Empirically Derived Markov Models and Prediction 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

ABSTRACT

Empirical dynamical modeling (EDM) is employed to determine if ENSO forecasting skill using monthly mean SST data can be enhanced by including subsurface temperature anomaly data. The Niño 3.4 index is forecast first using an EDM constructed from the principal component time series corresponding to EOFs of SST anomaly maps of the central and eastern tropical Pacific (32°N–32°S, 120°E–70°W) for the period 1965–93. Cross validation is applied to minimize the artificial skill of the forecasts, which are made over the same 29-yr period. The forecasting is then repeated with the inclusion of principal components of heat content of the upper 300 m over the northern tropical Pacific (30°N–0°, 120°E–72°W).

The forecast skill using SST alone and SST plus subsurface temperature is compared for lead times ranging between 3 and 12 months. The EDM, which includes the subsurface information, forecasts with greater skill at all lead times; particularly important is the second principal component of the heat content, which appears to contribute information on the transition phase between warm and cold ENSO events. The apparent improvement by including subsurface information, although robust, does not appear to be statistically significant. However, the temporal and spatial coverage of the subsurface data is limited, so this study probably underestimates the usefulness of including subsurface temperature data in efforts to predict ENSO. Finally, cross-validated forecasts using a Markov model that includes an annual cycle are shown to be less skillful than forecasts using a seasonally invariant Markov model. The reason for this appears to be that dividing the data yields an insufficient database to derive an accurate Markov model.

1. Introduction

Forecasts of ENSO conditions are provided by a number of groups using various modeling techniques in the *Climate Diagnostics Bulletin* on a monthly basis (Barnston and Ropelewski 1992; Ji et al. 1996; Cane et al. 1986; Chen et al. 1995; Penland and Magorian 1993; Barnett et al. 1993). The "linear inverse modeling" technique of Penland is relatively simple: predictions of sea surface temperature anomaly (SSTA) are made using a Markov model produced from observations of SST in the tropical Pacific from about 1950 to the present. The Markov model is effectively the average of the ensemble of matrices that transform the state vector (composed of the gridded SST anomaly data) from each month of the data record to the state vector a given number of months later. If 1) the true dynamics associated with ENSO are linear or weakly nonlinear, 2) the SST data

reflects most of the relevant dynamics in the evolution of SST, and 3) the driving effect of noise on the system is not so large as to preclude forecasting with useful lead time, then this empirically derived model (EDM) will be able to provide useful predictions of SST. In addition to simplicity, another advantage of the linear inverse modeling technique is that predictions are made using observational data alone, so questions regarding the validity of various model parameterizations are avoided.

The pioneers of this empirical modeling technique have reasoned that SST alone contains all of the relevant dynamics to a great extent (Penland and Sardeshmukh 1995, hereafter PS), and it is clear that useful predictions of up to about 9 months are made using this approach. On the other hand, it stands to reason that if thermocline depth anomalies and hence upper-level heat content is not in equilibrium with wind anomalies that are produced by (and are in equilibrium with) SST anomalies on the ENSO timescale, then SST alone cannot encompass all of the dynamics relevant to evolution of ENSO-related SST anomalies. Recent observational studies have highlighted the importance of subsurface temperature anomalies in the evolution of ENSO (Bigg and Blundell 1989; Kessler 1990; Tourre and White 1995;

Zhang and Levitus 1997). Thus it can be expected that heat content, if included in the state vector of the tropical Pacific, should improve prediction of SST by linear inverse modeling.

This paper presents an extension of the linear inverse modeling technique to include subsurface temperature data with SST data to determine the Markov model of the tropical Pacific on interannual timescales. The expanded Markov model will be used to forecast the Niño 3.4 index, which is the SST anomaly spatially averaged over the region (5°N–5°S, 170°–120°W). These forecasts will be compared to forecasts using only SST data to determine whether or not there is a discernible improvement by including subsurface temperature data. Work on this subject has recently been undertaken by Xue and Leetmaa (1997, hereafter XL) in which the EDM is built from observed SST and model-derived sea level data from 1980 to 1997. Xue and Leetma (1997) concluded that sea level anomalies (which are similar to subsurface temperature anomalies) contribute to predictive skill of ENSO. In addition, Latif and Graham (1992) applied a statistical prediction scheme to data from an oceanic GCM and concluded that heat content anomalies at 5°N contribute to predictive skill of equatorial Pacific SST. Our study, which involves strictly observational data, is complementary to these studies.

In section 2, a brief description of the data used in this study is presented, and the leading EOFs of SSTA and heat content anomaly are discussed. In section 3, the techniques for constructing the linear model are presented. A comparison of predictions using an EDM using SSTA alone and an EDM built from SSTA data plus heat content anomaly data is presented in section 4, including an examination of how predictions of warm events compare. In section 5, a number of important issues will be discussed, including 1) the sensitivity of our results to the meridional resolution of the datasets, and 2) the importance of separating data used for model-training and forecasting, a technique referred to as "cross validation." In section 5, we will also examine whether the forecasting skill can be improved if the invariant Markov model of sections 2–4 is extended to include an annual cycle. A summary of the findings is presented in section 6.

2. Data

a. SST data

The sea surface temperature data source is the Comprehensive Ocean–Atmosphere Data Sata (COADS) enhanced data made available by the Climate Diagnostics Center (CDC) of the Cooperative Institute for Research in Environmental Sciences (CIRES). The "enhanced sets" 1a and 1b together span the period 1950–95. The global data, supplied on a 2° × 2° grid, were consolidated on a 4° lat × 10° long grid spanning the region

32°S–32°N, 120°E–70°W. As in PS, the monthly mean data was low-pass filtered by smoothing using a 3-month running mean to better isolate interannual variability. Our focus will be on the period 1965–93; we felt the subsurface data were too incomplete before 1965, and the subsurface dataset ends in June of 1993 (see section 2b). The mean SST for each calendar month over the period January 1965 to June 1993 was removed from the corresponding calendar month, leaving anomalies.

