Seasonality in an Empirically Derived Markov Model of Tropical Pacific Sea Surface Temperature Anomalies*

SCOT D. JOHNSON, DAVID S. BATTISTI, AND E. S. SARACHIK

Department of Atmospheric Sciences, University of Washington, Seattle, Washington

(Manuscript received 8 January 1999, in final form 3 December 1999)

ABSTRACT

An empirically derived linear dynamical model is constructed using the Comprehensive Ocean–Atmosphere Data Set enhanced sea surface temperature data in the tropical Pacific during the period 1956–95. Annual variation in the Markov model is sought using various tests. A comparison of Niño-3.4 forecast skill using a seasonally varying Markov model to forecast skill in which the seasonal transition matrices are applied during opposite times of the year from which they were derived is made. As a result, it is determined that the seasonal transition matrices are probably not interchangeable, indicating that the Markov model is not annually constant. Stochastic forcing, which has been hypothesized to exhibit seasonality, is therefore not the sole source of the annual variation of El Niño–Southern Oscillation (ENSO) dynamics and the phase locking of ENSO events to peak during November.

1. Introduction

It is well documented that even with the climatological cycle removed, tropical Pacific SST anomaly variance varies annually, with a minimum in boreal spring and a maximum in November and December—the latter being coincident with the time of peak warm and cold El Niño–Southern Oscillation (ENSO) events. If it is assumed that the unforced dynamics of ENSO SST anomalies (SSTAs) produce no growing eigenmodes (so the variance is supported by high-frequency forcing associated with weather, Penland and Sardeshmukh 1995, hereafter PS95), the question arises as to whether the seasonal cycle in the anomalies exists in the internal dynamics, in the external forcing, or in both. This question has been addressed by Penland (1996, hereafter P96), in which it is concluded that the internal dynamics have little if any seasonality and the annual variation of SSTA variance is therefore due primarily to seasonality in the forcing. A recent study involving an intermediate coupled model (ICM) of ENSO that includes the seasonal cycle may support this result (Fluegel and Chang 1998), since it was found that including the seasonal cycle in the model dynamics does not improve forecasts and has no effect on the spring predictability barrier. In contrast, other ICM studies suggest that significant seasonality should be present in the internal dynamics due to seasonality of the basic state (Chen et al. 1997; Thompson 1998; Battisti and Hirst 1989; Blumenthal 1991).

This article presents the results of a study to determine whether or not the internal dynamics of the ENSO system, as modeled by a Markov model of SST anomalies, possess an annual cycle. Section 2 describes how the Markov model is produced in cases where the dynamics are annually constant and annually varying. In section 3, the τ test of PS95 is discussed, and the ability of the τ test to detect seasonality in a variety of Markov models built from observed and model-generated ENSO SST anomalies is evaluated. In section 4, better tests for seasonality are explored, primarily one in which a seasonally varying Markov model is employed to forecast the Niño-3.4 index (SSTA spatially averaged over the region 5°N–5°S, 170°–120°W). The conclusions are summarized and discussed in section 5.

2. The Markov model

With the assumption that nonlinearities in the internal dynamics and the external forcing can be represented by stochastic forcing, the equation for the time evolution of the SSTA state given by $\mathbf{x}$ is

$$\frac{d\mathbf{x}}{dt} = \mathbf{B}\mathbf{x} + \boldsymbol{\varepsilon}, \tag{1}$$

where the internal dynamics are represented as $\mathbf{B}$ and

* Joint Institute for the Study of the Atmosphere and Ocean Contribution Number 477.

Corresponding author address: Scot Johnson, Max Planck Institut für Meteorologie, Bundesstr. 55, 20146 Hamburg, Germany.
E-mail: johnson@dkrz.de

$\boldsymbol{\varepsilon}$ represents stochastic external forcing. If no stochastic forcing were present, then given a known initial state, all future states would be determined. Therefore, the internal (unforced) portion of the dynamics $\mathbf{B}$ also represents the deterministic portion of the dynamics. The general solution for $\mathbf{x}(t + \tau)$ given $\mathbf{x}(t)$ is

$$\mathbf{x}(t + \tau) = \exp\left[\int_t^{t+\tau} \mathbf{B}(t'')\, dt''\right]\mathbf{x}(t) + \int_t^{t+\tau} \exp\left[\int_{t'}^{t+\tau} \mathbf{B}(t'')\, dt''\right]\boldsymbol{\varepsilon}(t')\, dt'. \quad (2)$$

The first term on the right represents the purely deterministic contribution to the time evolution of $\mathbf{x}(t)$. The second term represents the incorporation and subsequent dynamical evolution of the external (stochastic) forcing. For seasonally invariant internal dynamics, that is, when $\mathbf{B}$ is constant, the deterministic term simplifies to

$$\mathbf{x}(t + \tau) = \exp(\mathbf{B}\tau)\mathbf{x}(t), \quad (3)$$

where $\exp(\mathbf{B}\tau)$ is sometimes referred to as the Green function, ''$\mathbf{G}(\tau)$.'' Note that $\mathbf{G}(\tau)$ is a ''Markov model'' since prediction requires knowledge of the state of the system at only a single time, and is empirically derived from the SSTA dataset according to

$$\mathbf{G}(\tau) = \langle \mathbf{x}(t + \tau)\mathbf{x}(t)^{\mathrm{T}} \rangle \langle \mathbf{x}(t)\mathbf{x}(t)^{\mathrm{T}} \rangle^{-1}, \quad (4)$$

where $\langle\ \rangle$ denotes the ensemble average. Each month throughout the data segment provides a member of the ensemble. The eigenanalysis of $\mathbf{G}(\tau)$ and the application to geophysical problems was pioneered by Hasselmann (1988) and von Storch et al. (1988). Hasselmann coined the phrase ''principal oscillation pattern'' (POP) in reference to the right eigenvectors of $\mathbf{G}(\tau)$:

$$\mathbf{G}\mathbf{u} = \mathbf{u}\boldsymbol{\gamma}, \quad (5)$$

which are empirical eigenmodes of the physical system. In von Storch et al. (1995) an overview is given of POP analysis and its applications in the study of geophysical systems. The eigenmodes are generally complex and occur as complex conjugate pairs and are accompanied by (right) eigenvalues of $\mathbf{G}(\tau)$ that also are complex and occur as complex conjugate pairs. Empirical eigenmodes without a complex conjugate mode can exist, in which case the eigenmode is not oscillatory, but purely decaying. Assuming that the time series is stationary, all eigenmodes of $\mathbf{G}(\tau)$ as defined by (4) decay in time (PS95), that is, all of the modes are stable.

