THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

THE GENERATION OF SYNTHETIC SEA SURFACE TEMPERATURE
DATA FOR THE EQUATORIAL PACIFIC OCEAN

By

JULIE CARON

A Thesis submitted to the
Department of METEOROLOGY
in partial fulfillment of the
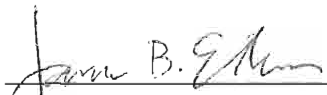requirements for the degree of
Master of Science

Degree Awarded:
Summer Semester, 1997

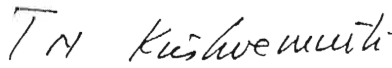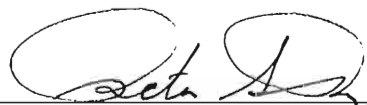The members of the Committee approve the thesis of Julie Caron defended on April 25, 1997.

James J. O'Brien
Professor Directing Thesis

Melody A. Owens
Committee Member

James B. Elsner
Committee Member

T. N. Krishnamurti
Committee Member

Approved:

Peter Ray, Chairperson, Department of Meteorology

This thesis is dedicated to my mother, Cleone Caron...

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Synthetic monthly sea surface temperature (SST) anomaly data are constructed using frequency domain analyses of significant principle components derived from the Reynolds reconstructed SST data in the equatorial Pacific Ocean. The model provides insight into the dominant physical processes contained in each component, and retains the relevant statistical properties of the original data, such as the mean, variance, and autocorrelation. Thus, numerous sets of synthetic SST anomaly data can be produced for the equatorial Pacific, which are statistically indistinguishable from the original SST anomaly data.

The spatial and temporal SST signatures of the biennial, intradecadal, and decadal pseudoperiodicities are reproduced, including their frequency and duration of occurrence. Specifically, the El Niño - Southern Oscillation (ENSO) warm and cold event signatures recur in the synthetic data at peak return periods of 2.4, 3.5, 5.0, and 6.4 years. Moreover, the anticipated return period of an extreme ENSO event with a maximum SST anomaly magnitude of $1.7^{\circ}C$ is approximately every 5 warm events, and every 7 cold events.

1. Introduction

Theoretical methods from which synthetic time series can be constructed provide a means to overcome the limitations of our inevitably short climate records. The algorithm devised in this study is a synthetic sea surface temperature anomaly generator, which successfully reproduces the equatorial Pacific SST anomaly patterns associated with the El Niño-Southern Oscillation and other identifiable SST anomaly signatures.

ENSO is defined quantitatively by its SST signature, using indices such as the JMA (discussed in section 4), but is more broadly defined as a coupled ocean-atmosphere phenomena in which anomalous warming (cooling) of sea surface temperatures in the eastern equatorial Pacific are preceded by anomalous westerly trade wind patterns (strengthened easterly) in the central equatorial Pacific, and the associated excitation of oceanic internal wave dynamics.

Extreme variability in the equatorial Pacific has important implications for the tropics as well as higher latitudes. Quantities such as air temperature, SST, and specific humidity in the western equatorial Pacific (WEP) are comparably large with respect to other latitudes due to intense, year-round solar radiation. The persistent trade winds across the equatorial Pacific basin, in conjunction with the warm SSTs ($\geq 25^{\circ}C$), result in very high surface latent heat fluxes and evaporation in the WEP. Anomalous SST signatures would thus imply a displacement of large heat and moisture transfers, corresponding convective perturbations, and associated atmospheric wave formation (Salby 1996). Such perturbations in the tropics have been shown to be associated with anoma-perturbations, and associated atmospheric wave formation (Salby 1996). Such perturbations in the tropics have been shown to be associated with anomalous patterns in pressure, temperature, and moisture in extratropical latitudes

1

(Yarnal 1985).

We develop a method which successfully reproduces the SST signatures associated with warm and cold events to study ENSO extremes. Algorithms for generating synthetic data are developed using frequency-domain analyses to extract information regarding the contribution of different frequency oscillations to the associated variability of the time series. Thompson and O'Brien (1973) developed a technique to produce realizations of wind stress data, given a known sample spectrum of wind speed. An analytic approximation of the kinetic energy spectrum was derived from the Fourier coefficients of wind speed. Random, antisymmetric phases, $(\theta(f) = \theta(-f))$, are applied to the analytic spectrum. The wind stress realization is obtained via the inverse Fourier transform. Recently, Theiler et al. (1992) used a frequency-domain method, although no analysis of the spectrum is done, to generate synthetic data from the original amplitude spectrum by simply randomizing phases under the aforementioned constraint. Theiler's method, called the method of surrogate data, was applied by Elsner and Tsonis (1993) to investigate the existence of nonlinearity in monthly sea level differences between Tahiti and Darwin. The data sets generated using these methods remain statistically indistinguishable from the observed realizations, thereby producing realistic time series from which valid statistical inferences can be made.

Thompson and O'Brien utilize a functional fit to the amplitude spectrum, thereby assuming that such an analytical approximation accurately captures the variance contribution with frequency needed to produce consistent realizathereby assuming that such an analytical approximation accurately captures the variance contribution with frequency needed to produce consistent realizations. The method developed in this study will take the work of Thompson

2

and O'Brien a step further. We assert that the analytic approximation to the amplitude function should be a deterministic variance distribution within the frequency bands attributable to physical processes, such as ENSO. The remaining portions of the amplitude function are modeled with the theoretical spectra for red and white noise processes, scaled to applicable amplitudes determined directly from the original spectrum. Thus, we allow the rest of the spectrum to represent random interactions between actual physical processes, correlated noise, or random noise. The method used to determine an appropriate functional fit, therefore, requires further spectral analysis than existing methods, and is described in section 3.

By validating ENSO indices of a sample of synthetic data against the Reynolds data, it is found that the frequencies of ENSO occurrences are retained, as are the approximate number of ENSO warm and cold events in a 40 year period. Furthermore, the synthetic data has approximately the same average ENSO event duration as the Reynolds data.

In the following pages, section 2 describes the Reynolds SST data set and the Empirical Orthogonal Function (EOF) technique that makes a frequency domain analysis of the Reynolds data possible. A description of the methods used to determine an appropriate theoretical model for the amplitude spectra is discussed in section 3. In section 4, ENSO indices are used to validate a sample set of synthetic data against the Reynolds data. Further, ENSO event extremes are quantified in terms of expected return periods for warm and cold events of a given SST magnitude. All findings are then summarized in section 5.

2. Data

The data used for this project are the Reynolds EOF reconstructed sea surface temperatures. The complete data set is global, has a 2° spatial resolution, and extends in time from 1950 to 1992. We utilize only the equatorial Pacific region (30S to 30N and 120E to 90W) over the years 1953 to 1992. The Reynolds monthly SST climatology on a 2° grid are used for the computation of SST anomalies in the aforementioned region (Smith et al. 1996).

Reynolds' reconstruction technique uses weekly SST fields from January 1982 to December 1993 (Reynolds and Marsico 1993). These data have been analyzed on a 1° grid using the method of Optimum Interpolation (OI). The OI technique incorporates both in situ and satellite data. The satellite data improves the overall spatial coverage of the SST data, while the in-situ data allows for the correction of satellite SST retrieval bias due to atmospheric aerosol content (Reynolds and Smith 1994). Reynolds' reconstruction methodology is described in Smith et al. (1996), and is summarized in the following sentences. The in-situ data which are found to be without systematic bias and represent sufficient spatial coverage, as determined by Smith et al. (1996), extend from 1950 to the present. However, inadequacies such as inconsistent spatial and temporal in-situ data coverage have resulted in gridded analyses which contain bullseyes in data rich regions and gaps in other regions. The improved spatial coverage provided by the OI data is utilized to fill spatial gaps and produce a smooth SST field for observations prior to 1982. The spatial components of an EOF analysis on the OI data are used as basis functions. These functions are fit smooth SST field for observations prior to 1982. The spatial components of an EOF analysis on the OI data are used as basis functions. These functions are fit to existing data, such as the comprehensive ocean atmosphere data set (Slutz

et al. 1985) from 1950 to 1982. The result is a reconstructed SST data set for the period 1950 to 1992 on the a 2° grid.

2.1 EOF analysis method

EOF analysis method is used by Reynolds for data reconstruction, as well as in this study to allow for the application of classical time series analysis. It is especially useful for data covering large spatial domains. The data are rewritten as orthogonal spatial and temporal components, thus extracting necessary information regarding the variability.

Consider some data y which vary both spatially as well as temporally. The EOF representation would be

$$y(s,t) = \sum_{j=1}^{M} T_j(t) S_j(s) \tag{1}$$

where t is time, s represents horizontal coordinate points, S_j are spatial fluctuations in the data (spatial patterns), T_j are temporal fluctuations (principle components), and M is the number of orthogonal patterns (equal to the dimension s or t, whichever is smaller). The method requires the construction of the covariance matrix of the original data, $\mathbf{C}$, and the solution of the eigenvalue problem

$$\mathbf{Cx} = \lambda \mathbf{x} \tag{2}$$

where λ is a matrix of M eigenvalues, and $\mathbf{x}$ is the matrix containing all x_j. The M elements of $\mathbf{x}$ are known as the eigenvectors. The temporal or spatial component which is selected as the eigenvector is normalized such that
The M elements of $\mathbf{x}$ are known as the eigenvectors. The temporal or spatial component which is selected as the eigenvector is normalized such that

$$\sum_k (E_j)_k (E_l)_k = \delta_{j,l} \tag{3}$$

5

where δ is the *Kronecker delta* (defined as unity where the indices are equal and it vanishes otherwise), k is a space or time index and E is either a spatial or temporal component. Using this relation, it can be shown that the sum of the squared elements of the non-normalized component are the variances of each EOF. The individual components of the EOF analysis are then ordered by descending percent variance contribution. The physical mechanisms which constitute the largest amount of variability in a region are associated with the largest eigenvalues.

The EOF analysis for this study is applied to the Reynolds SST anomalies in the selected equatorial Pacific domain. The spatial and temporal indices are 480 and 2,093 respectively. The spatial domain contains 2,250 grid points; however, 157 of them correspond to land, and are thus omitted from the data matrix. We find that the first principle component (PC1) constitutes approximately 46 % of the total variance in the SSTs for this region. The second principle component (PC2) constitutes approximately 10 % of the total variance.