An empirical orthogonal function (EOF) analysis of the smoothed anomalies was performed and the first two SSTA EOFs are shown in Figs. 1a and 1b. The principal component (PC) time series corresponding to the first and second EOFs are shown in Fig. 1c as the solid and dot–dashed lines, respectively. The first EOF, which accounts for 36% of the total SSTA variance, resembles the peak of the ENSO cycle, and the corresponding PC time series is highly correlated with the cold tongue index (SST anomaly averaged over 4°S–4°N, 180°–90°W) over the period 1965–93, the correlation being 0.95. The second EOF, accounting for 9% of the total SSTA variance, is comparable to that determined by PS using standard (not "enhanced") COADS data (cf. with their Fig. 1), although our warm patch is more constrained to the eastern equatorial region whereas theirs extends farther westward.

b. Subsurface temperature data

The subsurface temperature data were initially produced by the National Oceanographic Data Center (NODC). Temperatures of the upper 400 m were retrieved using expendable bathythemographs (XBTs) and Nansen bottles. This dataset was supplied to us by researchers at the Climate Diagnostics Center of CIRES who had formed monthly anomalies for the period 1940–93, and placed the data on a 4° × 4° grid and discretized it to 13 vertical levels (corresponding to depths of 0, 10, 20, 30, 50, 75, 100, 125, 150, 200, 250, 300, and 400 m). Our domain, 30°S–30°N, 120°E–72°W, is approximately the same as the domain for SST. The dataset is incomplete, in fact sampling in the Southern Hemisphere was particularly sparse. Spatial and temporal interpolation was used to guess the missing data. The temperature anomalies were then integrated over the top 300 m to produce a heat content anomaly (HCA) field that spanned the period 1965–93. Since the climatology was based on the period 1940–93, an annual march was present in the 1965–93 HCA data and this was subsequently removed, leaving monthly anomalies. The observations tapered off during the last months of 1993, so only January to June of 1993 is used. As with the SST anomalies, the HCA field was temporally smoothed using a 3-month running mean.

The first two heat content anomaly EOFs account for 33% and 11% of the total anomaly variance, respectively. The relative sparsity of the subsurface temperature data in comparison to the SST data, particularly
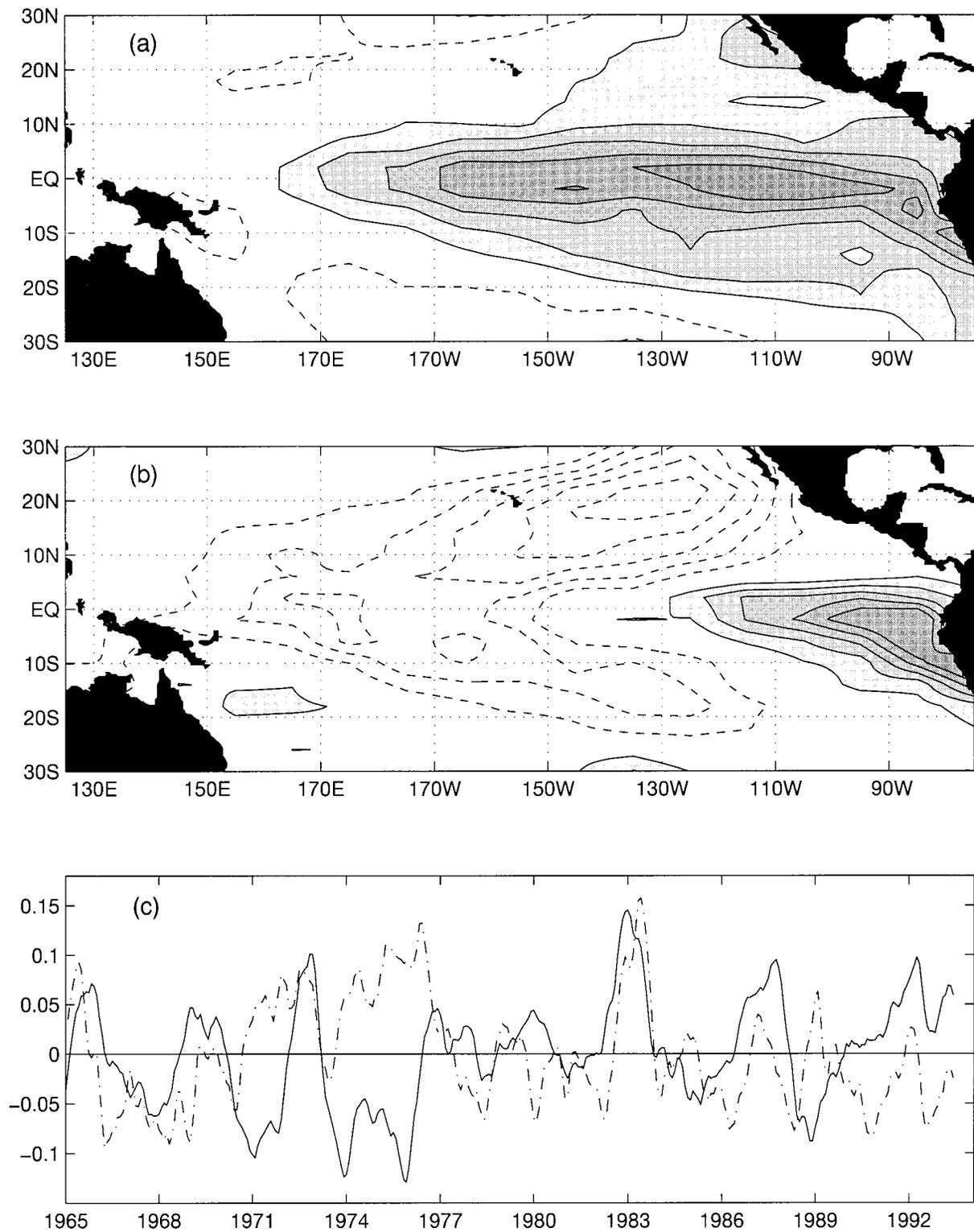
FIG. 1. (a) First and (b) second EOFs of Pacific (32°N–32°S, 120°E–70°W) SSTAs from Jan 1965 through Jun 1993 from COADS enhanced data. Solid (dashed) contours denote positive (negative) EOF amplitude. The zero contour line is suppressed. (c) Principal component time series corresponding to the first (solid) and second (dot–dashed) EOFs.