In section 3, the Markov model $\mathbf{G}(\tau)$ will be calculated from (4) only under the assumption that the internal dynamics are annually constant. Once $\mathbf{G}(\tau)$ is calculated, the dynamical operator $\mathbf{B}$ can be determined by means of a spectral decomposition of $\mathbf{G}(\tau)$ (Penland 1989) that proceeds as follows: first, the right and left eigenvectors of $\mathbf{G}(\tau)$ are determined:

$$\mathbf{G}\mathbf{u} = \mathbf{u}\boldsymbol{\gamma} \quad (5)$$

$$\mathbf{G}^{\mathrm{T}}\mathbf{v} = \mathbf{v}\boldsymbol{\gamma}. \quad (6)$$

The right and left eigenvectors are columns of the matrices $\mathbf{u}$ and $\mathbf{v}$, respectively, and $\boldsymbol{\gamma}$ is a diagonal matrix whose elements are the eigenvalues of $\mathbf{G}(\tau)$. The eigenvectors can be normalized such that $\mathbf{u}$ and $\mathbf{v}$ form a biorthonormal set:

$$\mathbf{u}\mathbf{v}^{\mathrm{T}} = \mathbf{u}^{\mathrm{T}}\mathbf{v} = \mathbf{I}, \quad (7)$$

where $\mathbf{I}$ is the identity matrix. (As a useful point to note, it follows that $\mathbf{v}\mathbf{u}^{\mathrm{T}} = \mathbf{v}^{\mathrm{T}}\mathbf{u} = \mathbf{I}$). Now $\mathbf{G}(\tau)$ can be written as

$$\mathbf{G}(\tau) = \mathbf{u}\boldsymbol{\gamma}\mathbf{v}^{\mathrm{T}}. \quad (8)$$

The columns of $\mathbf{u}$ and $\mathbf{v}$ are nonorthogonal right and left eigenvectors of $\mathbf{G}$ (5) and (6), which is in general nonnormal.

In the same manner $\mathbf{B}$ can be spectrally decomposed. The right and left eigenvectors of $\mathbf{G}(\tau)$ and $\mathbf{B}$ are the same:

$$\mathbf{B} = \mathbf{u}\boldsymbol{\beta}\mathbf{v}^{\mathrm{T}}, \quad (9)$$

where $\boldsymbol{\beta}$ is a diagonal matrix whose elements, the eigenvalues of $\mathbf{B}$, are the natural logarithm of the eigenvalues in $\boldsymbol{\gamma}$ divided by τ. To find the dynamical operator $\mathbf{B}$, the natural logarithm of the elements in $\boldsymbol{\gamma}$ are determined, divided by τ [i.e., $\boldsymbol{\beta} = \tau^{-1} \ln(\boldsymbol{\gamma})$], and substituted back into (9). The spectral decomposition of $\mathbf{B}$ and $\mathbf{G}$ will be employed as part of a test for seasonality in $\mathbf{B}$.

Since we are seeking annual variation in $\mathbf{B}$ (and therefore $\mathbf{G}$ also), It will not always be assumed that $\mathbf{B}$ is constant, but rather that $\mathbf{B}$ (and $\mathbf{G}$) may follow an annual cycle. The treatment of a geophysical case where the statistics of the system are assumed to be periodic was approached by Hasselmann and Barnett (1981) who applied a ''phase-averaged'' [or ''phase smoothed,'' OrtizBevia (1997)] technique to the statistical prediction of ENSO using SST and the trade wind field. The motivation behind using the phase-averaged technique is to limit the division of the empirical data (when sample sizes are reduced, significance of the results is reduced as well). The dataset can be divided into as few as three parts if $\mathbf{B}$ is modeled as

$$\mathbf{B}(t) = \mathbf{B}_0 + \mathbf{B}_1 \cos(\omega t) + \mathbf{B}_2 \sin(\omega t). \quad (10)$$

Determination of the seasonal dependence of $\mathbf{B}$ is not possible using the phase-averaged approach so a ''fixed-phase'' (cyclostationary) method is used instead, where season-specific versions of $\mathbf{G}$ are derived using (4) after the dataset is divided into seasonal ensembles for the calculation of $\mathbf{G}$ (Hasselmann and Barnett 1981). In section 4, forecasts will be made using seasonal and nonseasonal versions of (4). The seasonal $\mathbf{G}$ is defined in a very similar manner to the nonseasonal $\mathbf{G}$ because $\mathbf{B}$ is assumed to be constant during that portion of the year for which $\mathbf{G}$ is derived. In the first of two seasonal

constructions to be used, the dataset will be divided into 12 bins and $\mathbf{G}_i(3)$, the 3-month propagator starting from month i, will be calculated as

$$\mathbf{G}_i(3) = \langle \mathbf{x}(i + 12\nu + 3)\mathbf{x}(i + 12\nu)^{\mathrm{T}} \rangle \times \langle \mathbf{x}(i + 12\nu)\mathbf{x}(i + 12\nu)^{\mathrm{T}} \rangle^{-1}, \quad (11)$$

where $i = 1, 2, \ldots, 12$ and ν is any nonnegative integer. (Note that the ensemble in this case includes all values of ν, which ranges from 0 to $N - 1$ for N years of data). For example, $\mathbf{G}_1(3)$ and $\mathbf{G}_2(3)$ are the propagators from January to April and February to May, respectively. As a further example, the transition matrix from one March to the next is

$$\mathbf{G}_3(12) = \mathbf{G}_{12}(3)\mathbf{G}_9(3)\mathbf{G}_6(3)\mathbf{G}_3(3). \quad (12)$$

