stratified statistics, for the different seasons. The difficulty is that seasonally stratified statistics are based on less observations than the year-round statistics employed in this study. The first predictability results obtained during the course of this study were, in fact, constructed from seasonally stratified statistics and it

[2] Lower case letters refer to conclusions found in the text.

was from these results that the importance of artificial predictability was noticed. It is clear from such studies that reliable SST/SST predictors using more than about eight data variables cannot be constructed from seasonally stratified statistics (e). Owing to the short time scale of SLP anomalies it is, however, possible to construct a seasonally dependent estimator of SLP using a reasonable number of SST modes as data, and this has been done.

The procedure of testing for any increase in predictability of SLP by accounting for seasonally dependent processes was the following: Separate estimators were constructed for estimating SLP anomalies using SST data in cold months (November through April) and using SST in warm months. The resulting skill indices for estimating monthly SLP anomalies from monthly SST data are shown in Fig. 12. The results, aside from a general increase in skill resulting largely from increased artificial predictability, are similar to those obtained using year-round statistics.

Bearing in mind the limitations mentioned above, the results of the present predictability study may be easily stated. The persistence and regular pattern transformation of SST anomalies does allow their prediction using, as data, observations of SST (f). This is particularly true of three month average anomalies which can apparently be predicted six months in advance. The general features of such predictions are propagation to the east in the region of the Pacific West Wind Drift and a gradual transformation to larger scales (j). SST and SLP anomalies are sufficiently well correlated that SLP can be specified, with demonstrable skill, using as data SST observations from the same month (g). It is similarly possible to estimate prior SLP anomalies from SST data, the greatest skill being obtained in specifying the SLP anomaly one month prior to the time of observing the SST data. The relationship between SLP anomalies and the associated SST anomaly is a regular one described by geostrophic winds from the southeast overlying warm SST anomalies and *vice versa* (i). Despite a good simultaneous connection between SST and SLP it is not possible to predict future SLP anomalies from SST data using the method examined here (h). In fact, greater true skill in predicting SLP anomalies is achieved using SLP data than using SST data.

The connection of simultaneous SST and SLP anomalies is in substantial agreement with those found by Roden and Ried (1961) and by Namias (1972b). The lag of SST anomalies behind the associated SLP anomalies was also found by Roden and Groves (1960).

It is the belief of the author that the predictability results lead to the conclusion that the observed association of SST and SLP anomalies having approximately seasonal time scales is the result of atmospheric forcing of the ocean. This does not preclude possible back forcing, ocean to atmosphere, but does suggest that (1)

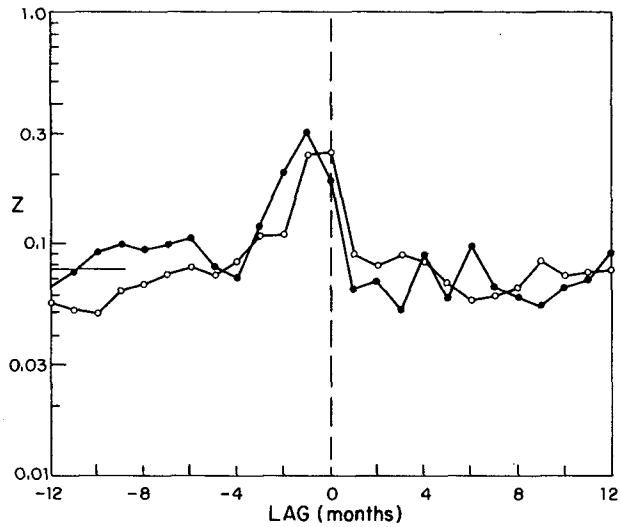

Fig. 12. Skill of estimating monthly SLP anomalies in month t+lag using as data ten SST modes from month t by separate estimators for cases when the SST data is from the cold months November through April (solid circles) or from warm months (open circles). The skill index Z is described in the text and the skill obtained for a single year-round estimator is shown in Fig. 7.

the effect is very weak, and (2) that ability to predict SST from SST plus ability to specify simultaneous SLP from SST observations does not mean that SLP can be predicted. SLP could not be predicted if, as is suggested here, the ocean has thermal memory and is subject to atmospheric forcing. The present state of SST would then be the result of recent atmospheric causes, allowing specification of simultaneous monthly average SLP anomalies, and the result of prior forcing whose effects persist, allowing predictability of SST from prior SST.