in the Southern Hemisphere, led us to expect that the HCA field would be noisy in comparison to the SST field. The drop in the variance of successive HCA EOFs is quite similar to that for SST, however, which suggests that the HCA field is not noisier than the SST field.

The first two EOFs for the heat content anomaly and the corresponding PCs are shown in Fig. 2a and Fig. 2b, respectively. The EOFs are very similar to the first two heat content EOFs determined by Zhang and Levitus (1996, their Fig. 9), although they found that the explained percentages of variance of the first two EOFs were about 27% and 16%, in contrast to our 33% and 11%. The first two PCs are shown in Fig. 2c. It is apparent that the second EOF (dot–dashed line) lags the first EOF (solid line) fairly consistently throughout the nearly 30 yr of the record. Indeed, the lag correlation peaks at a value of 0.61 for the time series of the second EOF lagging behind that of the first EOF by 9 months. This corroborates the suggestion by Zhang and Levitus (1997) that the second EOF of heat content represents the transition phase of an ENSO eigenmode that is apparent in the heat content anomaly. There is no corresponding behavior in the EOFs of SSTA; although during the 1982–83 warm event it appears that the second SSTA EOF might lag the first EOF, it does not do so consistently. The correlation between the first and second SSTA EOFs does not exceed 0.3 at any lag.

3. Empirically derived model

The modeling approach employed here has been applied to the tropical Pacific in a number of previous studies (Blumenthal 1991; Penland and Magorian 1993; Xue et al. 1994; PS). It is assumed that the deterministic (or internal) dynamics of the coupled tropical system on ENSO timescales are well represented as linear and that nonlinearities can be approximated as white noise (temporally, but not necessarily spatially) and so can be included alongside observational error in a noise-forcing term. The dynamical equation takes the form of a stochastic differential equation:

$$\frac{d\mathbf{x}}{dt} = \mathbf{B}\mathbf{x} + \boldsymbol{\epsilon},$$ (1)

where $\mathbf{x}$, the state vector of the system, might include SSTA information only or the state of both SSTA and heat content anomaly fields. Lowercase bold letters denote vectors, and capital bold sansserif letters denote matrices. The deterministic portion of the dynamics is represented by the square matrix $\mathbf{B}$, and $\boldsymbol{\epsilon}$ represents stochastic external forcing. In addition, $\boldsymbol{\epsilon}$ is a column vector of the same length as the state vector $\mathbf{x}$. The general solution for $\mathbf{x}(t + \tau)$ given $\mathbf{x}(t)$ is

$$\mathbf{x}(t + \tau) = \exp\left[\int_t^{t+\tau} \mathbf{B}(t'') \, dt''\right] \mathbf{x}(t)$$

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

The first term on the right is the deterministic contribution, and for seasonally invariant deterministic dynamics, that is, using the simplification that $\mathbf{B}$ does not depend on the time of the year, the deterministic term simplifies to

$$\mathbf{x}(t + \tau) = e^{\mathbf{B}\tau}\mathbf{x}(t).$$ (3a)

The matrix operator,

$$\mathbf{G}(\tau) = e^{\mathbf{B}\tau},$$ (3b)

is sometimes referred to as the propagator or the Green function (Penland and Magorian 1993). When applied to the state vector of the system at any time $\mathbf{t}$, it gives a prediction of the state at the later time $\mathbf{t} + \tau$. Assuming that the deterministic dynamics are seasonally invariant, then while $\mathbf{G}(\tau)$ depends explicitly on τ, $\mathbf{B}$ is independent of τ.

If seasonality were included, different Markov models would be built for different seasons of the year. In other words, $\mathbf{B}$ would be taken as constant during any given month or season. An example of seasonal EDMs is provided by Blumenthal (1991) in which four separate transition matrices (one for each season) are built from output of the Zebiak–Cane model (Zebiak and Cane 1987). It has been determined in a previous study that there is a statistically significant annual variation in $\mathbf{B}$ as it is empirically derived from the observed SSTA over the period 1950–95 (Johnson et al. 1999, manuscript submitted to *J. Climate,* hereafter JBS). In regard to forecasting using the empirically derived dynamics, however, including seasonality does not significantly improve predictions. The reason for this is that the improvement of prediction by including seasonality is negated by the reduction in the skill of the determination of each seasonal $\mathbf{B}$, which results from the reduction in the size of the data ensemble that trains each seasonal EDM [by a factor of 4 (12) when determining four seasonal (monthly) EDMs]. We assume that this trade-off exists also with respect to the dynamics associated with heat content, although we have not checked this. Since our interest is to test improvement of prediction, we therefore neglect seasonality in the derivation of the EDMs in this study. To support our contention that this is appropriate, a comparison of the predictive skill of an annually varying SST-only Markov model versus a constant SST-only Markov model is presented in section 5.