In the second of the two seasonal constructions to be used, the dataset will be divided into only four bins (so that the training samples are three times larger than in the 12-bin case) and $\mathbf{G}_i(3)$ will be calculated according to

$$\mathbf{G}_i(3) = \langle \mathbf{x}(i' + 12\nu + 3)\mathbf{x}(i' + 12\nu)^{\mathrm{T}} \rangle \times \langle \mathbf{x}(i' + 12\nu)\mathbf{x}(i' + 12\nu)^{\mathrm{T}} \rangle^{-1}, \quad (13)$$

where $i = 1, 2, 3, 4$ denotes the season, $i' = \{1\ 2\ 3\} + 3x(i - 1)$ denotes the month, and ν is a nonnegative integer. The ensemble includes all values of ν, which ranges from 0 to $N - 1$ for N years of data, plus the 3 values of i' for each given season (denoted by i). For example, $\mathbf{G}_1(3)$ and $\mathbf{G}_2(3)$ are the propagators from January to April (which is the same as from February to May and from March to June), and from April to July (same as from May to August and from June to September), respectively. The transition matrix from one April to the next is

$$\mathbf{G}_2(12) = \mathbf{G}_1(3)\mathbf{G}_4(3)\mathbf{G}_3(3)\mathbf{G}_2(3). \quad (14)$$

Forecasts using season-specific propagators applied during the appropriate time of year will be compared to forecasts when the propagators are applied during the opposite (supposedly the incorrect) time of year to determine whether or not the propagators truly are season specific or if applying the propagators at different times of the year alters the forecast skill.

3. The τ test

Penland (1996) has suggested that the deterministic dynamics contain little or no annual variation. This conclusion is based on the results of the "τ test," as it was called by PS95, which is one of several tests they conducted to determine the validity of fitting SSTA data to a Markov model. To perform the τ test, $\mathbf{G}(\tau)$ is first calculated directly from monthly mean SSTA using (4) at values of τ ranging from 1 to 15 months. Next, spectral decomposition is employed to derive $\mathbf{B}$ from each of the 15 versions of $\mathbf{G}$. If the linear model is a good fit to the data, and if there is no annual variation in the internal dynamics, then (3) applies and $\mathbf{B}$ has no dependence on τ, and should be roughly constant for all values of τ. PS95 presented (in their Fig. 12) the Euclidean norm of $\mathbf{B}$ versus τ of a Markov model of Pacific SST anomalies. The Euclidean norm, which is the square root of the sum of the squares of the matrix elements, is convenient since it is a scalar measure of $\mathbf{B}$. In general, the norm of $\mathbf{B}$ versus τ will fluctuate if there is variation in the matrices $\mathbf{B}$. Note that the τ test is not a rigorous statistical test in the same sense as, for example, the χ^2 test (Spiegel 1988). The τ test involves judgment of how flat the profile of norm($\mathbf{B}$) is versus lead time τ. PS95 observed that the norm of $\mathbf{B}$ for values of τ up to about nine months is approximately constant, and consequently concluded that the matrices $\mathbf{B}$ are approximately constant versus prediction lead time τ.

As previously mentioned, one potential reason for this τ test to fail is annual variation in the internal dynamics where none is assumed to be present. P96 noted that if the internal dynamics vary sinusoidally (characteristic of an annual cycle), then the empirically derived $\mathbf{B}$ will vary with τ. Another cause for failure of the τ test is the appearance of an eigenmode whose period of oscillation is nearly twice the value of τ. When such a mode appears, the spectral decomposition may yield left and right eigenvectors that tend to be larger in magnitude than when such a mode is not present. This, in turn, can induce a fluctuation in the norm of $\mathbf{B}$ from one value of τ to the next and thereby cause failure of the τ test. Since the eigenmodes usually have periods longer than about 20 months, this situation is generally not expected to arise for values of τ less than about 10 months. In practice, however, such modes appear frequently, although the reason for the appearances is not clear.

We repeat this τ test using SST data from the Comprehensive Ocean–Atmosphere Data Set (COADS) "enhanced" set that spans the years 1956–95. The data are processed in a manner similar to PS95. SST data of the Pacific basin (30°S–30°N, 120°E–70°W) are consolidated onto a $4° \times 10°$ latitude–longitude grid. The data are subjected to a 3-month running mean smoother and the climatology is removed, where climatology is determined at each grid point as the average of the data at that grid point over the period 1956–95 for each of the 12 calendar months. We derived the nonseasonal Markov model by applying (4), and then decomposed the Markov model $\mathbf{G}(\tau)$ to determine $\mathbf{B}(\tau)$.

Figure 1 shows the Euclidean norm of $\mathbf{B}$ versus τ for three nonseasonal Markov models constructed using different numbers of EOFs. The solid (dashed) [dot–dashed] line corresponds to a Markov model constructed using 10 (11) [15] EOFs, which account for 62.0% (63.5%) [68.3%] of the total variance. The symbol "π" is plotted at each value of τ in the three curves where one or more eigenmodes are present whose period is approximately equal to 2τ. The appearance of the high-frequency mode causes the norm of $\mathbf{B}$ to fluctuate, and

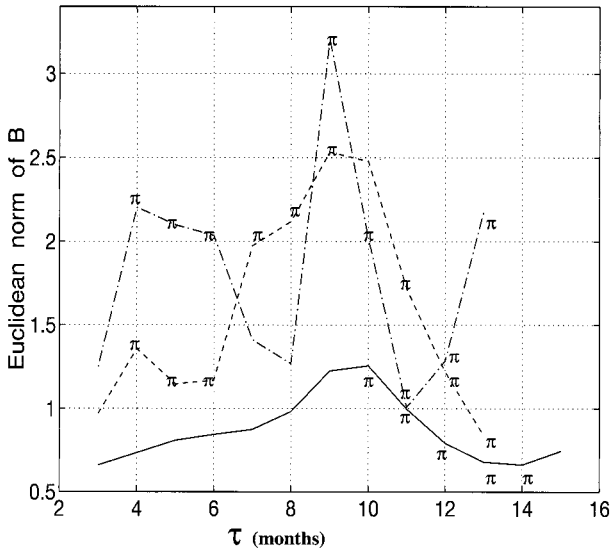

FIG. 1. Euclidean norm of **B** vs regression lag time τ for Markov model empirically derived from (4) using Pacific (32°N–32°S, 100°E–70°W) SST anomalies from Jan 1956 through Dec 1995. Markov models are built from 15 EOFs (68.3% of variance retained, dot–dashed line), 11 EOFs (63.5% of variance retained, dashed line), and 10 EOFs (62.0% of variance retained, solid line). The symbol π is used to label the lead times for which eigenmodes with period $2^*\tau$ erroneously appear in the Markov models.

this is apparent in all three profiles of norm(**B**). It is found, however, that at all points where the π mode appears in Fig. 1, the mode is spurious (i.e., the mode is not a consistent element of the modeled system). This is determined to be the case by examining the modes at values of τ where no π mode is present, and extrapolating the periods of the modes to find the periods that are to be expected at other values of τ. At all points where the π mode appears, a mode with this period is not expected. Unfortunately, there appears to be little that can be done to suppress the appearance of the π modes. One remedy is to reduce the number of principal components used in making the Markov model as much as possible (as demonstrated by the solid curve in Fig. 1). Reducing the dimensionality of the model reduces the number of eigenmodes and hence reduces the chances of an appearance by a π mode.