An additional argument in favor of the idea that observed seasonal SST anomalies are the result, not the cause, of SLP anomalies, comes from the general pattern relating the two associated fields. The observed relation is one of northward flowing geostrophic winds overlying warm SST anomalies. In their numerical modeling study of atmospheric response to SST forcing, Salmon and Hendershott (1975) find the dominant atmospheric response to be a high pressure region south of a cold SST anomaly. The observed pattern is that the SST anomalies are generally south of the SLP anomaly and centered longitudinally about the position of maximum north/south wind rather than the maximum pressure. Salmon and Hendershott do, however, find a realistic SLP/SST connection when they allow the ocean to respond to the atmosphere through heat fluxes. This result together with results of the study by White and Clark (1975) show that SST anomalies may be the result of anomalous heat fluxes. This could explain the observation that southeast geostrophic winds, which are likely to bring north warm moist air and result in an anomalous heat flow to the ocean, are associated with

warm SST anomalies. It is, however, equally plausible that northward oceanic currents, driven by southeast geostrophic winds, cause warm SST anomalies through advection in the northward gradient of oceanic temperature.

Because of the vital importance of possible prediction of environmental factors, both atmospheric and oceanic, the results of this study cannot be taken as final. Clearly, the suggested oceanic response to atmosphere can be exploited to provide an improved ability to predict upper ocean thermal structure and work toward this end is under way. Further, it is essential to continue exploration of possible ways of predicting short-term climate fluctuations in the atmosphere and SST cannot be ruled out as a potential predictor. The present study must be expanded upon before it is to be accepted that the ocean is not a significant factor in determining atmospheric variability on seasonal time scales. It is, however, clear that future claims of predictability from oceanic variables must rest on more than demonstration of a connection between ocean and atmosphere.

Acknowledgments. I wish to especially thank Jerome Namias for his help and interest in this work. Despite the controversial nature of the results, he has continued in every way possible to encourage and improve this study. This work was supported by the Office of Naval Research, under Contract N00014-75-C-0260 as part of the NORPAX project.

APPENDIX A

The Namias Data Sets

The analysis described here was performed on "finished" data sets made available through the generosity of Jerome Namias and Robert Born of Scripps. It must be emphasized that their effort associated with obtaining the basic data and generating the finished data sets greatly exceeds that expended in the analysis reported here.

The SLP data were originally obtained from the Long Range Prediction Group of the National Meteorological Center as one month averages on a 5° diamond grid (i.e., 20°N–140°W, 20°N–150°W, . . . , 25°N–145°W, 25°N–155°W, etc.). The data were transferred onto the regular 5° square grid used by Namias' group using linear interpolation from the four nearest diamond grid points to fill in the square grid.

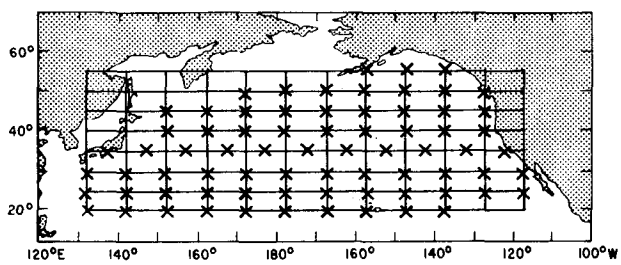

FIG. 13. The grids of SST and SLP data. Ten degree by five degree SLP averages are centered at grid intersections and SST averages are centered at crosses.

The SST data were obtained primarily from the National Marine Fisheries Service in the form of averages over one month and 2° squares. Because this grid spacing is not a submultiple of 5° and because data were sometimes "bad" or missing the following analysis scheme was employed. The 2° data were subjectively analyzed to produce "maps" contourable with a 1°F contour interval. During this stage "bad" points were altered and missing data filled in where feasible. These "corrected" data were linearly interpolated onto a 1° grid and 25 values averaged to provide area averages on the standard 5° grid.

It is recognized that the subjective aspects of this procedure may be regarded as undersirable and that the quality of the basic ship intake temperature reports is not high (see Saur, 1963). Robert Bernstein and Richard Wert of Scripps are presently gathering the raw ship reports and objectively analyzing them as part of the NORPAX program. This should eventually lead to a more precise determination of data quality but it is believed that for present purposes the Namias data are adequate and the analysis method, based on empirical eigenfunctions, is reasonably insensitive to the geographically local errors associated with raw data quality and subjective analysis.