2.2   The Spectra of PCs 1 and 2

The raw amplitude spectra of each PC are of primary interest for the identification of the pseudoperiodicities and associated physical processes which contribute to the variability. For example, the largest spectral peaks of PC1 are mainly attributable to ENSO. There are pronounced peaks at frequencies near 0.01670 and 0.02355 cpm, which correspond well with the timing of SST anomalies in the eastern equatorial Pacific (EEP) according to the delayed action oscillator theory of ENSO (Philander 1990, Allan et al. 1996). The theory holds lies in the eastern equatorial Pacific (EEP) according to the delayed action oscillator theory of ENSO (Philander 1990, Allan et al. 1996). The theory holds that a 12-15 month delay between Rossby wave-induced SST anomaly maxima

6
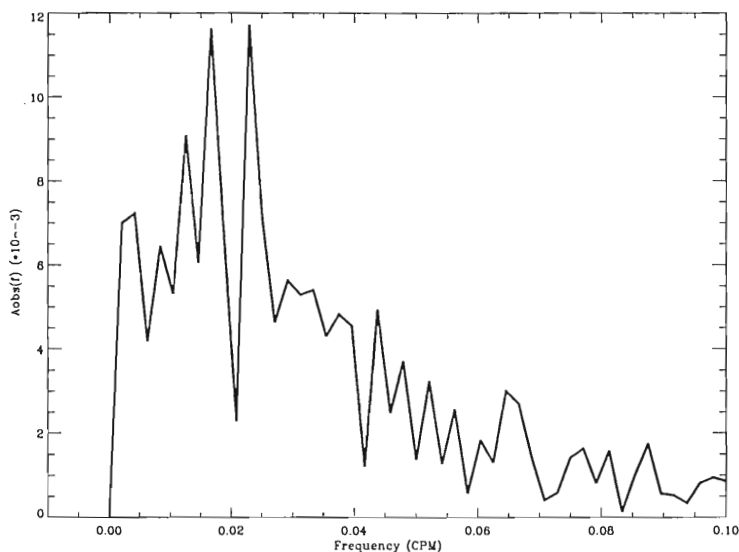
Figure 1: Plot of the amplitude spectrum of principle component 1; $Aobs(f) = \sqrt{(a_i(f)^2 + b_i(f)^2)}$ where a_i, b_i are the Fourier transform coefficients. Frequency is in units of cycles per month (cpm). Note the amplitude peaks which represent intradecadal $f \approx 0.023\ CPM$, $\approx 0.017\ CPM$, $\approx 0.013\ CPM$ ($\approx 3.6\ years$, $\approx 4.9\ years$, $\approx 6.4\ years$) pseudoperiodicities can be linked to the occurrence frequencies of ENSO.

and minima exists in the EEP. The SST maximum occurs during an ENSO warm event. The central/eastern equatorial Pacific westerly wind anomaly is associated with deepening of the thermocline due to the excitation of both down-welling Kelvin waves at the equator and off-equatorial Rossby waves that further enhance downwelling (or weaken upwelling). The SST minimum is indicative of an ENSO cold event. The upwelling Rossby waves reflect off the western boundary of the equatorial Pacific and propagate eastward as upwelling Kelvin waves, reversing the sign of the SST anomaly in the EEP. (Allan et al. 1996). Rossby waves which favor upwelling are thus excited, enhancing the EEP up-welling and corresponding cooling of the SSTs. The negative feedback caused Rossby waves which favor upwelling are thus excited, enhancing the EEP up-welling and corresponding cooling of the SSTs. The negative feedback caused by wave dynamics initially serves to break down the warm event, but later leads
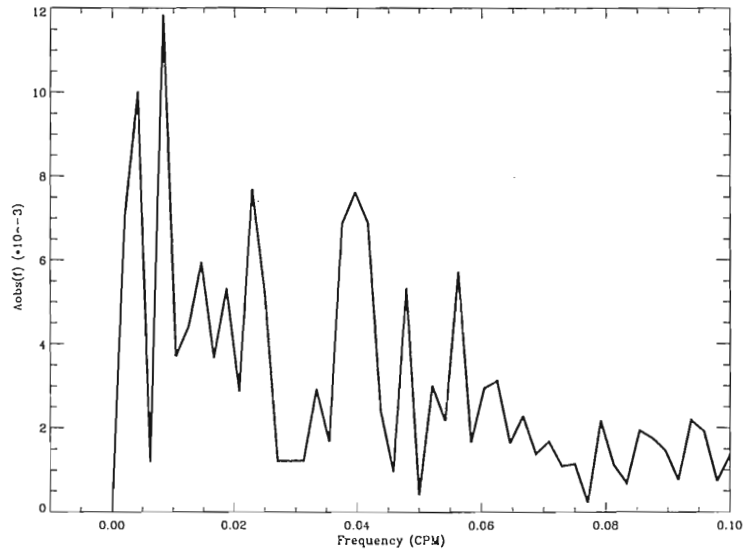
Figure 2: Plot of the amplitude Spectrum of principle component 2. Note the amplitude peaks which represent decadal $f \approx 0.008\ cpm$ ($\approx 10\ years$), intradecadal $f \approx 0.023\ cpm$ ($\approx 3.6\ years$), and biennial $f \approx 0.04\ cpm$ ($\approx 2\ years$) pseudoperiodicities.

to enhanced cooling of SSTs in the EEP, upwelling, and a reversal of the west-erly wind anomaly. Based on the speed of wave propagation and the timing of excitation, the process typically results in SST maxima and minima occurring 12-15 months apart (Allan et al. 1996).

The dominant physical mechanisms represented by spectral peaks in PC2 are ENSO, as well as the decadal and quasibiennial oscillations(Figure 2). The physics of the quasibiennial oscillation (QBO) are not well known; however, the QBO is detected in the SST fluctuations of the equatorial Pacific region, as well as in the troposphere and stratosphere (Salby 1996). The peak in the raw ampli-tude spectrum of PC2 corresponding to approximately 10 years ($f \approx 0.008\ cpm$) supports a fluctuation in SST patterns in the equatorial Pacific on decadal time tude spectrum of PC2 corresponding to approximately 10 years ($f \approx 0.008\ cpm$) supports a fluctuation in SST patterns in the equatorial Pacific on decadal time scales (Figure 1). There is evidence that the amount of solar radiation entering

8

the earth's atmosphere (and associated surface heating) varies, possibly due to sunspot activity, on an approximately decadal time scale (Barnett 1990). Thus, the decadal SST variability could be a result of the sunspot cycle. Enfield and Luis Cid S. (1991) argues that variations in ENSO characteristics on decadal time scales are inversely related to high and low sunspot activity; when sunspot activity is low, ENSO occurrence is found to be more frequent. It is also possible that the variability is related to internal ocean dynamics. Although the physical mechanisms are not known, the decadal variability of SSTs in the tropics remains pronounced (Oceanic Interdecadal Climate Variability 1992).

The three aforementioned quasiperiodicities are considered to prevail over any other SST anomaly fluctuations, such as the intraseasonal, seasonal, or other oscillations in the first two principle components. The amplitude peaks associated with interannual (ENSO), biennial, and decadal oscillations are comparably large relative to other spectral peaks, and thus account for the most variance. Thus, the term *select physical processes* throughout this paper refers to the interannual, quasibiennial, and decadal oscillations collectively.

3. Methodology

The goal in constructing synthetic time series from existing, observed SST data is to determine the important characteristics contained in the observed set and develop a theoretical model function from which those data could have been produced. We approach the model development by first determining the number of Principle Components needing to be modeled based on what meaningful information is contained in the probability structure of the amplitude spectra. We determine mathematical functions that describe both the amplitude and phase spectra in the frequency domain for those EOFs which contain signal information; that is, they contain information of SST fluctuations for dominant physical processes in this region. These are called *significant EOFs* or correspondingly, *significant principle components* (PCs), and are determined, in part, by using methods of EOF truncation.

One common method of EOF truncation assumes that the components which remain distinguishable from noise will possess a steep slope relative to adjacent eigenvalues when plotted on a Scree Graph (Figure 3). The eigenvalues which produce a nearly horizontal line on this graph are most likely components of uncorrelated noise, and they do not represent relevant signals in the data (Wilks 1995). The Scree Graph for Pacific equatorial SSTs implies a truncation point at the second principle component.

Another method of truncation determines which modes describe the selected physical processes discussed in section 2, by analyzing the Fourier amplitude functions of each PC. The SST fluctuations associated with these physical pro- physical processes discussed in section 2, by analyzing the Fourier amplitude functions of each PC. The SST fluctuations associated with these physical processes are identified as peak values in the amplitude function at the correspond-
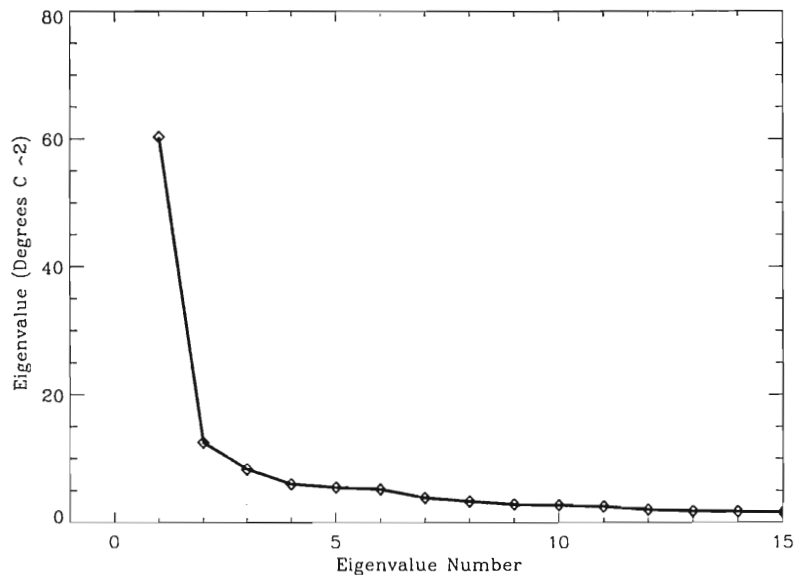
Figure 3: Plot of the eigenvalues (representing variance contribution ($^{\circ}C^2$)) as a function of eigenvalue rank. The slope of the line between the eigenvalues is related to the amount of signal information in those EOFS. The eigenvalue itself is known as the *observed variance* of the component.

ing frequencies. Recall from section 2 that the ENSO exists in the amplitude function of PC1, while the ENSO, biennial and decadal variabilities are found in PC2 (Figure 6). No other large amplitude peaks are identified as recognizable and dominant physical processes in the amplitude functions of PCs 3 and higher. Therefore, the first and second components are categorized as *significant PCs* because their spectra contain peaks attributable to physical processes. This result is consistent with the Scree truncation.