It can be seen in Fig. 1 that where π is not present, in particular in the profile of the 10-EOF Markov model, norm(**B**) is fairly constant. The flatness of norm(**B**) in this curve is comparable to that of Fig. 12 of PS95. It is this flatness that is the basis of the conclusion of P96 that the dynamics contain negligible seasonality. The first π mode in the norm(**B**) profile of the 10-EOF Markov model appears at $\tau = 10$, and it appears that this π mode induces a fluctuation in norm(**B**) at $\tau = 9$ months (and even at $\tau = 8$ months) as well. Assuming that the profile of norm(**B**) is not influenced by a π mode for τ less than 8 months, we note that the variation in the norm(**B**) from $\tau = 3$ to $\tau = 7$ is 32.4%. Perhaps, since there are no π modes in this range, then this variation may be caused by annual variation in **B**.

For comparison, two ENSO Markov models of intermediate complexity were tested for uniformity of norm(**B**). The first Markov model was derived using output from a linear ocean–atmosphere model (LOAM) designed as a linearized version of the Battisti model (Thompson 1998; Battisti 1988) which, in turn, is a variant of the Zebiak and Cane (ZC) model (Zebiak and Cane 1987). The second Markov model is derived using output of the Zebiak and Cane (1987) ICM of ENSO. Both models contain a seasonally varying basic state, and hence both models possess seasonally varying internal dynamics (seasonality in both of these models is documented in Thompson and Battisti 2000; Zebiak 1984; and Blumenthal 1991). The Markov model analysis proceeds in the same way for both cases: each ICM is run for a period of about 100 years and the SST anomaly data are collected, an EOF analysis is performed, and a seasonally invariant Markov model is derived using (4) from a selected number of the top principal component time series. The Markov model $\mathbf{G}(\tau)$ is then decomposed into internal dynamics matrices $\mathbf{B}(\tau)$ and the Euclidean norm of **B** versus τ is examined.

Norm(**B**) versus τ for LOAM is shown in the top of Fig. 2. The solid (dashed) line corresponds to a Markov model derived using 2 (4) retained EOFs, accounting for 97.6% (99.4%) of the total variance. Even with just two retained EOFs, a fairly skillful Markov model can be derived. This is due to the fact that LOAM has little activity besides the ENSO mode, and the first two EOFs correspond to two phases of the ENSO mode 90° apart, which is all that is required to make a Markov model of a single cyclic mode. In the 2-EOF case, norm(**B**) changes by only 3.5% between $\tau = 1$ and $\tau = 20$. [It is apparent that beginning at $\tau = 20$, norm(**B**) climbs due to the approach of the point at which half the mode period is reached, which occurs at $\tau = 22$ months]. When higher numbers of EOFs are retained, a high-frequency π mode quickly appears, as shown by the dashed line corresponding to four retained EOFs. The π mode is spurious, however. No real mode exists in LOAM with a period so short.

Norm(**B**) versus τ for the Markov model derived from output of the ZC model is shown in Fig. 2b. The solid (dashed) line corresponds to three (five) retained EOFs, accounting for 89.3% (93.8%) of the total SSTA variance. As with LOAM, a reasonably good Markov model can be derived using very few EOFs of the ICM output. When three EOFs are retained, only one oscillating mode can exist since oscillating modes always occur as complex conjugate pairs (and these two eigenmodes therefore refer to the same empirical mode of the system). This oscillating mode has a period of about four years and is identified as the ENSO mode. The third eigenmode is purely decaying. When five EOFs are retained, two oscillating modes and one purely decaying

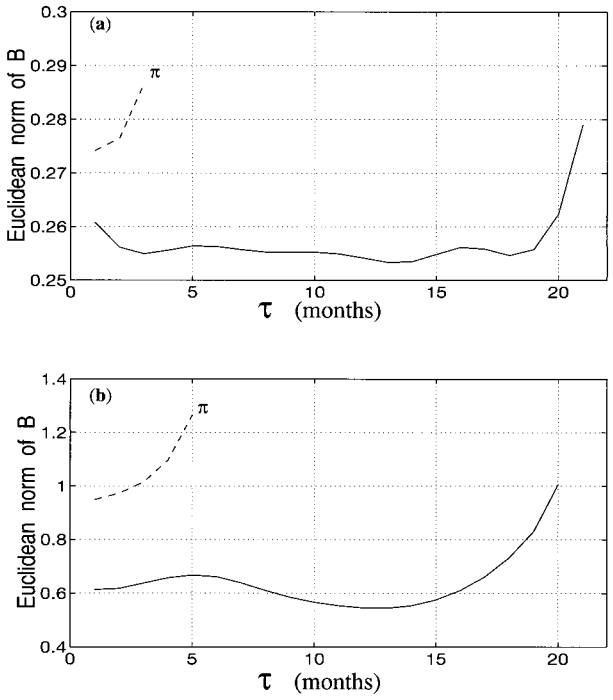

FIG. 2. Euclidean norm of **B** vs regression lag time τ for Markov models derived from (4) using model output of two ENSO models of intermediate complexity, both having annually varying background. (a) EDMs are built using first 4 SST EOFs of LOAM (99.4% variance retained, dashed line) and first 2 SST EOFs of LOAM (97.6% variance retained, solid line). (b) EDMs are built using the first 5 SST EOFs of the Zebiak–Cane (ZC) model (93.8% variance retained, dashed line) and first 3 SST EOFs of the ZC model (89.3% variance retained, solid line).

mode occur. In this case, one oscillating mode is identified as the ENSO mode. The second oscillating mode has a period of about 10–12 months and appears to be consistent with the ''mobile mode'' identified by Mantua and Battisti (1995). Note that in this case the π mode that appears at $\tau = 5$ months may not be spurious but real to the modeled system. Either way, the high-frequency mode causes norm(**B**) to fluctuate, precluding the ability to examine norm(**B**) for seasonality in the dynamics when the Markov models of ZC output are larger than 3×3. Norm(**B**) for the 3-EOF Markov model varies by 22.5% between $\tau = 1$ and $\tau = 17$ months.