Before analysis began, the Namias data sets were further averaged onto a grid with 5° latitude spacing and 10° longitude spacing. In cases when one SST datum was missing the available datum was used as the average; there remained occasional 5°×10° regions of missing data in the SST fields. The 5°×10° grids of SST and SLP data are shown in Fig. 13.

APPENDIX B

Empirical Orthogonal Functions

The purpose of this appendix is to review those properties of empirical eigenfunctions which make them useful in analyzing large data sets. The discussion proceeds in the context of a time series of maps of some scalar variable ϕ. Each map consists of the value $\phi(x,t)$ at N positions x, and there are T such maps, each for a distinct time t. In the discussion the mean $\langle\ \rangle$, analogous to the statistical mean, is defined as

$$\frac{1}{T}\sum_{t=1}^{T},$$

and the inner product, denoted by a dot, is defined by

$$\mathbf{f}\cdot\mathbf{g}=\sum_{x=1}^{N} f(x)g(x).$$

The discussion centers around representing the data

in terms of orthogonal functions according to

$$\phi(x,t)=\sum_n a_n(t)f_n(x),$$

where

$$\mathbf{f}_n\cdot\mathbf{f}_m=\delta_{nm}.$$

Specifically, the orthogonal function representation involves using for the functions $\{\mathbf{f}_n\}$ the "empirical eigenfunctions" $\{\mathbf{v}_n\}$ which are the eigenvectors of the matrix

$$C(x,y)=\langle\phi(x,t)\phi(y,t)\rangle$$

according to

$$\mathbf{C}\cdot\mathbf{v}_n=\lambda_n\mathbf{v}_n,\quad \mathbf{v}_n\cdot\mathbf{v}_n=1.$$

The orthogonality of these functions is easily shown. By convention the $\{\mathbf{v}_n\}$ are ordered so that $\lambda_n\geqslant\lambda_{n+1}$. The fact that $\mathbf{C}$ is the sum of inner products insures that it is a non-negative matrix so that all $\lambda_n\geqslant 0$.

The principal virtues of the empirical eigenfunctions are (1) they provide the most efficient method of compressing data, (2) they may be regarded as uncorrelated modes of variability of the field, and (3) they simplify understanding the procedures of minimum mean square error estimation. These points are considered in sequence below.

The empirical functions are the most efficient representation of the data in the sense that, for a fixed number of functions $M<N$, no approximate representation

$$\hat{\phi}(x,t)=\sum_{n=1}^{M} a_n(t)f_n(x)$$

can produce a lower mean square error

$$E=\langle[\phi-\hat{\phi}]\cdot[\phi-\hat{\phi}]\rangle$$

than is obtained using $\mathbf{f}_n=\mathbf{v}_n$. Elementary least-squares procedures show that, for a given set $\{\mathbf{f}_n\}$, E is minimized by taking

$$a_n(t)=\phi(t)\cdot\mathbf{f}_n,$$

which is independent of M, and that the minimum is

$$E=\langle\phi\cdot\phi\rangle-\sum_{n=1}^{M}\mathbf{f}_n\cdot\mathbf{C}\cdot\mathbf{f}_n.$$

The point of interest here is that the error E can take on no smaller value than that obtained when $\mathbf{f}_n=\mathbf{v}_n$. To see this, note that because the $\{\mathbf{v}_n\}$ form a complete set any $\mathbf{f}_n$ can be represented as

$$\mathbf{f}_n=\sum_m^N\gamma_{mn}\mathbf{v}_m,$$

where, according to the orthogonality of the $\{\mathbf{f}_n\}$ and the $\{\mathbf{v}_n\}$,

$$\sum_n^N\gamma_{nm}\gamma_{nl}=\sum_n^N\gamma_{mn}\gamma_{ln}=\delta_{ml}.$$

Also note that any orthogonal set of functions generated by linear combination of the M $\mathbf{f}_n$ functions

$$\mathbf{f}_n'=\sum_m^M\gamma_{nm}'\mathbf{f}_m$$

leads to the same mean square error as the original $\mathbf{f}_n$ since

$$\sum_m^M\gamma_{nm}'\gamma_{nl}'=\delta_{ml},$$

and therefore

$$\sum_{n=1}^{M}\mathbf{f}_n'\cdot\mathbf{C}\cdot\mathbf{f}_n'=\sum_n^M\sum_m^M\sum_l^M\gamma_{nm}'\mathbf{f}_m\cdot\mathbf{C}\cdot\mathbf{f}_l\gamma_{nl}=\sum_n^M\mathbf{f}_n\cdot\mathbf{C}\cdot\mathbf{f}_n.$$