A Kolmogrov-Smirnov (K-S) goodness of fit test is applied to the spectra of PCs 3 through 11 to determine whether or not the remaining PCs can be considered random noise. The range 3 through 11 is selected because the slope of PCs 3 through 11 to determine whether or not the remaining PCs can be considered random noise. The range 3 through 11 is selected because the slope between the eigenvalues on the Scree diagram does not appear to be horizontal,

11

and thus may not simply represent uncorrelated noise. The K-S test compares the magnitude of the maximum difference between the integrated spectrum of the data and that of pure random noise (Priestly 1981). A confidence limit is selected apriori. The null hypothesis states that the sample cumulative distribution function (cdf) of the data will not be distinguishable from the theoretical cdf of a random process. PCs 3-11 *are* distinguishable from white noise at the 95 % confidence level. Thus, PCs 3-11 are not simply uncorrelated noise. The spectra contain much larger amplitude peaks at lower frequencies than the peaks at higher frequencies. We deduce that some amount of temporal correlation exists in PCs 3-11; however, the amplitude peaks at low frequency cannot be attributed to any obvious and well described properties. Thus, we have selected the term *red noise* to describe that portion of the temporal fluctuation that is assumed to be neither deterministic nor random. Red noise is selected because it is a common term describing correlated noise; characterized by a spectrum containing larger variance at lower frequencies. For example, the spectra of autoregressive processes with coefficients greater than zero are known as red noise processes (Wilks 1995).

We expect the spatial and temporal variations in the data associated with EOFs of low percent variances to be purely random. Therefore, PCs 12 through 480 represent white noise. These PCs constitute only 15 % of the total variance in the system. Individually their observed variances are less than or equal to $2.0^{o}C^{2}$, with an average observed variance of only $0.04^{o}C^{2}$. Additionally, on the Scree diagram the lines become most nearly horizontal at PCs 12 and higher. $2.0^{o}C^{2}$, with an average observed variance of only $0.04^{o}C^{2}$. Additionally, on the Scree diagram the lines become most nearly horizontal at PCs 12 and higher. For these reasons, we assume that random numbers, rescaled by equipartitioned

original variance can represent the random fluctuations in both space and time for PCs 12 through 480.

A distinction has been made, for the purpose of EOF truncation, between the portions of the amplitude functions which represent select physical processes, correlated noise, and stochastic noise. Thus, we will develop mathematical functions to denote the three separate amplitude models: a *deterministic amplitude model*, a *red noise model*, and a *white noise model* to correspond respectively to these three distinctions.

3.1   White noise amplitude model

We develop a theoretical model for the portion of the spectra which represents uncorrelated noise. Since random noise is by definition uncorrelated from one time step to the next, it is manifested in the amplitude spectrum as high frequency information. Thus, the white noise amplitude model is taken to be the mean value of the original amplitude function for the frequency band from f_c to the Nyquist frequency, $\frac{1.}{2 months}$. The value f_c is that frequency between low frequency red noise or deterministic (physical) signal and higher frequencies in the amplitude function. The determination of the frequency dividing correlated and uncorrelated information is chosen subjectively from the spectrum of each PC. We select f_c to be large enough so that we do not model relevant physical information as uncorrelated noise. Let M represent the sample mean of the amplitude values from the original spectrum in this defined white noise frequency range. Then, the white noise model is defined over all frequencies such that

range. Then, the white noise model is defined over all frequencies such that

$$W(f_i) = \begin{array}{ll} M, & |f_i| > 0 \\ 0, & f_i = 0 \end{array} \qquad (4)$$

13

where i is the frequency index. From this model, the time domain white noise variance can be computed. The remaining variance is attributed to either red noise processes or deterministic processes.

3.2   Red noise amplitude model

The modeled red noise represents the non-deterministic portion of the PC which still contains some temporal correlation. An autoregressive model was selected as the basis for the red noise amplitude model. The general order autoregressive model is described as a regression of the variable X over time,

$$X_t = \sum_{k=1}^{K} \phi(k) X_{t-k} \tag{5}$$

where K denotes a chosen maximum lag, from k=1,2,3,...K, and $\phi(k)$ are the autoregressive coefficients.

An estimate to the autocorrelation function is needed to determine the necessary coefficients, ϕ. We are interested in the approximate correlation of *only* the red noise over time. The sample autocorrelation functions of the PCs cannot be used, as they contain information from the deterministic processes as well as red noise. Thus, we select an indicative red noise autocorrelation function which drops off exponentially such that

$$\rho = exp(-\frac{k}{\gamma}) \tag{6}$$

where k=0,1,2,3,...K represents the lag, and γ represents a reference lag. Such an estimate of the autocorrelation function represents a temporal correlation for the first few lags and approaches zero correlation at higher lags. The autoregression estimate of the autocorrelation function represents a temporal correlation for the first few lags and approaches zero correlation at higher lags. The autoregression coefficients are determined from the autocorrelation function by solving the

14

Yule-Walker equations for all lags (Wilks 1995). The matrix equation

$$\mathbf{R\Phi} = \mathbf{r} \tag{7}$$

consists of the autocorrelation matrix

$$
\mathbf{R} = \begin{pmatrix}
\rho(0) & \rho(1) & \cdots & \rho(k-1) \\
\rho(1) & \rho(2) & \cdots & \rho(k-2) \\
\rho(2) & \rho(3) & \cdots & \rho(k-3) \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\rho(k-1) & \rho(k-2) & \cdots & \rho(0)
\end{pmatrix}, \tag{8}
$$

the column vector of the autoregressive coefficients

$$
\mathbf{\Phi} = \begin{pmatrix}
\phi(1) \\
\phi(2) \\
\cdot \\
\cdot \\
\phi(K)
\end{pmatrix}, \tag{9}
$$

and the autocorrelation column vector

$$
\mathbf{r} = \begin{pmatrix}
\rho(1) \\
\rho(2) \\
\cdot \\
\cdot \\
\rho(K)
\end{pmatrix} \tag{10}
$$

The result is that the coefficients, ϕ, are determined for all lags (k = 1,2,3,...K), and they are used to fit an autoregressive model, initially in the time domain. The Fourier Transform of the autoregressive model represents a preliminary red noise amplitude model in the frequency domain.

The red noise model must be scaled to match the time domain variance of the red noise processes being modeled. Thus, we need to quantify the time domain variance of the red noise process for each PC. To do this, we developed the red noise processes being modeled. Thus, we need to quantify the time domain variance of the red noise process for each PC. To do this, we developed a method called the *overlap technique*. The modeled red noise can be scaled

15

to the peak value of the original amplitude function in the low frequency range 0.0 to 0.007 cpm.

The time domain variance of this model represents an upper bound guess for red noise, $\sigma^2_{R_{max}}$, as it is the amplitude that the red noise would have if there were no other deterministic or random information in the frequency band. Using the white noise variance for the PC of interest, we can compute a model deterministic function with the remaining variance. The variance of
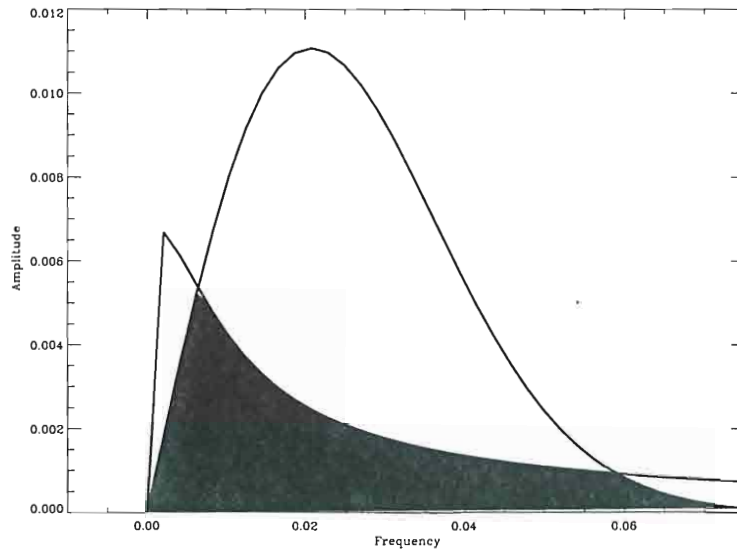
Figure 4: The shaded region is an example overlap region for the maximum possible red noise variance and the corresponding deterministic amplitude model. The time domain variance of the overlap region is $\Delta\sigma^2_R$.

the overlap region of these two functions denotes the largest possible red noise

deviation, $\Delta \sigma^2$ (Figure 4). We determine the lower bound red noise variance

the overlap region of these two functions denotes the largest possible red noise

deviation, $\Delta \sigma^2_R$ (Figure 4). We determine the lower bound red noise variance

to be

$$\sigma^2_{R_{min}} = \sigma^2_{R_{max}} - \Delta \, \sigma^2_R \tag{11}$$

The red noise variance is taken to be the average of its upper and lower bounds. Thus, the deterministic variance is the remaining portion of the total time domain variance for the significant PCs

$$\sigma^2_{det} = \sigma^2_{total} - \sigma^2_R - \sigma^2_w \tag{12}$$

where σ^2_w denotes the white noise variance and σ^2_{total} denotes the total time domain variance for the PC being modeled.

3.3 Deterministic amplitude model

The amplitude peaks which occur at or near frequencies of the select physical processes, will be modeled with rescaled probability density functions (PDFs)in the frequency domain (Bendat and Piersol 1986). PDFs are selected because the area under the PDF curve provides meaningful probability information for each frequency. Thus, each of the select physical processes are considered to be aperiodic and are best represented as deterministic distributions about their peak occurrence frequency. A good example is the well-known range of 2 to 7 years for the return interval of an ENSO event. An amplitude peak for ENSO would reflect the *distribution* of occurrence probability for each frequency in this range, rather than at a single frequency. We argue that such an amplitude distribution over the frequency range for physical processes is deterministic.