Over the usable range of the profiles of norm(**B**) versus τ, the norm(**B**) for the EDM of LOAM varies by 3.5%, the norm(**B**) for the EDM of the ZC model varies by 22.5%, and for the EDM of the COADS SSTA data the norm(**B**) varies by 32.4%. The dynamics of LOAM and ZC are seasonally varying by construction, and since the norm(**B**) for the Markov model derived from COADS data varies by a greater amount, it is plausible that significant seasonality also exists in the internal dynamics of the Markov model fitted to observational data. It is therefore difficult to conclude based upon the profiles of norm(**B**) using Markov models of observational data that there is no seasonality in the internal dynamics (as PS95 did). We suggest, based on the above examples and other studies we have made, that the τ test is not well suited to the task of seeking seasonality, and other methods of testing for seasonal variation in the Markov model should be sought.

4. Seasonality in Markov models of equatorial Pacific SST

Three further tests were performed on Markov models to detect seasonality. The first test was to directly measure the rms difference between corresponding elements of 12 (seasonally varying) Markov models from one calendar month to the next. One-month 10×10 Markov models were produced from the first 10 principal components of COADS SST data and the standard deviation of a given element over the 12 months was calculated. Then, the dataset was split into four sections of 10 years each. The standard deviation of the set of four January–February transitions was compared to the standard deviation between different calendar months previously calculated, and this was repeated for the February–March transitions, etc. Although it was found that the difference between different months was usually greater than the difference between the same month built from different decades of the data, the results were not statistically significant. The number of years of SST data appears to be too small to derive a consistent seasonally varying Markov model.

A second test was employed to find seasonality, which compares the skill of prediction of nonseasonal Markov models to that from the seasonal Markov models. If the internal dynamics vary seasonally, then a seasonal Markov construction should make better forecasts than a nonseasonal Markov model. We devised two different seasonal constructions whose predictive skill of the Niño-3.4 index would be compared to that of a nonseasonal construction. The nonseasonal construction was simply a calculation of the $\tau = 3$-month Markov model from the 40-yr length of data using (4) assuming no seasonality. To make predictions, the 40-yr dataset was split into 38-yr Markov model training periods and 2-yr verification periods. The Niño-3.4 index during the 2-yr period was predicted using a Markov model trained by the other 38 years. This process was repeated 20 times to make forecasts during all 20 2-yr periods in the 40-yr dataset. Since the spans of data to be predicted were kept separate from data used for training, the predictions are said to be ''cross validated'' and might be called ''forecasts.'' If cross validation is not used, the predictions would demonstrate artificial skill that would exceed the true skill (Davis 1976). It is hoped that the cross-validated method we employ minimizes artificial prediction skill.

In the first seasonal construction, the data was split into 12 bins as explained in section 2: one bin contained all January–April transitions ($\tau = 3$ months), the second

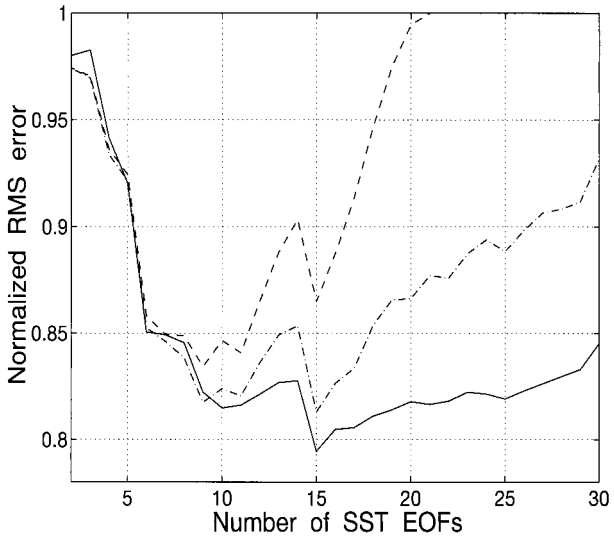

FIG. 3. The rms error of prediction of the Niño-3.4 index over the period 1956–95 vs number of retained EOFs of SST anomaly. The Markov prediction models are derived using (4) with $\tau = 3$ months using COADS (enhanced set) SST. In all cases in this study using Markov models of COADS data, a cross-validating procedure is used so that training periods and verification periods of the Markov models never overlap. The rms error is averaged over forecast lead times of 3, 6, 9, and 12 months, and is normalized such that prediction using the climatology would have rms error equal to one. Solid line; the Markov model is nonseasonal. Dot–dashed line; the Markov model is seasonal such that 4 models are built for transitions beginning in each of four seasons. Dashed line; the Markov model is seasonal such that 12 models are built for transitions beginning in each of the calendar months.