In practice, the empirically derived model, $\mathbf{G}(\tau)$, is derived from the record of the state of the system according to

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

where $\langle \ \rangle$ denotes an ensemble average. It is apparent that assuming linear, stable dynamics and building a Markov prediction model boils down to something very similar to applying linear regression. In linear regression, when seeking to predict with a lead time τ, $\mathbf{G}(\tau)$ is calculated specifically at that lead time τ since $\mathbf{G}(\tau)$ is the best least squares fit linear model of the data to
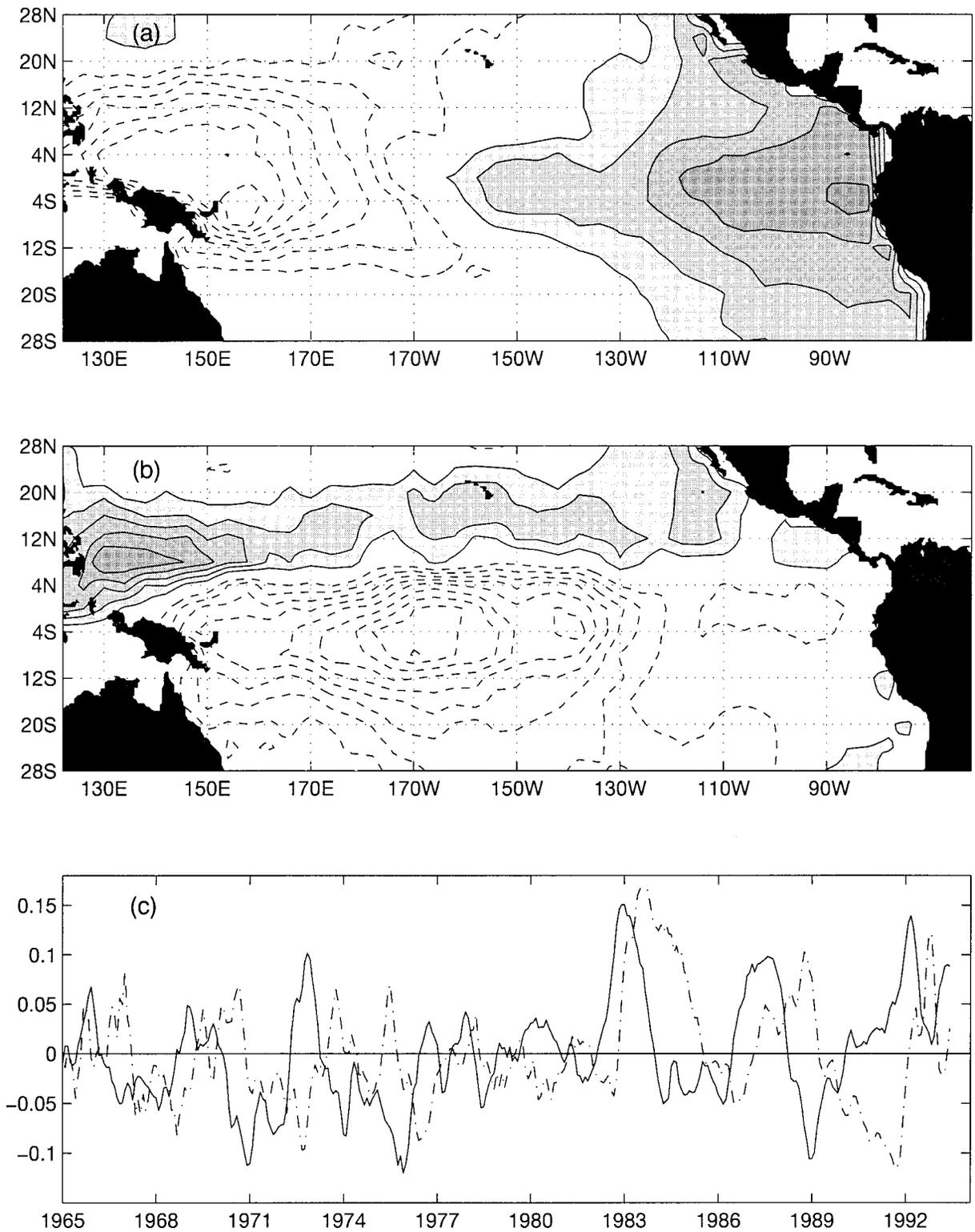
FIG. 2. Same as Fig. 1 except for HCA, which is temperature anomaly integrated between 0- and 300-m depth. The data domain is 30°N–30°S, 120°E–76°W.

determine the state at time $\mathbf{t} + \tau$ given the state at the previous time $\mathbf{t}$. With the Markov assumption, in contrast, $\mathbf{G}$ is calculated for a given lead time generally other than τ. The resulting propagator is then spectrally decomposed to determine $\mathbf{B}$ via (3b), and subsequently $\mathbf{G}(\tau)$ for other lead times τ is calculated using (3b) (Penland 1989). For example, from (3b) it is apparent that if $\mathbf{G}(5)$ is first empirically derived, $\mathbf{G}(10)$ will be given by $\mathbf{G}(5)^2$ or $[\mathbf{G}(5)][\mathbf{G}(5)]$. In addition, $\mathbf{G}(11)$ would be given by $\mathbf{G}(5)^{11/5}$.