The three PDFs which will be used have been determined to closely fit distribution over the frequency range for physical processes is deterministic.

The three PDFs which will be used have been determined to closely fit the shape of the distribution about the frequency of the amplitude peak. The

17

Gaussian PDF is used, as are the Rayleigh and the Maxwell PDFs. The latter two distributions are specific cases of the more general Wiebull PDF (Figure 5). PDFs have the important characteristic that
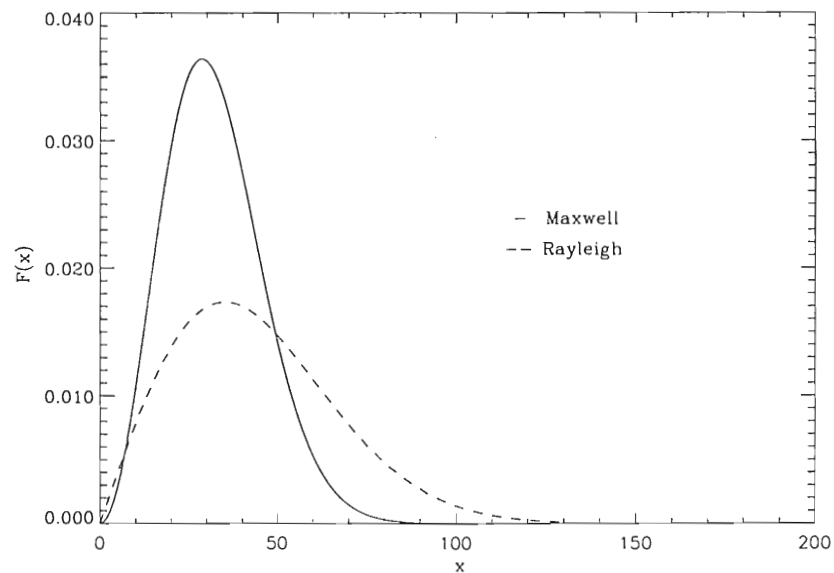
Figure 5: Example Rayleigh and Maxwell PDFs. The Maxwell distribution has a sharper peak and drops off more quickly than the Rayleigh.

$$\sum_x F(x)\Delta x = 1 \tag{13}$$

where $F(x)$ represents the distribution and Δx is the sampling interval. Equation (13) reiterates that the area under the PDF is a probability. Moreover, the shape of the distribution of probability over a given frequency range marks the important difference between these PDFs.

3.4 Specific model development

To construct the model for each individual PC, we will utilize the aforementioned
3.4 Specific model development

To construct the model for each individual PC, we will utilize the aforementioned amplitude models as needed to represent the probability distribution over the

18

frequencies of select physical processes. Further, we compute the total time domain sample variance directly from the original time series of each PC. The corresponding modeled time series must retain this variance.

For the significant principle components, the modeled amplitude function is comprised of a root mean square sum of all three amplitude models. Much of the significant variability occurs in the frequency range 0.0 to 0.1 $month^{-1}$, corresponding to time scales longer than 10 months. We assume that all three portions of the amplitude model overlap in this range, contributing to the larger variance at low frequencies. Physically, this assumption could imply an interaction between deterministic processes, linearly correlated noise, and purely random noise.

For PC1, the ENSO signal has peaks near frequencies of 0.02355, 0.01670, and 0.0130 (3.5, 5, and 6.4 years respectively) which have amplitude distributions best modeled with rescaled Gaussian probability density functions (PDFs)

$$G(f) = (\frac{1}{(\sigma_y)(\sqrt{(2\pi)}})exp-(\frac{(f-f_o)^2}{2\sigma_y^2}) \tag{14}$$

where f_o is the Gaussian peak frequency. Moreover, the ENSO signal also has a peak at the 0.0353 frequency (2.4 years); however, the rate at which the amplitude peak drops off with increasing frequency is most closely reproduced with a rescaled Maxwell PDF

$$M(f) = (\frac{f^2}{c^3})\sqrt{(\frac{2}{\pi})}exp-(\frac{f^2}{2c^2}) \tag{15}$$

where f are the frequencies and c is a constant given by:

$$M(f) = (\frac{}{c^3})\sqrt{(\frac{}{\pi})}exp-(\frac{}{2c^2}) \tag{16}$$

where f are the frequencies and c is a constant given by:

$$c = \frac{f_{oM}}{\sqrt{3}} \tag{16}$$

19

with f_{o_M} representing the Maxwell peak frequency. The amplitude at this peak is rescaled to match the amplitude of the original peak.

For PC2, the decadal peak was modeled with a rescaled Maxwell PDF, because the amplitude peak is relatively sharp and drops off with increasing frequency at a rate close to that of the Maxwell distribution. The remaining ENSO signal peaks at 3.6 years and was modeled with a rescaled Rayleigh PDF, due to its slow rate of drop-off and wide peak.

$$R(f) = (\frac{f}{c^2}) exp-(\frac{f^2}{2c^2}) \tag{17}$$

$$c = f_{oR}{}^2 \tag{18}$$

The two year peak was modeled with a rescaled Gaussian PDF, as the distribution in this range is approximately symmetric about the peak. The final amplitude function for the PC being modeled is obtained by computing the root mean square addition of the individual rescaled PDFs.

The models for the amplitude functions of PCs 3 through 11 are comprised of two parts. First, the white noise in each component is modeled using the aforementioned f_c to f_{Nyq} methodology. Since no deterministic amplitudes contribute to these components, the low frequency peaks are modeled using the red noise model methodology; however, the determination of the red noise variance is simply

$$\sigma_R^2 = \sigma_{total}^2 - \sigma_w^2 \tag{19}$$

thereby eliminating the need for the overlap technique to model these compo-
$$\sigma_R^2 = \sigma_{total}^2 - \sigma_w^2 \tag{19}$$

thereby eliminating the need for the overlap technique to model these components.

After modeling PCs 3 through 11 with an amplitude function, we model the corresponding phase functions for each. The phase functions of the original PCs are given by the Fourier coefficients a_i and b_i, such that:

$$\theta = ATAN(\frac{-b_i}{a_i}) \quad for \quad a_i > 0 \tag{20}$$

$$\theta = ATAN(\frac{-b_i}{a_i}) - \pi \quad for \quad a_i < 0, \ b_i > 0 \tag{21}$$

$$\theta = ATAN(\frac{-b_i}{a_i}) + \pi \quad for \quad a_i < 0, \ b_i \geq 0 \tag{22}$$

These phases are contained in the interval $-\pi, \pi$, and are antisymmetric about the zero frequency; $\theta(f) = -\theta(-f)$. No obvious correlation between adjacent phases was found in these spectra. Thus, the final model phases are justifiably random numbers in this interval. The model phase function is $e^{j\theta(f)}$. Thus, in the frequency domain we multiply our model amplitude by our modeled phases for each PC. The inverse Fourier transform for each yields a synthetic time series. These are the modeled temporal components.

Finally, a model is needed to represent the random spatial and temporal SST fluctuations of PCs 12 through 480. A single model PC can represent the necessary random fluctuations in both space and time and can be rescaled to match the *observed variance* of all 468 remaining components. The observed variance is obtained from the sum of the squared spatial patterns of all remaining EOFS. Thus, we construct a single model spatial pattern that consists of Gaussian random noise with equipartitioned observed standard deviation over each point in our spatial domain. By simply generating Gaussian random numbers and rescaling by the observed standard deviation for all grid points in the each point in our spatial domain. By simply generating Gaussian random numbers and rescaling by the observed standard deviation for all grid points in the domain, for all 480 months, a final modeled EOF is produced.

We are now able to produce a complete data set of synthetic SST anomalies for the equatorial Pacific by projecting the modeled temporal components on the associated spatial patterns. By taking the product of the spatial and temporal parts of the first 11 modeled PCs, we project the modeled temporal variation onto the original spatial patterns. The true standard deviation is now modeled in the first 11 respective EOFs. By adding the final modeled EOF to EOFs 1-11, we have constructed a complete synthetic data set for the same spatial domain and temporal extent as the Reynold's Reconstructed SST anomalies.

4. Results

The modeled *significant PCs* are compared to the corresponding PCs from the Reynolds data to validate the success of the individual component models. Then, a sample of complete synthetic data sets are compared to the complete Reynolds data set. The ENSO SST signature for the JMA domain will be used as a standard of comparison. The frequencies of ENSO events are reproduced. Moreover, the average number of respective warm and cold events are retained in the synthetic data for a forty year period. Statistical inference is made regarding the expected return period of an extreme ENSO event. The extreme ENSO warm event reaching an SST anomaly magnitude of 2^oC occurs more frequently than an equivalent magnitude cold event.

4.1 Comparison of Model PCs with original PCs

By determining an analytical approximation to the amplitude spectra of PCs 1 and 2, we have constructed a theoretical amplitude function from which the observed data can be derived, and from which statistically consistent synthetic data are derived. We have developed, via spectral analysis, a useful variance decomposition for the original amplitude spectrum into two or three distinct amplitude functions for each PC being modeled. As a result, we have gained a much better understanding of the possible physical mechanisms which compose the amplitude structure and what amount of the variability is likely attributable to noise; either correlated or uncorrelated.