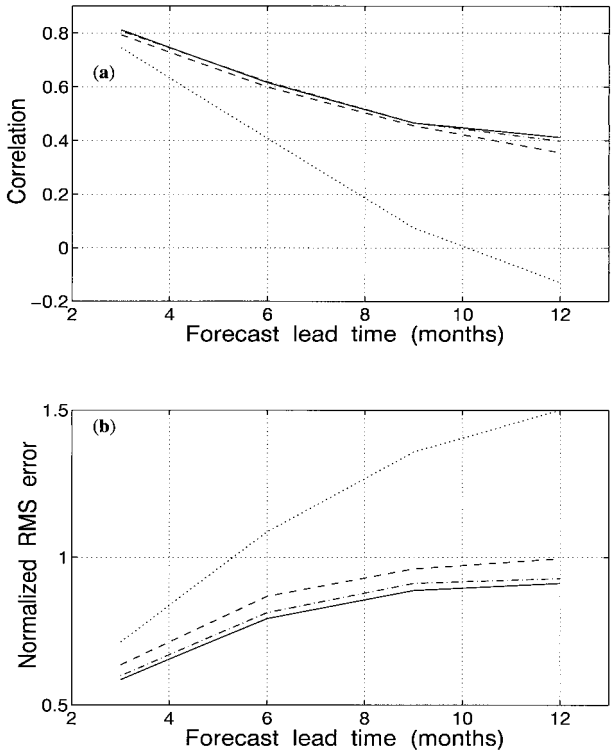

FIG. 4. (a) Correlation between the forecasted and observed Niño-3.4 time series from 1956 to 1995 vs lead time of predictions. Forecasts are made with lead times of 3, 6, 9, and 12 months using a single nonseasonal Markov model (solid line), using 4 separate Markov models applied during each of the four seasons (dot–dashed line), and using 12 separate Markov models applied during each of the 12 calendar months (dashed line). Forecasting by persistence is shown by the dotted line. (b) Same as in (a), for rms error of the forecasts normalized such that rms error of forecasting using climatology equals one at all lead times.

bin contained all the February–May transitions ($\tau = 3$ months), and so on. Twelve different 3-month propagators were thus calculated, 1 for each of the 12 bins. Each bin contained 38 sample transitions (one sample for each of the 38 yr of data used for training). The January–April Markov model would then be used to predict the January–April transitions during the 20 2-yr verification periods, and so on for the other 11 Markov models. A sample size of 38 transitions is a rather small set, so an additional seasonal construction was devised that would increase the training sample size: the data was divided into only four bins. The first bin contained all January–April transitions plus all February–May transitions plus all March–June transitions. The second bin contained all April–July transitions plus all May–August transitions plus all June–September transitions, and so on. The Markov model trained by the ($3 \times 38 =$) 114 transitions in the first bin would be used to make all 3-month predictions beginning in January, February, and March during the cross-validated verification periods, and similarly for the other three Markov models. Predictions were made using the $\tau =$ 3-month propagator, applied during the appropriate season, for lead times of 3, 6, 9, and 12 months.

Figure 3 shows the rms error of the Niño-3.4 forecasts averaged over lead times of 3, 6, 9, and 12 months versus the number of SST anomaly EOFs used in the construction of the Markov models. The solid line corresponds to the nonseasonal forecasts, the dot–dashed line corresponds to the 4-bin forecasts, and the dashed line corresponds to the 12-bin forecasts. The rms error is normalized such that forecasts using climatology have an rms error always equal to 1. Persistence forecasts (averaged over lead times of 3, 6, 9, and 12 months) generally have an rms error greater than 1 (not shown). In both seasonal and nonseasonal cases, about nine EOFs are required in the Markov models to maximize skill and as more EOFs are retained, error grows. The lowest error, just under 0.8, is had by the 15-EOF nonseasonal model. It is apparent that including seasonality in the Markov model does not lead to improvement in forecasts.

Figures 4a and 4b reflect this in further detail. The correlation between true and predicted Niño-3.4 indices and the rms error of Niño-3.4 forecasts are shown, respectively, versus lead time. Solid, dot–dashed, and dashed lines correspond to the nonseasonal, 4-bin seasonal, and 12-bin seasonal Markov models, respectively.

The dotted line corresponds to forecasts by persistence. The profiles shown correspond to Markov models built using the number of SST EOFs that minimizes error, which is 15 EOFs for the nonseasonal and 4-bin seasonal constructions, and 9 EOFs for the 12-bin seasonal construction (Fig. 3 shows that the error is minimized at these numbers of EOFs). As was apparent in Fig. 3, it is clear that the seasonal construction does not provide a better forecast than the nonseasonal construction. In comparison to the persistence forecasts, the nonseasonal versus seasonal skills appear not to differ significantly. Two possible explanations of this result are that (i) there is no seasonality in the internal dynamics and therefore predictive skill is the same for seasonal and nonseasonal constructions, and (ii) seasonality is present in the internal dynamics, and this would lead to better forecast skill by the seasonal models if not for the fact that the reduction of the training sample size (which is necessary to construct a seasonal Markov model) reduces skill and cancels the benefit of seasonality. The effect of this reduction of the training sample is visible in Fig. 3 as the steeper increase in error as more EOFs are included for the seasonal models. It appears that the 4-model seasonal case in particular may have achieved better skill than the nonseasonal model if not for this effect. It is necessary to devise a test for seasonality where the reduction of the training sample is not a factor, as is the case in the next method.

A third approach to determine whether or not the internal dynamics possess seasonal variation is to compare the forecast skill of the seasonal model as previously applied with the forecast skill of the seasonal model in a case in which the propagators are applied during the opposite calendar month for which they were constructed to be applied. For example, in the four-bin seasonal construction, the 3-month propagator derived and applied to forecast start months of January, February, and March, is applied in the second case during July, August, and September, and similarly for other 3-month propagators throughout the year. If the internal dynamics do not vary seasonally, then there will be no significant difference in forecast skill between the two cases. In other words, Markov models for different seasons would be interchangeable. On the other hand, if the dynamics do possess an annual variation, then Markov models for given seasons should not be able to be applied during the opposite season without showing diminished forecast skill.

Figures 5a and 5b show rms forecast error (averaged over lead times of 3, 6, 9, and 12 months) of the 12-bin and 4-bin seasonal Markov models versus the number of SST anomaly EOFs used in construction of the Markov models. In Fig. 5a (Fig. 5b), the dashed line (dot–dashed line) is the rms error of the correctly applied 12-bin (4-bin) model [same as the dashed line (dot–dashed line) in Fig. 3]. The dotted line in Fig. 5a (Fig. 5b) is the rms error of the 12-bin (4-bin) seasonal model applied during the opposite calendar month (season).