When applying this scheme to monthly Pacific data, we choose τ to be approximately 5 months in (4) for the following reasons: If τ is chosen small, such as $\tau = 1$, then error in $\mathbf{G}(1)$ would multiply 10 times in the calculation of $\mathbf{G}(10)$, for example, where it would multiply only twice if $\tau = 5$ in (4). In addition, DelSole (1998) has shown that if τ in (4) is small, the Markov model is erroneous since at small τ the time series being fitted to the model is differentiable in time, whereas a true first-order Markov process is never time differentiable. Choosing τ to be large, such as $\tau = 10$, is also avoided since at large lead times all predictions tend toward zero, so it is expected that the dynamics cannot be as reliably derived, given the limited duration of the training data, as when τ is smaller. In addition, growth of the effects of nonlinearities in the internal dynamics may worsen the linear approximation when larger lead times are used in (4).

We have found that determining $\mathbf{G}$ from $\mathbf{B}$ is quite laborious, since, among other difficulties (Penland 1989), the eigenvector pairs usually need to be manually rearranged so that the left (right) eigenvector has the same column (row) index as the corresponding eigenvalue in the eigenvalue matrix. For this reason, we determine $\mathbf{G}(\tau)$ for 3–12 months using an approximate method of first determining $\mathbf{G}(3)$, $\mathbf{G}(4)$, $\mathbf{G}(5)$, and $\mathbf{G}(6)$ and then multiplying two of these to find $\mathbf{G}(7-12)$, for example $\mathbf{G}(8) = [\mathbf{G}(4)][\mathbf{G}(4)]$, $\mathbf{G}(9) = [\mathbf{G}(5)][\mathbf{G}(4)]$, and $\mathbf{G}(10) = [\mathbf{G}(5)][\mathbf{G}(5)]$. The task of determining $\mathbf{B}$ is effectively skipped. Note that our modeling formulation is not Markovian in the strictest sense, where only one lead time (instead of four) is selected but the results are insensitive to the approximation.

It has been suggested that at least 2 months of (monthly mean) data are required to accurately determine each element in $\mathbf{G}(\tau)$ (specifically, 1 month of data is required to determine each element of $\mathbf{G}(\tau)$ and 1 month is required to determine each element of the same-sized noise covariance matrix; Penland 1989; c. Penland 1997, personal communication). According to this criterion, the size of $\mathbf{G}(\tau)$ is limited to be on the order of 13 × 13 elements. For this reason, $\mathbf{G}(\tau)$ is built in a greatly reduced "principal component space," where the state vector at each month is given by the PCs during that month of the leading 10 or so EOFs of the SSTA data matrix spliced together with, in the case that heat content anomaly information is desired, a number of the leading PCs of HCA.

4. Prediction of Niño 3.4 SSTA

Predictions are made of the Niño 3.4 index, which is the SSTA spatially averaged over the region (5°N–5°S, 170°–120°W). (Since our grid boxes are 4° wide meridionally, our Niño 3.4 region actually extends between only 4°N and 4°S.) Testing propagators (EDMs) of different size is necessary because it is unknown in advance how many EOFs should be kept in order to retain the relevant dynamics. Therefore, (4) was applied to state vectors containing from 2 to 40 SST-only elements plus between 0 and 10 HCA elements to produce (square matrix) propagators containing 4–2500 elements. Keeping too few EOFs omits important dynamics. Keeping too many EOFs not only degrades the propagator by including noise, but also decreases the accuracy by which each element of the propagator can be determined by increasing the number of elements to be determined. Numerous trials reveal that the optimal number of EOFs to be kept is usually that for which about 70% of the variance is retained, which varies between about 10 and 20 EOFs for observational fields, as opposed to model-generated data, for which the number of EOFs is usually smaller, since models generally contain a smaller number of modes (see, e.g., Thompson 1998; Blumenthal 1991). This is to be expected since models have no measurement error and invariably contain simplified dynamics.

Propagators were determined for prediction lead times ranging between $\tau = 3$ months and $\tau = 12$ months. Predictions with $\tau = 1$, 2 months lead time are omitted because the 3-month running mean smoothing of the data artificially inflates the skill at these lead times. Predictions beyond 12 months give a nearly zero anomaly because the eigenmodes [also known as principal oscillation patterns; Hasselmann (1988)] are purely decaying. This decaying of the eigenmodes represents a limitation of empirically derived models that does not exist, in principle, for dynamical models. The Zebiak–Cane model (Zebiak and Cane 1987), for example, is used to predict SST anomalies associated with ENSO more than one year in advance. These forecasts are currently presented in the *Climate Diagnostics Bulletin.*

A cross-validation method following Penland (1997) is employed to separate the training period of the EDM (which should include as much of the data as possible) from the verification period. For this reason, we refer to our predictions as "forecasts" as opposed to "hindcasts" (Davis 1976). The cross-validating technique proceeds as follows; the 28.5-yr period is broken into 14 segments of approximately 2 yr each. Predictions are made for each time segment using a propagator built using the other 13 segments. Three-month predictions are made of the Niño 3.4 index for months 4 (or February 1965) to 342 (June 1993), 4-month predictions are made for months 5–342, and so on.

Skill of prediction is measured by the correlation of the true time series of Niño 3.4 and the time series of
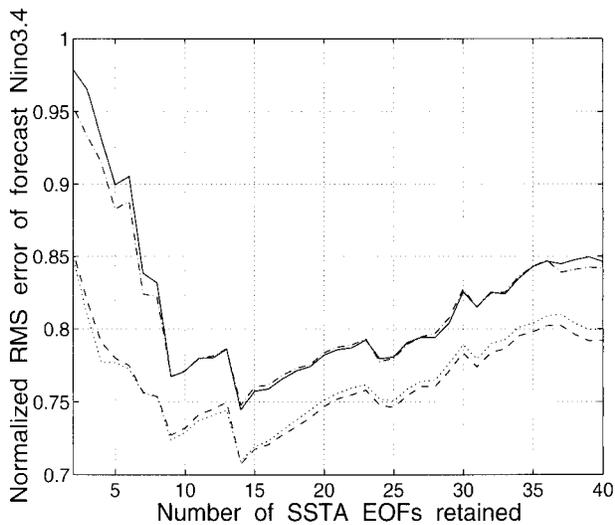
FIG. 3. Rms error of Niño 3.4 index forecasts averaged over lead times of 3, 4, . . . , 12 months plotted vs the number of principal components of SST used in deriving the Markov model. The solid line corresponds to forecasting by a Markov model using SST anomaly data alone. The dot–dashed, dashed, and dotted lines correspond to the cases in which the first principal component, first two PCs, and first three PCs of HCA are added to PCs of SST anomaly to construct the Markov model, respectively. The rms error is normalized such that predictions by climatology have an rms error of 1.