The model amplitude functions for PCs 1 and 2 are analytical approxi-
to noise; either correlated or uncorrelated.   good fit to the

The model amplitude functions for PCs 1 and 2 are analytical approximations to the original amplitude functions which represent a good fit to the
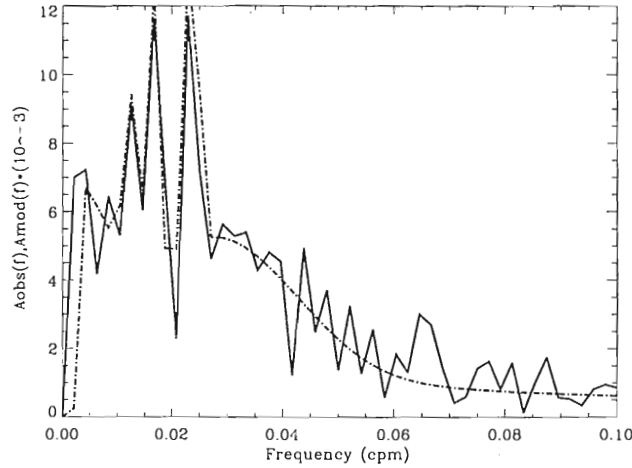
Figure 6: Amplitude Spectrum for the first principle component with Modeled Amplitude Spectrum Overlay. The solid line is the amplitude spectrum for the original PC1. The dotted line is the modeled amplitude spectrum.

relevant spectral peaks and to the overall distribution of red and white noise amplitudes over frequency (Figures 6, 7). The model for PC1 represents a good fit to the amplitude distribution corresponding to the ENSO pseudoperiodicities; which constitute the largest percent variance for all ENSO events in the Reynold's data. Moreover, the largest amount of variance in the spectrum is modeled with deterministic ENSO peaks, while much less variance is attributed to the red and white noise processes. The model for the spectrum of PC2 captures the amplitude distributions of the decadal and biennial pseudoperiodicities, as well as the remaining ENSO pseudoperiodicity. The ENSO peak in PC2 constitutes much less variance than the ENSO peaks of PC1. The red noise peaks, in both PC1 and PC2, give accurate variance contribution at frequencies smaller than 0.007 cpm, which could conceivably represent the leakage of longer-scale signals into the spectra of the 40 year set.

quencies smaller than 0.001 cpm, which contained for each synthetic time series of longer-scale signals into the spectra of the 40 year set.

The statistics of the original data are retained for each synthetic time series
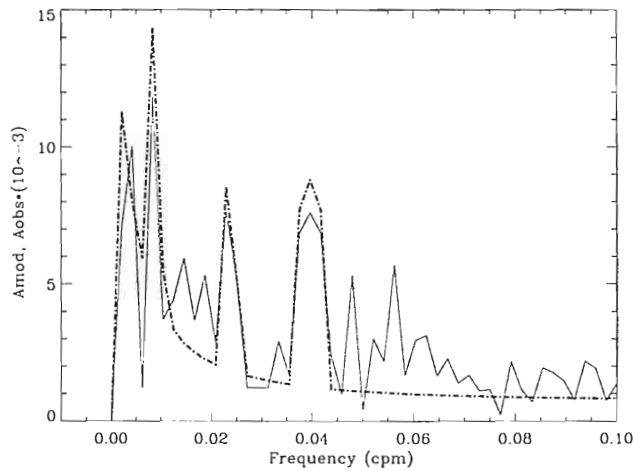
Figure 7: Amplitude spectrum for the second principle component with modeled amplitude spectrum overlay. The solid line is the amplitude spectrum for the original PC1. The dotted line is the modeled amplitude spectrum.

generated from this method. The modeled PCs retain the normalization crite-ria given by equation (3) in section 2, and the final reprojected EOFS retain the original variances. The red, white, and deterministic amplitude models are defined to be zero at $f = 0$; the frequency domain equivalent of a zero tempo-ral mean. Thus, any small, non-zero mean resulting from the computation of anomalies from climatology data is removed for the model development process and added in at the end. The autocorrelation function is an important statistic of the realization, as it represents the dependence of the value of the data at one instant with the value separated by an interval, (or lag), τ. Periodicities occurring in the time series are manifested as sinusoidal fluctuations with in-creasing lag in the autocorrelation function. Both modeled components closely reproduce the temporal correlation of the Reynold's data out to approximately 100 lags (Figures 8 and 9). The autocorrelation function (ACF) for PC1 con-reproduce the temporal correlation which fluctuate between 40 - 60 month (3-6 100 lags (Figures 8 and 9). The autocorrelation function (ACF) for PC1 con-tains the ENSO pseudoperiodicities which fluctuate between 40 - 60 month (3-6
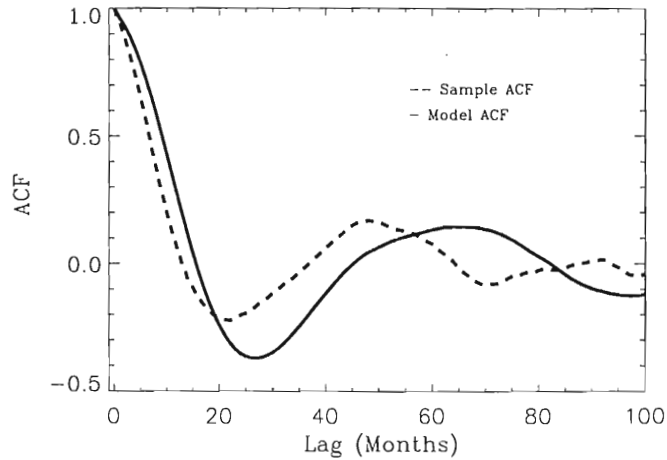
25

Figure 8: Autocorrelation function for the first principle component with modeled auto-correlation function overlay.

year) lags, and are also correlated near 80 lags (7 years) (Figure 8). Likewise, the autocorrelation function for PC2 shows pseudoperiodicity, with peak correlations associated with the decadal, ENSO, and biennial oscillations at lags 20 to 30 months (1.6-2.5 years), 50 months (4 years) and 65 months (5.4 years) (Figure 9).

4.2 Complete Data Set Comparison

A consistent point of comparison is necessary to determine the success of the model in producing synthetic data which is a valid reproduction of the original data. Ten independent synthetic SST time series are produced, and the criteria for the Japan Meteorological Agency (JMA) index is applied.

The JMA index is computed as a five month running mean of spatially averaged SST anomalies for the region 4° S to 4° N and 150° W to 90° W. From this index, an ENSO warm event is identified by having an SST anomaly value eraged SST anomalies for the region 4° S to 4° N and 150° W to 90° (known this index, an ENSO warm event is identified by having an SST anomaly value greater than or equal to 5° C for a minimum of 6 consecutive months (Japan
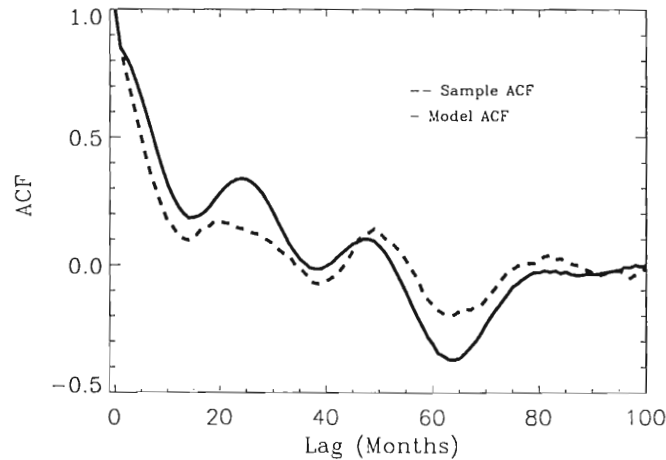
26

Figure 9: Autocorrelation function for the second principle component with modeled autocorrelation function overlay.

Meteorological Agency 1991). Since the synthetic data do not necessarily correspond to particular months or years, we do not include the JMA criteria that October, November, December be three of the 6 months in the event. Further, following Sittel (1994) a symmetric definition will be utilized for the identification of ENSO cold events.

The Reynold's JMA time series contains 11 warm events and 8 cold events. The average number of ENSO events from the JMA time series for the ten runs of synthetic data are 10 and 7, for warm and cold events respectively. 95 % of the mean number of events from the synthetic data will fall within the range 9.4 and 11.0 for a warm event and between 6.3 and 7.5 for a cold event. Thus, the Reynolds values for the mean number of events fall within the range of two standard deviations from the mean number of synthetic events. Moreover, the average duration of ENSO events for Reynold's data is 10 months for cold events and 13 months for warm events. In the synthetic data, the average duration of average duration of ENSO events for Reynold's data is 10 months for cold events and 13 months for warm events. In the synthetic data, the average duration of an ENSO event is 12 months for a cold event and 15 months for a warm event.

Thus, the JMA time series for the synthetic data closely reproduces the ENSO frequency and duration of the original data.
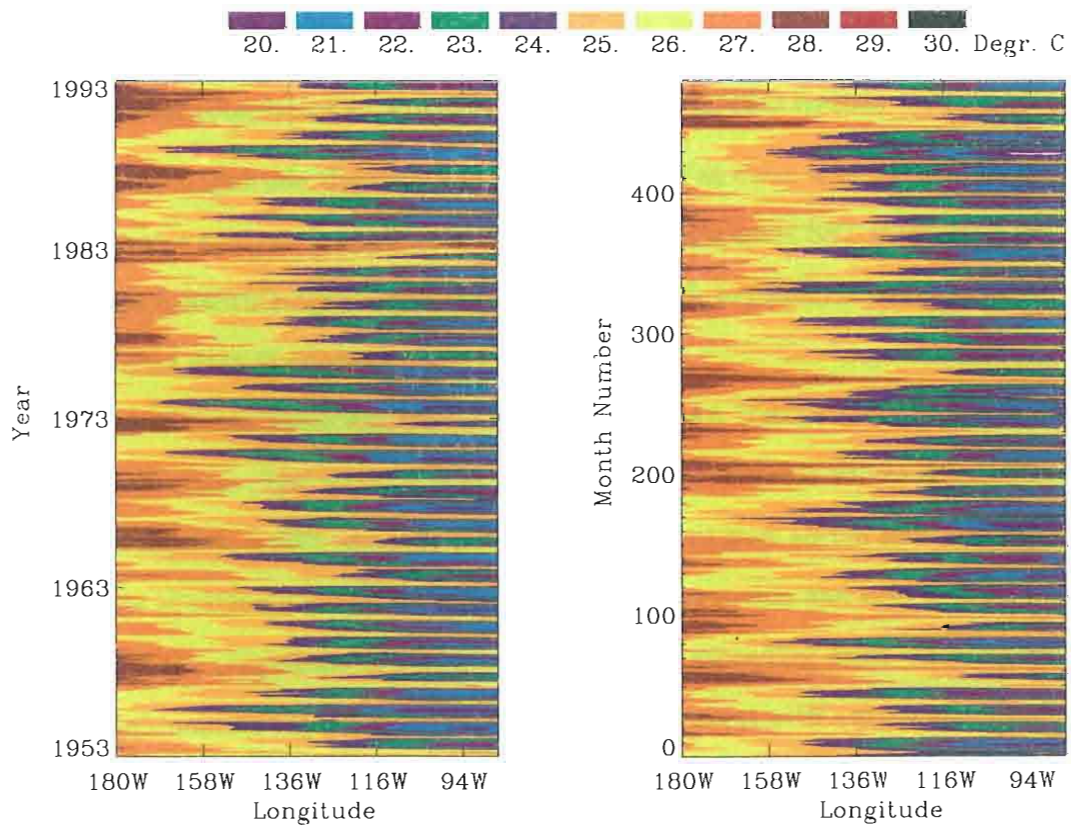

Figure 10: (a). Hovmuller diagram of Reynolds SSTs (in degrees Celsius) along the equator. (b). Hovmuller diagram of synthetic SSTs (in degrees Celsius) along the equator.