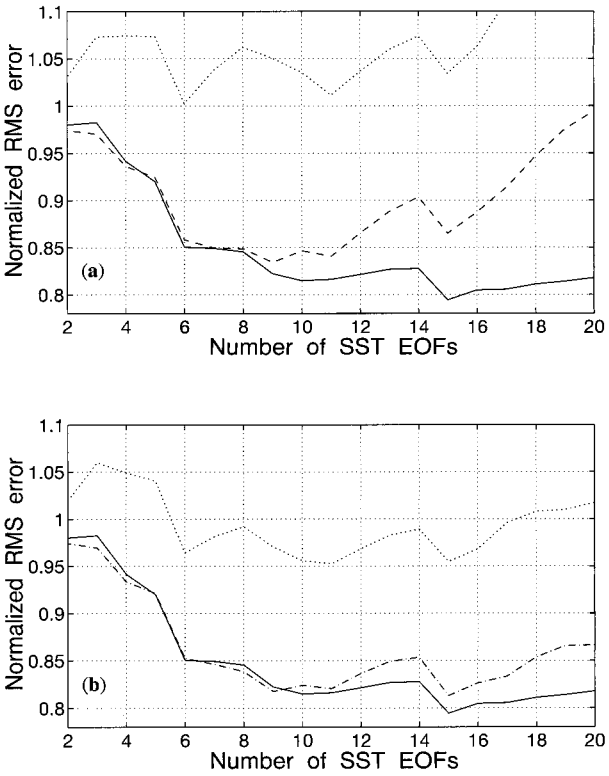

FIG. 5. Normalized rms error of the Niño-3.4 index forecasts vs number of leading EOFs of SST used in making the Markov models. (a) Twelve separate Markov models applied during each of the 12 calendar months (dashed line, same as in Fig. 4). The dotted line corresponds to the error when the 12 propagators are applied during the opposite time of year from which they were intended to be applied. For comparison, the error using nonseasonal models is shown (solid line, same as solid line in Fig. 3). (b) Four separate Markov models are applied during each of the four seasons (dot–dashed line, same as in Fig. 3). The dotted line corresponds to the error when the four propagators are applied during the opposite time of year from which they were intended to be applied. The error using nonseasonal models is also shown (solid line, same as solid line in Fig. 3).

The solid line in Figs. 5a and 5b is the error of the nonseasonal model (same as the solid line in Fig. 3), shown here for the purpose of comparison. It is clear for both seasonal constructions that applying the Markov models at opposite times of year results in substantially larger forecast error than when the Markov models are applied during the time of year that they were trained for.

A test is made of the statistical significance of the difference between forecast errors of the rightly and wrongly applied seasonal models in Fig. 5. It was noted that forecasts of Niño-3.4 during the 40-yr (480 month) time series yield a 480-member set of forecast errors that are approximately normally distributed about zero. If one method of forecasting is significantly worse than another method, then the standard deviation of the set of forecast errors of the worse method will be larger than the standard deviation of the better set. The ''f statistic'' is used to measure the significance of the dif-

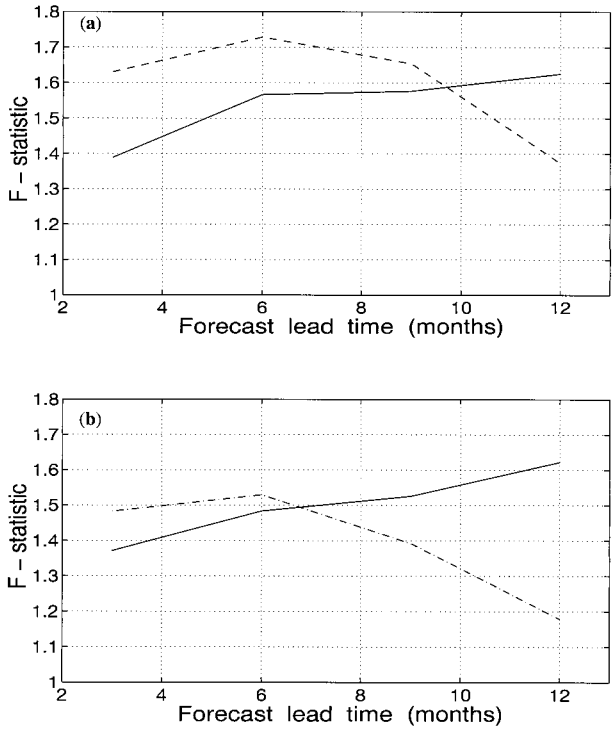

FIG. 6. The f statistic vs forecast lead time τ that measures the statistical significance of the difference between the rms forecast errors of the seasonal Markov models applied during the correct and opposite times of year. The 90% significance level (solid line) is also shown. (a) and (b) present the f statistic vs lead time for the case of 12 and 4 Markov models, respectively.

ference between the standard deviations of two sets (Spiegel 1988). The dashed line of Fig. 6a and the dot–dashed line of Fig. 6b show the f statistic versus lead time for the 12-bin and 4-bin seasonal cases, respectively. It is apparent in both cases that the difference in prediction skill (as measured by the f statistic) climbs to a peak at a lead time of about 6 months and then decreases at longer lead times. This behavior is to be expected: at short lead time the rms error of any of the different prediction schemes is smaller, and as a result, the degree of certainty that the prediction schemes have different skill is reduced. In addition, as lead time grows, all predictions worsen and tend to approach the level of skill of prediction using climatology, so the distinction between different schemes disappears at long lead time as well.

The solid lines in Fig. 6 are the levels above which the f statistic is significant at the 90% level. The number of independent elements (degrees of freedom) of the sets of forecast errors is calculated using the commonly applied technique by Leith (1973). The lag–one month autocorrelation of the 480-member time series of forecast errors is used to estimate the effective number of degrees of freedom, which ranges from about 30 to 60. It is hoped that this technique also corrects for the fact that prior to the analyses, the monthly anomalies were smoothed using a 3-month running mean, which will cause predictions from 1 month to the next to have some ''artificial'' similarity.

Since the lag–one month autocorrelation of the forecast time series monotonically increases with forecast lead time, the number of degrees of freedom of the error samplings correspondingly decreases, and consequently the level for which the f statistic is 90% significant increases as lead time increases. Defining the 90% significance levels was inexact, and at steps during the calculation where approximation was necessary, the most conservative possibility was used, so the 90% level in as high as possible in Fig. 6. We conclude that the difference in forecast skill of the rightly and wrongly applied seasonal model is fairly significant, indicating that the internal dynamics of the ENSO system indeed possess an annual cycle.