the forecast, and by the rms difference between the two time series. The rms difference between the two time series is normalized by dividing by the rms of the true anomaly time series itself. This provides a useful perspective—a normalized rms (or nrms) error equal to one is exactly what would result from predicting the anomaly using climatology (predicting zero anomaly always). Good correlation but large nrms error results when the phasing of the Niño 3.4 time series is better simulated than the amplitude. During times of zero anomaly, for example, error of prediction is not measured by the correlation at all, but is captured by rms difference. Low correlation but small nrms error tends to indicate that the error of prediction is largest when the anomalies are highest. Forecasting using climatology is an extreme example that illustrates this point. Clearly, the greater the anomaly, the greater the forecast error. But while the ensemble nrms error is always equal to one (which is lower than persistence nrms error at lead times longer than 6 months), the correlation is always zero (which is always worse than for persistence forecasts at all lead times). For the forecasts reported in this study, we find that strong correlation is always accompanied by small nrms error and vice versa, indicating that forecast error is equally distributed between periods with warm and cold events and periods without events.

Figure 3 shows the rms error of Niño 3.4 index forecasts, normalized to climatology, averaged over the 10 lead times of 3–12 months. The solid line corresponds to the case in which only SST anomalies are used to forecast Niño 3.4. As expected, predictive skill increases

versus number of retained EOFs to a maximum, after which the higher number of retained EOFs degrades the predictive skill. The dot–dashed line corresponds to the case in which the first principal component of HCA is added to 2–40 PCs of SST anomaly to construct the Markov model, the dashed line shows the error when the first two PCs of HCA are added, and the dotted line corresponds to when the first three PCs of HCA are added. When only the first principal component of HCA is added for constructing the Markov model, little change is seen in the forecasts. This is probably due to the fact that the first PC for HCA is nearly identical to the first PC for SST (the correlation between the first PC of SST anomaly and of HCA is 0.92; hence, both PCs approximately represent the amplitude of the mature phase of the ENSO). Adding the first PC of HCA only adds redundant information so predictions are nearly unchanged. When the first two PCs of HCA are added, however, there is a clear reduction in forecast error, but further addition of PCs of HCA does not lead to further improvement in the forecasts (up to 10 PCs were added in tests; forecast error beyond 3 added PCs of HCA is not shown).

It was noted in reference to Fig. 2 that the second PC of heat content anomaly lags the first by 9 months with a correlation of 0.61. Although this is not a very strong correlation, the second PC of HCA comes closer to representing a "transitional" phase of ENSO than the second PC of SSTA and it may be for this reason that it provides seemingly useful information for predicting ENSO that is not present in SSTA data. Figure 2 shows that variance in the second EOF of HCA extends meridionally over the full domain, from 30°N to 30°S. The improvement of forecasting by inclusion of the second HCA EOF, if it proves to be significant, might be taken to corroborate the suggestion that delayed oscillator theory, in which only equatorially confined wave processes contribute to the development and demise of warm events (Battisti 1989), should be modified to include important contributions of off-equatorial processes to the evolution of ENSO events (Zhang and Levitus 1997). This is not necessarily the case, however, since our results do not necessarily imply that dynamical processes outside the equatorial region (from 8°N and 8°S) are contributing to the evolution of ENSO events. Recall that the time series associated with the second EOF of HCA aids prediction (not the spatial EOF itself) and that this time series *lags* the Niño 3.4 index. This PC cannot, therefore, represent a contribution to the *development* of warm events, but rather corresponds to *consequences* of the warm events (which, as shown by the structure of the EOF, involve latitudes away from the equator). A lagging time series can aid the Markov model, nonetheless, by helping to determine the direction that the Niño 3.4 index is taking. For example, if the time series is higher (lower) than the Niño 3.4 index, it may indicate that the Niño 3.4 index is decreasing (increasing). Thus, Markov predictions of warm events

can be aided by including a consequence of warm events, namely the second EOF of HCA. In addition, since this EOF most likely represents a *consequence* of ENSO events, the apparent forecast improvement by adding it to the Markov model cannot be taken to suggest that delayed oscillator theory requires modification to include extratropical processes.

It is interesting to note in Fig. 3 that when the first two PCs of heat content anomaly are added to fewer than 14 PCs of SST anomaly, namely to up to the first 6, the improvement in forecast skill is much more dramatic. So if only the first six EOFs of SST were used to construct the Markov model, it would be concluded that inclusion of HCA information, namely the second EOF, boosts forecast skill significantly. However, it is apparent in Fig. 3 that the seventh and ninth SST anomaly EOFs contain useful forecasting information, which, when included, reduces the apparent contribution by the second HCA EOF. It is difficult to conclude that the seventh and ninth SST EOFs contain the information provided by the second HCA EOF, however, since the correlation between the second PC of HCA and the seventh and ninth PCs of SST is only 0.31 and 0.16, respectively. For comparison, the correlation between the second PC of HCA and the third and eighth PCs of SST, neither of which appear to contribute appreciably to forecast skill, is 0.51 and 0.38.