The ENSO events are characterized by an east or west displacement of isotherms which disrupt the usual regions of intense tropical convection. In a warm event, the 28^oC isotherm can be displaced to the central and eastern equatorial Pacific, and in an extreme case such as the 1982-83 El Niño, the a warm event, the 28^oC isotherm can be displaced to the central and eastern equatorial Pacific, and in an extreme case such as the 1982-83 El Niño, the

28

isotherm can be displaced as far as 90W (Figure 11). Tropical convection has been associated with sea surface temperatures equal to or in excess of 28^oC (Enfield 1989). During a cold event, there is a displacement of the 28^oC isotherm toward the western equatorial Pacific.

4.3   Characteristics of Future, Unobserved ENSO events

When the magnitude, duration and/or frequency of an event is of interest, an extreme value analysis is applicable. We utilize the maximum SST magnitude per event to constitute our sample of extreme values. A representative distribution of extreme data is selected.

The extreme value analysis is done on the JMA indices for all 10 sets of synthetic data. The method is applied to a set of extreme values selected to be the maximum SST anomaly magnitude per event over all ten sets. Separate samples of extreme values are determined for warm events and cold events. By plotting the magnitude of the extreme values in ascending order on a log-log plot, information regarding the return period of an event with a given magnitude can be determined (Gumbel 1958). For a warm event, an SST magnitude of 2^oC can be expected to occur approximately every 8 warm events (Figure 11a). For a cold event, an SST magnitude of 1.8^oC can be expected to occur approximately every 8 cold events (Figure 11b).

According to Enfield and Luis Cid S. (1991), the most probable return period for the ENSO warm event is 3-4 years, while that of the Strong (S) and Very Strong (VS) category events (Quinn and Neal 1987), is 9-12 years. We find that events with SST anomaly magnitudes of 1.4^oC occur in the range of 9-Strong (VS) category events (Quinn and Neal 1987), is 9-12 years. We find that events with SST anomaly magnitudes of 1.4^oC occur in the range of 9-12 years, if we assume an ENSO return period of 3-4 years(Figure 11a). Lau
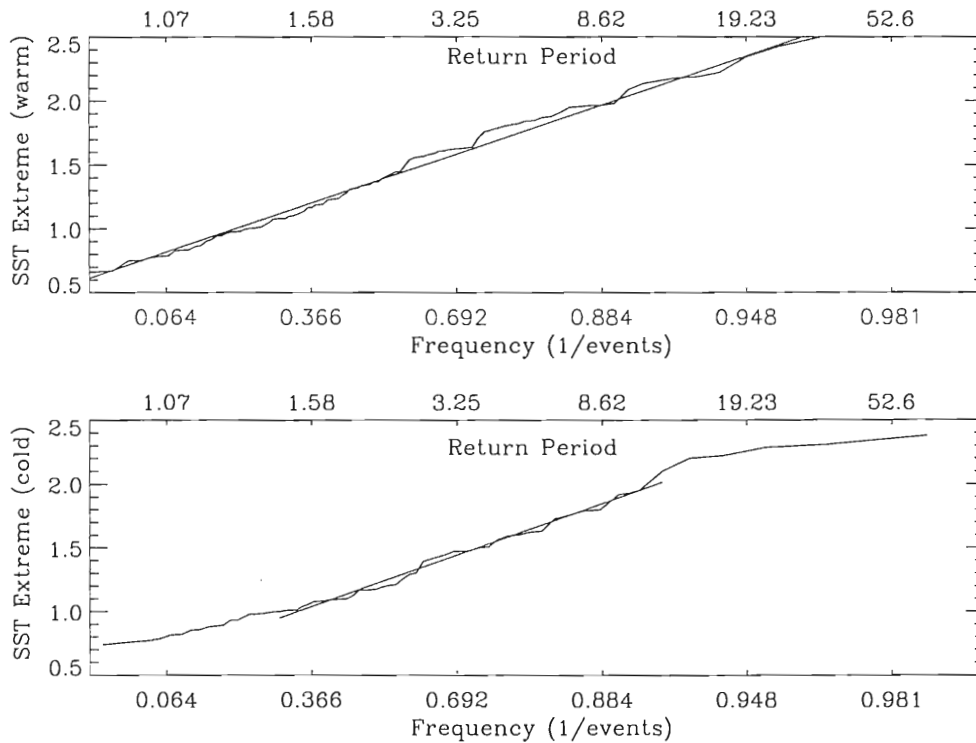
Figure 11: (a). Plot of the return period of a given SST anomaly magnitude per warm event. (b). Plot of the return period of a given SST anomaly magnitude per cold event.

(1985) utilizes statistical methods to determine the average return period of Super-ENSOs to be every 30 to 40 years. A warm event with a maximum SST anomaly magnitude of 2.0^oC is expected to occur every 7 to 10 events; corresponding to a range of 30-40 years for an ENSO return period of 4 years. For a cold event, the value of 1.7^oC will occur at approximately 30-40 years, for an ENSO return period of 4 years.

5. Conclusions

The principle components of the Reynolds reconstructed SST data set are used to determine the dominant physical processes in SST anomaly fluctuations via spectral analysis. An EOF analysis of the Reynolds data determines and ranks the spatial and temporal modes. The dominant physical processes are identified by their spectral peaks and the percent variance contribution. PC1 describes 46 % of the total variance, and contains the ENSO pseudoperiodicities occurring on time scales between 2.4 and 6.5 years. PC2 contains the ENSO (3.6 years), biennial, and decadal pseudoperiodicities as large amplitude peaks. PCs 3 through 11 contain correlated red noise and band-limited white noise. Thererfore, the amplitude functions can be divided into three distinct parts, based on the results of the spectral analysis and the determination of the significant EOFS. The deterministic, red noise, and white noise amplitude models are the separate analytical functions developed to approximate the parts of the spectra for the principle components via the methodology described in section 3. Random phases are applied to the spectral amplitude models, permitting the generation of numerous statistically indistinguishable SST anomaly data sets.

The time series generating method is useful for producing a large sample of SST anomaly data for statistical testing. Furthermore, by retaining the statistics of the observed data (specifically the ACF), it is possible to make valid statistical inferences about processes like the ENSO. In section 4, we found that the return period of an extreme ENSO warm event having a maximum SST anomaly magnitude of $2^\circ C$ occurs approximately every 8 warm events.
the return period of an extreme ENSO warm event having a maximum SST anomaly magnitude of $2^\circ C$ occurs approximately every 8 warm events.

Further use for these synthetic data could be as forcing input in coupled

31

ocean-atmospheric or atmospheric numerical models. The envelope of atmospheric response to various SST anomaly forcing associated with the ENSO and other pseudoperiodicities can be understood by combining the work of this study and the current modeling studies of coupled air-sea interaction.

Appendix A. Final model amplitude functions for PCs3 through 11

The final amplitude models for components 3 through 11 are root mean square sums of the red noise and white noise amplitude models which best represent the variance drop-off in the frequency domain.
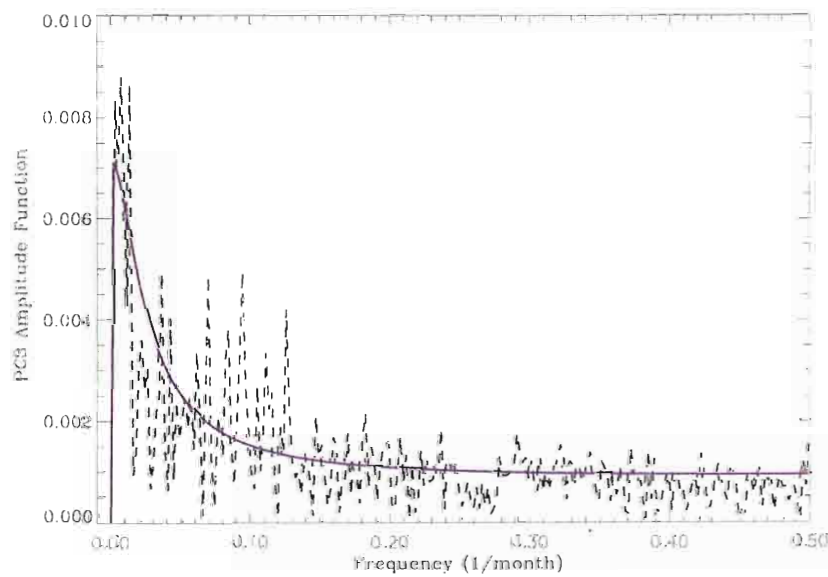

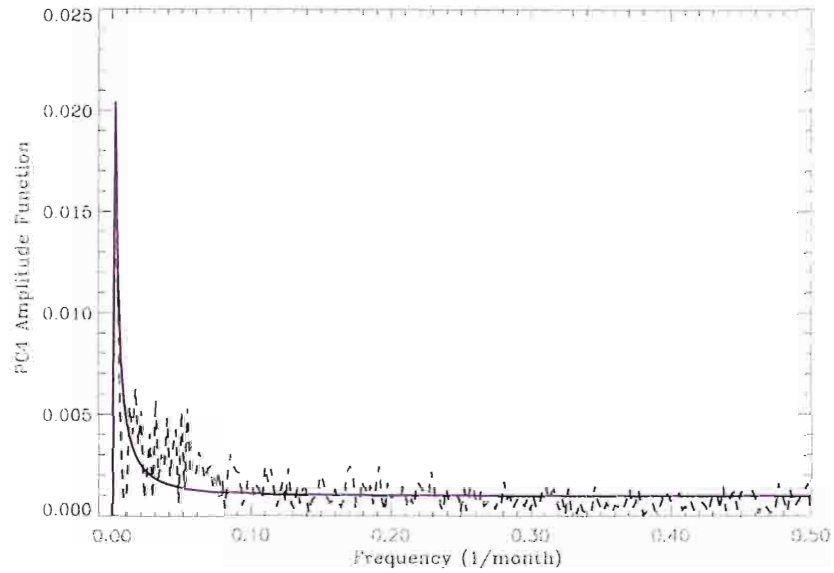Figure 12: Plot of the amplitude function of principle component 3, with model amplitude function overlay.