It follows then that every set $\{\mathbf{f}_n'\}$ of M functions is associated with a special set $\{\mathbf{f}_n\}$ with $\gamma_{nm}=\delta_{nm}\Gamma_n$ for $n,m\leqslant M$ and both have the same minimum mean square error

$$E=\langle\phi\cdot\phi\rangle-\sum_{n=1}^{M}\sum_{m=1}^{N}\sum_{l=1}^{N}\gamma_{nl}\gamma_{nm}\mathbf{v}_l\cdot\mathbf{C}\cdot\mathbf{v}_m$$

$$=\langle\phi\cdot\phi\rangle-\sum_{n=1}^{M}\left[\Gamma_n^2\lambda_n+\sum_{m=M+1}^{N}\gamma_{nm}^2\lambda_m\right].$$

Noting the descending magnitudes of the $\{\lambda\}$ and that

$$\Gamma_n^2=1-\sum_{m=M+1}^{N}\gamma_{nm}^2,$$

it is seen that

$$E\geqslant\langle\phi\cdot\phi\rangle-\sum_{n=1}^{M}\left[\lambda_n+(\lambda_{M+1}-\lambda_n)\sum_{m=M+1}^{N}\gamma_{nm}^2\right]$$

$$\geqslant\langle\phi\cdot\phi\rangle-\sum_{n=1}^{M}\lambda_n.$$

Clearly, this overall minimum mean square error is achieved by taking $\mathbf{f}_n=\mathbf{v}_n$, in which case $\Gamma_n=1$ and $\gamma_{nm}=0$ for $m>M$.

The second point of interest concerning the empirical eigenfunction representation

$$\phi(x,t)=\sum_n a_n(t)v_n(x)$$

is the conceptual virtue that the functions $\mathbf{v}_n$ may be regarded as uncorrelated modes of variability. This follows immediately from

$$\langle a_na_m\rangle=\langle\mathbf{f}_n\cdot\phi(t)\mathbf{f}_m\cdot\phi(t)\rangle=\mathbf{f}_n\cdot\mathbf{C}\cdot\mathbf{f}_m=\lambda_n\delta_{nm}.$$

Thus it is seen that $\langle a_na_m\rangle=0$ if $n\neq m$ and that the eigenvalues λ_n are the portion of the overall variance $\sum_x\langle\phi^2(x)\rangle$ associated with the corresponding eigenvector.

It should be pointed out in this respect that empirical eigenfunction analysis is very similar to the widely used Fourier analysis of finite length stationary time series. In fact, if the process is stationary in the variable

x, so that $C(x,y)=C(x-y)$, then the empirical functions are trigonometric functions and the eigenvalues are the conventional Fourier spectral estimates multiplied by the bandwidth.

One final aspect of empirical eigenfunctions is that they can simplify understanding the theory of minimum mean square error estimation. To demonstrate this feature the specific problem of estimating a signal from noisy data is considered. The example is, itself, interesting but the purpose here is to illustrate a much more general procedure. The problem is to make an estimate $\hat{\phi}$ of some signal $\phi(x)$ from noisy data $d(x)=\phi(x)+\epsilon(x)$ when the noise ϵ is uncorrelated between different x and has uniform variance $\langle\epsilon^2\rangle$. According to the Gauss-Markov theorem the minimum expected mean square error is achieved by the estimator

$$\hat{\phi}(x)=\sum_y \sum_z \langle d(y)d(z)\rangle^{-1}\langle d(z)\phi(x)\rangle d(y),$$

where $\langle dd\rangle^{-1}$ is the inverse of the data-data covariance matrix and $\langle d\phi\rangle$ is the covariance of data and signal. The sum over x of the minimum mean square error $\langle(\phi-\hat{\phi})^2\rangle$ is

$$\sum_x \langle\phi^2(x)\rangle-\sum_x \sum_y \sum_z \langle\phi(x)d(y)\rangle\langle d(y)d(z)\rangle^{-1}\langle d(z)\phi(x)\rangle.$$

In the case of interest here $\langle d(x)\phi(y)\rangle=\langle\phi(x)\phi(y)\rangle$ since the noise is uncorrelated with the signal.