It has been suggested that weighting the HCA data more heavily than the SST anomaly data will further improve the forecasts (XL). In a manner similar to the approach taken by XL, we have tested this by splicing the SST and HCA data together before determining the EOFs, having increased the magnitude of the HCA field so that it plays a larger role in determining the EOFs. In contrast to XL, we found that weighting the observed HCA more heavily resulted in degrading the forecast skill. In addition, we examined the forecast skill of Niño 3.4 using HCA information alone. The Markovian approach is inappropriate in this case, so linear regression is used to predict SSTA. The normalized rms prediction error averaged over lead times of 3–12 months is 0.82 (0.76) using 15 EOFs of HCA (SSTA) to forecast the Niño 3.4 index. As the number of EOFs of the predictor and predictand fields is varied, the forecast error using HCA as the predictor field always exceeds the error when SSTA is used as the predictor. It is also determined that Markovian forecasting of SST using HCA alone is less skillful than forecasting by linear regression with HCA as the predictor (not shown). In conclusion, it appears that it is beneficial to augment the SSTA with portions of the HCA information, but that SST anomaly taken as a whole is a better predictor of SST anomaly than HCA, taken as a whole, is a predictor of SST anomaly.

The correlation and nrms error of the two EDMs are shown versus lead time of the forecast in Figs. 4a and 4b, respectively. The solid line represents forecast skill by the top 14 PCs of SST anomaly alone; the dashed
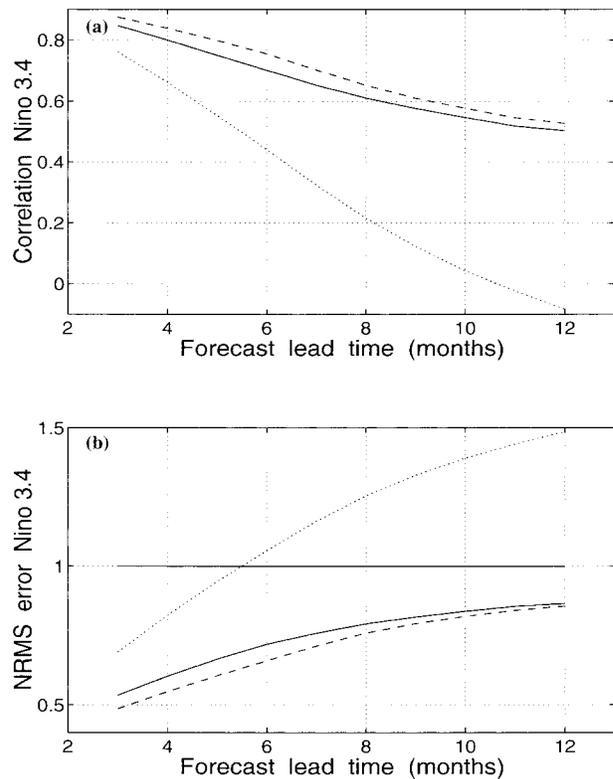

FIG. 4. (a) Correlation of the observed Niño 3.4 (4°N–4°S, 170°W–120°W) index for the period Jan 1965–Jun 1993 with the forecast time series using an empirically derived Markov model of 14 PCs of SST anomaly alone (solid line) and 14 PCs of SSTA plus 2 PCs of heat content anomaly (dashed line) vs prediction lead time. The dotted line corresponds to the persistence forecast, which is the lag-τ autocorrelation of the Niño 3.4 index. (b) Normalized rms forecast error of Niño 3.4 index. Solid, dashed, and dotted lines are the same as in (a). The rms error is normalized (hence ''nrms'') such that prediction using climatology has nrms error equal to one at all τ.

line represents the skill when these 14 SST PCs are augmented by the first 2 PCs of HCA. The dotted line represents skill of persistence forecasts, and the error of forecasting shown in Fig 4b is normalized such that forecasting by climatology has an error of exactly one. Figure 4 shows that 1) the Markov skill far exceeds the skill of persistence and climatology and 2) the improvement in skill by including the second EOF of heat content anomaly boosts the forecast skill by a small amount at all lead times. Our results are consistent with the findings of XL of better prediction by inclusion of sea level anomaly.