It can be seen in Fig. 6 that the f statistic better exceeds the level of 90% significance in the case of the monthly models (Fig. 6a) than do the seasonally averaged models (Fig. 6b). This is due to the fact that a monthly Markov model is separated by a gap of 5 months from the opposite month of the year, while a seasonally averaged Markov model is separated from the opposite season by a gap of only 3 months. The seasonally averaged models are therefore less distinguished from the models of the opposite time of year. Similar experiments using two biannual Markov models (to represent the warm and cold seasons) show a difference in forecast skill between correct and opposite times of year that is, correspondingly, less significant than the profiles of Fig. 6 (not shown).

5. Summary and discussion

The τ test of Penland and Sardeshmukh (1995), designed to test the ''goodness of fit'' of a linear, stable dynamical framework to equatorial Pacific SST anomalies, is examined as a test to find seasonality in the internal dynamics. It is suggested that the τ test is limited as a tool for determining whether or not the deterministic (internal) dynamics possess an annual cycle.

The skill of prediction of the Niño-3.4 index by Markov models with and without annual variation is employed to provide a more direct means of testing for seasonality. It is found that forecasting skill of two different seasonal Markov models is significantly diminished when the Markov model built for one season is applied during the opposite time of the year. It is therefore concluded that seasonality is present in the internal dynamics.

The finding that the internal dynamics of ENSO has a seasonal cycle agrees with the analyses of several intermediate coupled atmosphere–ocean models (for example, the Zebiak and Cane model and its variants), but it is opposite to the findings of Penland (1996), who concluded on the basis of the τ test that little or no annual variation is present in the internal dynamics. As

we have shown, however, the τ test is an indirect and unreliable test for seasonality and a more direct test using prediction skill suggests otherwise. More recently, Fluegel and Chang (1998) have used an intermediate coupled model of ENSO which includes the seasonal cycle to find that whether or not seasonality is included in the internal dynamics, the spring predictability barrier remains. On this basis they suggest that the spring predictability barrier results from seasonal information contained in the initial conditions rather than from the seasonally varying background, which suggests that seasonal variation in the internal dynamics is not a factor in the prediction skill of their model. This is a curious result that we can not explain, since we find that the seasonal cycle has a clear and strong impact on the predictive skill of our statistical model.

It may be, however, that seasonal variation in the internal dynamics of ENSO anomalies, even if it is certain to exist, does not have as strong an impact on ENSO predictability as initial conditions. Furthermore, in the view that ENSO is stochastically driven, it may also be true that weather noise contains a nontrivial seasonality for establishing the timing of ENSO event peaks (which occur around November). Important questions remain regarding the nature of these various factors and their relative importance to prediction; however, this study suggests that a true representation of the internal dynamics guiding the evolution ENSO anomalies will include a seasonal cycle.

Acknowledgments. This work was supported by a grant from the NOAA/Office of Global Programs to the Stanley P. Hayes Center of the University of Washington. We would like to thank Joe Barsugli, Cecile Penland, and Prashant Sardeshmukh for many gracious and enjoyable conversations and for their assistance. This publication is funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement NA67RJ0155. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies.

REFERENCES

Battisti, D. S., 1988: Dynamics and thermodynamics of a warming event in a coupled tropical atmosphere–ocean model. *J. Atmos. Sci.,* **45,** 2889–2919.

——, and A. C. Hirst, 1989: Interannual variability in a tropical atmosphere–ocean model: Influence of the basic state, ocean geometry, and nonlinearity. *J. Atmos. Sci.,* **46,** 1687–1712.

Blumenthal, M. B., 1991: Predictability of a coupled ocean–atmosphere model. *J. Climate,* **4,** 766–784.

Chen, Y.-Q., D. S. Battisti, T. N. Palmer, J. Barsugli, and E. S. Sarachik, 1997: A Study of the predictability of tropical Pacific SST in a coupled atmosphere–ocean model using singular vector analysis: The role of the annual cycle and the ENSO cycle. *Mon. Wea. Rev.,* **125,** 831–845.

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

Fluegel, M., and P. Chang, 1998: Does the predictability of ENSO depend on the seasonal cycle? *J. Atmos. Sci.,* **55,** 3230–3243.

Hasselmann, K., 1988: PIPs and POPs—The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res.,* **93,** 11 015–11 021.

——, and T. P. Barnett, 1981: Techniques of linear prediction for systems with periodic statistics. *J. Atmos. Sci.,* **38,** 2275–2283.

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

Mantua, N. J., and D. S. Battisti, 1995: Aperiodic variability in the Zebiak–Cane coupled ocean–atmosphere model: Air–sea interactions in the western equatorial Pacific. *J. Climate,* **8,** 2897–2927.

OrtizBevia, M. J., 1997: Estimation of the cyclostationary dependence in geophysical data fields. *J. Geophys. Res.,* **102,** 13 473–13 486.

Penland, C., 1989: Random forcing and forecasting using principal oscillation pattern analysis. *Mon. Wea. Rev.,* **117,** 2165–2185.

——, 1996: A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D,* **98,** 534–558.

——, and P. D. Sardeshmukh, 1995: The optimal growth of sea surface temperature anomalies. *J. Climate,* **8,** 1999–2024.

Spiegel, M. R., 1988: *Theory and Problems of Statistics.* McGraw-Hill, 504 pp.

Thompson, C. J., 1998: Initial conditions for optimal growth in a coupled ocean–atmosphere model of ENSO. *J. Atmos. Sci.,* **55,** 537–557.

——, and D. S. Battisti, 2000: A linear stochastic dynamical model of ENSO. Part I: Model Development. *J. Atmos. Sci.,* **13,** 2818–2832.

von Storch, H., T. Bruns, I. Fischer-Bruns, and K. Hasselmann, 1988: Principal oscillation pattern analysis of the 30- to 60-day oscillation in general circulation model equatorial troposphere. *J. Geophys. Res.,* **93,** 11 022–11 036.

——, G. Buerger, R. Schnur, and J.-S. von Storch, 1995: Principal oscillation patterns: A review. *J. Climate,* **8,** 377–400.

Zebiak, S. E., 1984: Tropical atmosphere–ocean interaction and the El Niño/Southern Oscillation phenomenon. Ph.D. thesis, MIT, 261 pp. [Available from MIT, 77 Massachusetts Avenue, Cambridge, MA 02139-4307.]

——, and M. A. Cane, 1987: A model El Niño–Southern Oscillation. *Mon. Wea. Rev.,* **115,** 2262–2278.