Figure 13: Plot of the amplitude function of principle component 4, with model amplitude function overlay.
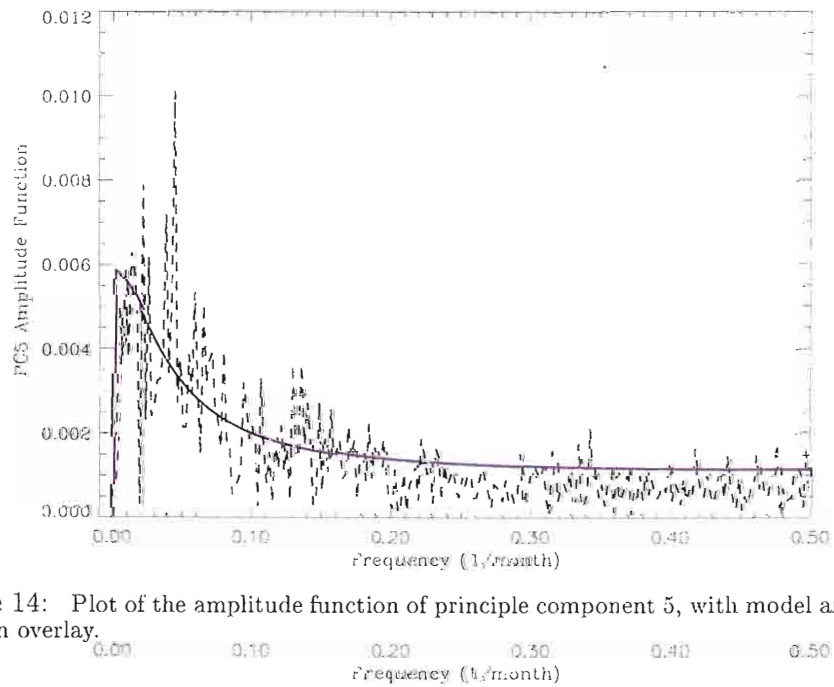

Figure 14: Plot of the amplitude function of principle component 5, with model amplitude function overlay.

34

Figure 15: Plot of the amplitude function of principle component 6, with model amplitude function overlay.

Figure 16: Plot of the amplitude function of principle component 7, with model amplitude function overlay.
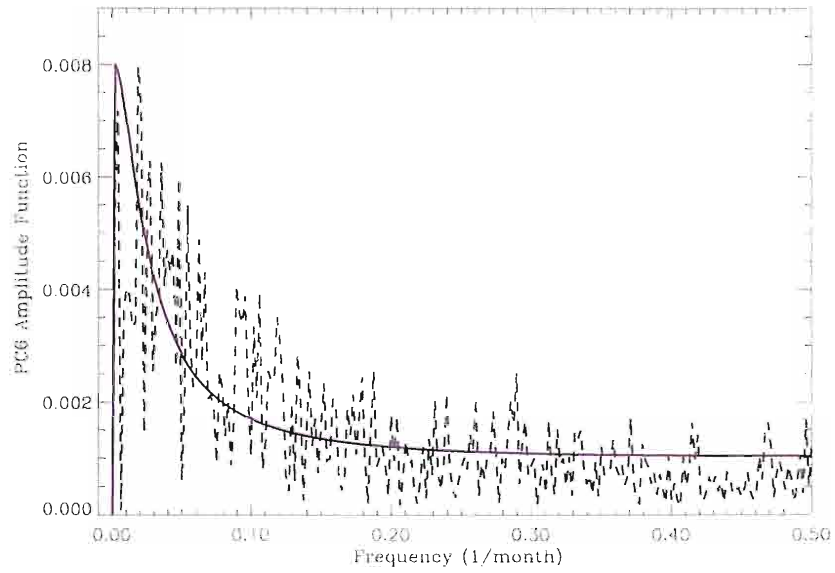
Figure 17: Plot of the amplitude function of principle component 8, with model amplitude function overlay.
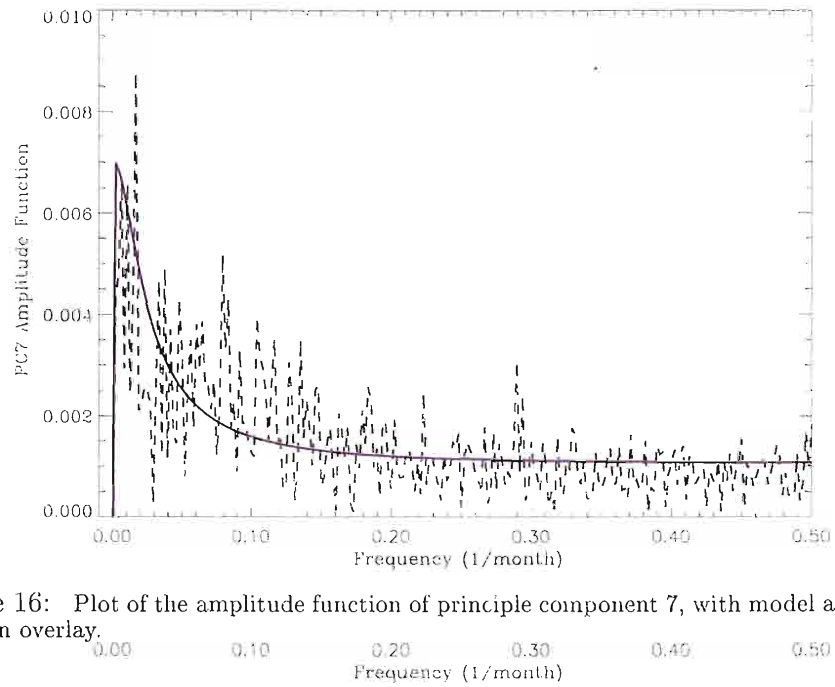

Figure 18: Plot of the amplitude function of principle component 9, with model amplitude function overlay.
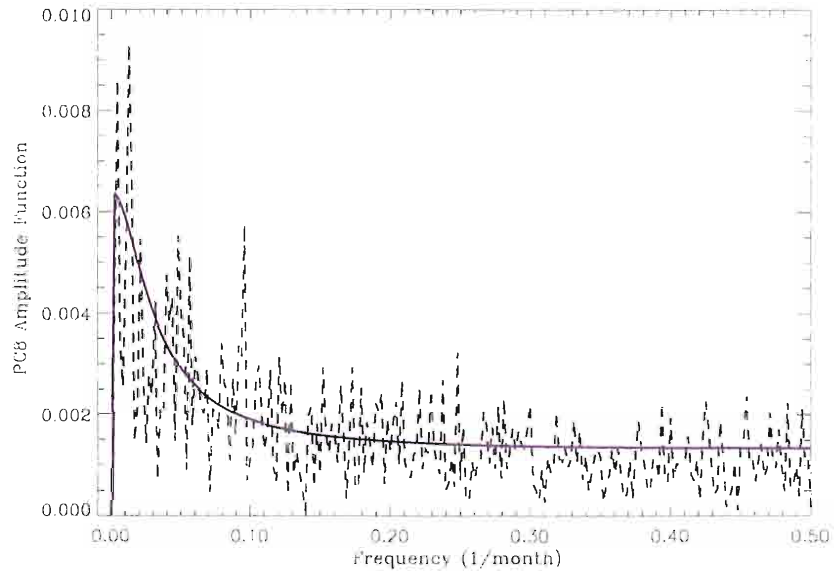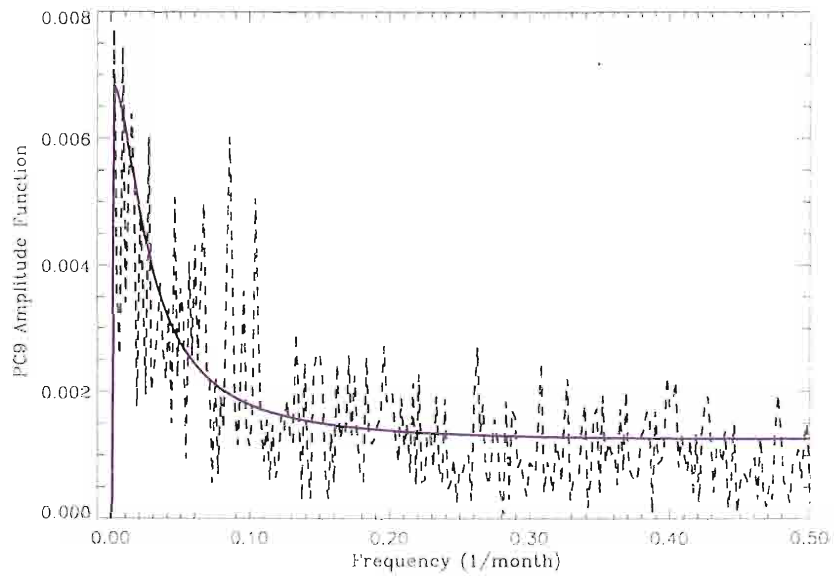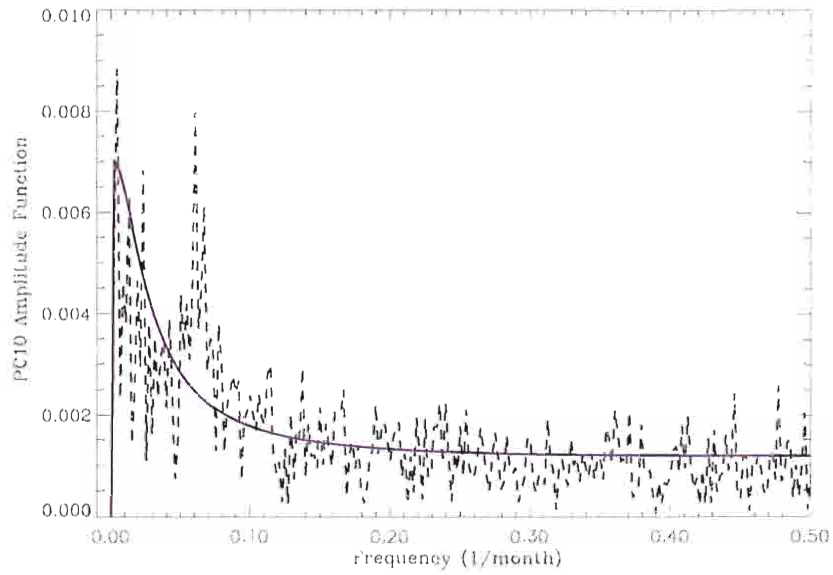
Figure 19: Plot of the amplitude function of principle component 10, with model amplitude function overlay.
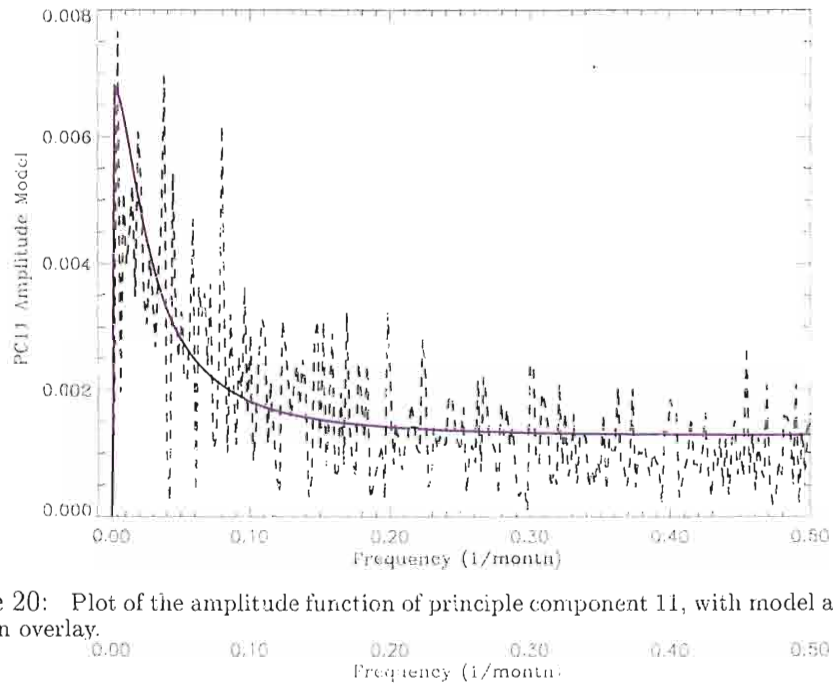