The virtue of empirical eigenfunctions in examining minimum mean square error estimators is that the covariances are conveniently expanded in terms of these functions and this simplifies examination of the performance of the estimator. Specifically,

$$\langle\phi(x)\phi(y)\rangle=\sum_n \lambda_n f_n(x)f_n(y),$$

$$\langle\phi(x)\phi(y)\rangle^{-1}=\sum_n \lambda_n^{-1}f_n(x)f_n(y),$$

and since

$$\langle d(x)d(y)\rangle=\langle\phi(x)\phi(y)\rangle+\langle\epsilon^2\rangle\delta_{xy},$$

$$\langle d(x)d(y)\rangle^{-1}=\sum_n [\langle\epsilon^2\rangle+\lambda_n]^{-1}f_n(x)f_n(y).$$

Substituting these and taking account of the orthgonality of the $\{f_n\}$ gives a much simplified result for the error introduced by the noise, namely

$$\sum_x \langle[\phi(x)-\hat{\phi}(x)]^2\rangle=\sum_n \lambda_n-\frac{\lambda_n^2}{\lambda_n+\langle\epsilon^2\rangle}.$$

In a similar manner it can be shown that the mean square error associated with estimating a_n, the amplitude of the eigenvector $\mathbf{f}_n$ in the signal, is

$$\langle[a_n-\hat{a}_n]^2\rangle=\lambda_n-\frac{\lambda_n^2}{\lambda_n+\langle\epsilon^2\rangle}.$$

This provides one *a priori* criterion for determining how many eigenfunctions should be used to represent a data set; when the variance $\langle a_n^2\rangle=\lambda_n$ approaches the noise variance the estimates of the amplitude become unreliable and the function may be deleted without seriously deteriorating the representation of the signal.

The examples considered above are far from exhaustive and, following similar procedures, it is possible to examine many others including signal extraction from data with spatially correlated noise, estimation of interpolated values to fill in missing data, and statistical extrapolation or prediction. In particular, the discussion in Section 2 is greatly simplified by the use of empirical functions to represent the prediction data.

REFERENCES

Favorite, F., and D. R. McLain, 1973: Coherence in transPacific movements of positive and negative anomalies of sea surface temperature, 1953–60. *Nature*, **244**, 139–143.

Leith, C..E., 1973: The standard error of time-average estimates of climatic means. *J. Appl. Meteor.*, **12**, 1066–1069.

Liebelt, P. B., 1967: *An Introduction to Optimal Estimation*. Addison-Wesley, 273 pp.

Lorenz, E. N., 1959: Empircal orthogonal functions and statistical weather prediction. Report No. 1, Statistical Forecasting Project, Dept. Meteorology, MIT.

Munk, W. H., 1960: Smoothing and persistence. *J. Meteor.*, **17**, 92–94.

Namias, J., 1972a: Large-scale and long-term fluctuations in some atmospheric and oceanic variables. Nobel Symposium 20, D. D. Dyrsen and D. D. Jagner, Eds., Almqvist and Wiksell, Stockholm, 27–48.

——, 1972b: Space scales of sea-surface temperature anomalies and their causes. *Fish. Bull.*, **70**, 611–617.

O'Connor, J. F., 1961: Mean circulation patterns based on 12 years of Northern Hemisphere data. *Mon. Wea. Rev.*, **89**, 211–227.

Robinson, M. K., 1975: *Atlas of North Pacific Ocean Monthly Mean Temperature and Mean Salinites of the Surface Layer*. Naval Oceanographic Office, pp. 262.

Roden, G. I., and G. W. Groves, 1960: On statistical prediction of ocean temperatures. *J. Geophys. Res.*, **64**, 249–263.

—— and J. L. Reid, 1961: Sea surface temperature, radiation and wind anomalies in the North Pacific Ocean. *Rec. Oceanogr. Works Japan.*, **6**, 36–52.

Salmon, R., and M. C. Hendershott, 1975: Large-scale air-sea interactions with a simple general circulation model. *Tellus* (in press).

Saur, J. F. T., 1963: A study of the quality of sea water temperatures reported in logs of ships' weather observations. *J. Appl. Meteor.*, **2**, 417–425.

White, W. B., and N. E. Clark, 1975: On the development of blocking ridge activity over the central North Pacific. *J. Atmos. Sci.*, **32**, 489–502.