Figure 20: Plot of the amplitude function of principle component 11, with model amplitude function overlay.

Figure 20: Plot of the amplitude function of principle component 11, with model amplitude function overlay.

37

Appendix B. Red Noise Model Parameters for PCs1 through 11

Given equation (6) in section 3.2, a separate value for the cut-off lag Γ, is needed for each PC, to determine the sample ACF corresponding to the red noise portion of the amplitude spectrum. The value selected depends on the magnitude of the amplitude peak(s) at frequencies less than 0.007 cpm, as well as the rate of drop-off of the low frequency amplitudes. Moreover, a value X_o is needed to serve as a first guess to the AR time series model. The time domain variance changes as the value X_o changes, and thus it is essentially a *scale parameter* for the red noise model. Both X_o and Γ have been selected for PCs 1-11 as shape and scale parameters for the respective red noise amplitude models (Table 1).

Table 1: The final values for the selected lag, gamma, and for the initial value of the Autoregression, Xo.

Component	gamma	Xo
PC1	20	0.170
PC2	54	0.124
PC3	9	0.401
PC4	54	0.150
PC5	5	0.617
PC6	9	0.452
PC7	9	0.392
PC8	7	0.461
PC9	8	0.433
PC10	8	0.447
PC11	8	0.431

Appendix C. List of Acronyms

For a necessary reference to the text of this thesis, a list of acronyms is included (Table 2).

Table 2: The acronyms and their associated meanings are listed.

Acronym	Meaning
ACF	Autocorrelation Function
CEP	Central Equatorial Pacific
EEP	Eastern Equatorial Pacific
ENSO	El Niño - Southern Oscillation
EOF	Empirical Orthogonal Function
JMA	Japan Meteorological Agency
OI	Optimum Interpolation
PC	Principle Component
PDF	Probability Density Function
QBO	Quasi-Biennial Oscillation
SST	Sea Surface Temperature
WEP	Western Equatorial Pacific

References

Allan, R., J. Lindesay, and D. Parker, *El Niño Southern Oscillation and Climatic Variability*, CSIRO, 1996.

Barnett, T., A solar-ocean relation: fact or fiction?, *Geophysical Research Letters*, **16**, 803–806, 1990.

Bendat, J. S., and A. G. Piersol, *Random Data Analysis and Measurement Procedures*, Wiley-Interscience, 1986.

Elsner, J. B., and A. A. Tsonis, Nonlinear dynamics established in the ENSO, *Geophysical Research Letters*, **20**, (2), 213–216, 1993.

Enfield, D. B., and Luis Cid S., Low-frequency changes in El Niño - Southern Oscillation, *Journal of Climate*, **4**, 1137–1146, 1991.

Enfield, D. B., El Niño past and present, *Reviews of Geophysics*, **27**, (1), 159–187, 1989.

Gumbel, E. J., Extreme value analysis of hydrologic data, in *Proc. Hydr. Symp.*, no. 5, 1958.

Japan Meteorological Agency, Climate charts of sea surface temperatures of the western north pacific and the global ocean, Marine Department, Japan Meteorological Agency, Tokyo, Japan, 51 pp., 1991.

Lau, K.-M., Elements of stochastic-dynamical theory of the long-term variability of the el niño-southern oscillation, *Journal of Atmospheric Sciences*, **42**, (14),

Lau, K.-M., Elements of stochastic-dynamical theory of the long-term variability of the el niño-southern oscillation, *Journal of Atmospheric Sciences*, **42**, (14), 1552–1558, 1985.

Oceanic Interdecadal Climate Variability, IOC Technical Series 40, UNESCO, 1992.

Philander, S. G., *El Niño, La Niña, and the Southern Oscillation*, Academic Press, Inc., 1990.

Priestly, M., *Spectral Analysis and Time Series*, Academic Press, Inc., 1981.

Quinn, W. H., and V. T. Neal, El niño occurrences over the past four and a half centuries, *Journal of Geophysical Research*, **92**, (C13), 14449–14461, 1987.

Reynolds, R. W., and D. C. Marsico, An improved real-time global sea surface temperature analysis, *Journal of Climate*, **6**, 114–119, 1993.

Reynolds, R. W., and T. Smith, Improved global sea surface temperature analyses using optimum interpolation, *Journal of Climate*, **7**, (6), 1994.

Salby, M. L., *Fundamentals of Atmospheric Physics*, Academic Press, 1996.

Sittel, M. C., Marginal probabilities of the extremes of enso events for temperature and precipitation in the southeastern united states, technical report 94-1, Center for Ocean Atmospheric Prediction Studies (COAPS)-The Florida State University, 1994.

Slutz, R., S.J.Lubker, J. Hiscox, S. Woodruff, R. Jenne, D. Joseph, P. Steurer, and J.D.Elms, *COADS Comprehensive Ocean-Atmosphere Data Set Release 1*, CIRES University of Colorado, 1985.

1, CIRES University of Colorado, 1985.

Smith, T. M., R. W. Reynolds, R. E. Livezey, and D. C. Stokes, Reconstruction

of historical sea surface temperatures using empirical orthogonal functions, *Journal of Climate*, **9**, 1996.

Theiler, J. S., A. Longtin, B. Galdrikian, and J. Farmer, Testing for nonlinearity in time series: the method of surrogate data, *Physica D*, **58**, 77–94, 1992.

Thompson, J. D., and J. J. O'Brien, Time-dependent coastal upwelling, *Journal of Physical Oceanography*, **3**, (1), 33–46, 1973.

Wilks, D. S., *Statistical methods in the atmospheric sciences*, Academic Press, 1995.

Yarnal, B., Extratropical teleconnections with el niño-southern oscillation (enso) events, *Progr. Phys. Geog.*, **9**, 315–352, 1985.

BIOGRAPHICAL SKETCH

Degrees

B.S. 1995, The Florida State University

Experience

1992: Radio Meteorologist - WTAL

1992-93: Meteorological observational experience as FSU weather station employee

1994: Pseudo-stress analyst for Pacific and Indian Ocean wind data as MASIG employee

1994-95: Research assistant in MASIG, working on digital and log data comparisons for the METEOR research vessel

1995: Undergraduate thesis work: EOF analysis of TOGA-COARE Temperature and Dew Point data to fill spatial data gaps

1995-96: TOGA-COARE diurnal cycle atlas

1995-96: Modeling of Kelvin waves in Lake Michigan

1996-97: This thesis.

1996-97: This thesis.

Honors

Academic Scholarship - University of MN at Morris

AMS 75th Anniversary Campaign Scholarship

University Fellowship - CU

Memberships

American Meteorological Society (AMS)

Chi Epsilon Pi - Meteorology Honor Society

Personal information

Born May 5, 1972 in St. Cloud, Minnesota

Publications

1994 Comparison of Digitally Transmitted Surface Meteorology Data and
Cruise Report for DBMeteor: Cruise 18. Report WOCEMet 95-1.
COAPS.

1996 Mean Diurnal Patterns for Select Stations in the TOGA COARE
Region. August. Coare Report 96-2. (with Shawn Smith and David
Legler) COAPS.