To further examine the difference in forecasting by including heat content anomaly, the predictions of the Niño 3.4 index over the period January 1965 to June 1993 are shown in Fig. 5. The dotted line corresponds to the observed Niño 3.4 index, and the solid and dashed lines correspond to the forecasts for a lead time of 6 months for the EDMs of SSTA alone and for SSTA and heat content anomaly combined, respectively. It is apparent in Fig. 5 that most warm and cold events are
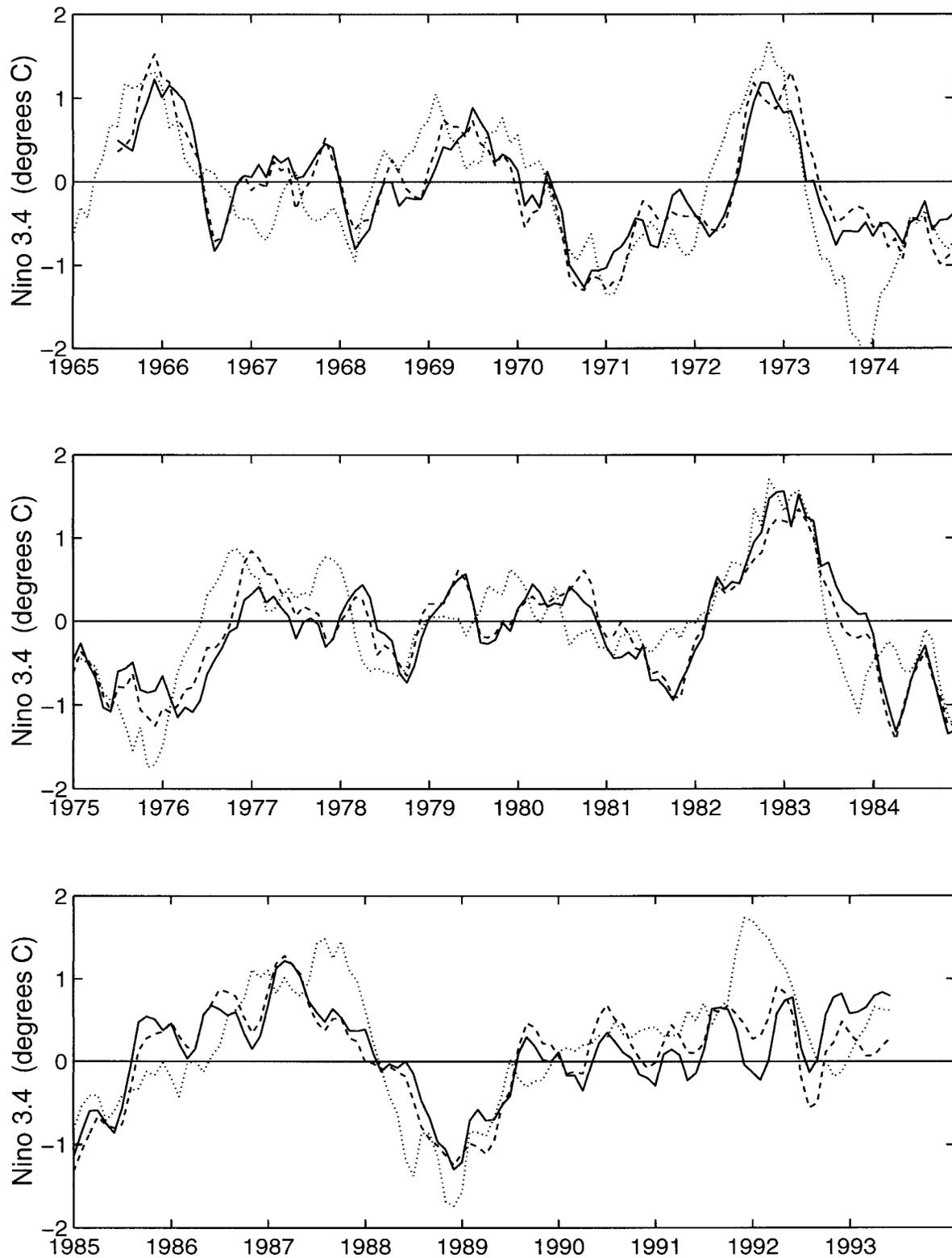
Fig. 5. The 6-month forecasts of Niño 3.4 SSTA index using the 14 PC SSTA-only EDM (solid line) and using the EDM of 14 PCs of SSTA plus 2 PCs of heat content anomaly EDM (dashed line) over the period Jan 1965–Jun 1993. The dotted line is the observed Niño 3.4 index.

fairly well forecasted by both EDMs. Exceptions include the cold event of 1974, and the warm events of 1987 and at the end of 1991. The apparent lack of skill for both EDMs in the early 1990s is common to most model predictions of ENSO and corroborates the hypothesis that during the early 1990s the mechanisms associated with the ENSO mode were altered or absent. It has been suggested that the decadal mode was stronger than the interannual variability during this time (Goddard and Graham 1997; Latif et al. 1997). Stochastic variability has also been shown to produce decadal variations in an intermediate coupled model of ENSO (Fluegel and Chang 1998, manuscript submitted to *J. Geophys. Res.*).

An improvement by adding heat content is visible in some instances, such as the warm event of 1976 and the warm early 1990s. Overall, however, the predictions made by the EDMs with and without the second PC of HCA appear throughout the 29-yr period to be more similar to each other than to the true Niño 3.4 index. Indeed, while the correlation between the observed Niño 3.4 time series and the 6-month predictions by the EDMs of SSTA and SSTA plus heat content are 0.70 and 0.76, respectively, the correlation between the two forecasted time series is 0.93. Thus, it is clear that predictions do not greatly change when heat content anomaly data are included in the construction of the EDM. This probably reflects a close link between surface and subsurface temperature anomalies.

In the interest of determining how forecast skill during different phases of ENSO compare, we examine the forecasting skill specifically of positive SST anomalies and how adding heat content anomaly affects these forecasts. For this purpose, warm events were classified as all months for which the Niño 3.4 index exceeded 0.7°C. Figure 6 represents as solid line segments the warm months for which the prediction of the Niño 3.4 index is examined in detail. Figure 7 shows the normalized rms forecast error versus lead time for warm SST anomaly predictions using 14 PCs of SST anomaly alone (dotted line) and using 14 SST anomaly PCs plus two HCA PCs (dot–dashed line). The solid and dashed lines show the nrms error over the entire index and so are the same as from Fig. 4b. Figure 7 shows that 1) predictions of positive SST anomalies are more skillful than predictions for all temperatures on average, and 2) in comparison to the whole record, at most lead times during warm events there is more improvement in skill when subsurface temperature data is added to SST (exceptions include lead times of 7, 8, and 9 months).

Figures 3, 4, and 7 indicate that the Niño 3.4 forecast, averaged over the period 1965–93, is robustly improved by adding heat content anomaly information. An improvement is apparent over the entire range of lead times investigated. It would be interesting to test the statistical significance of the apparent improvement. The set of about 340 monthly forecast errors are approximately normally distributed with a mean value of nearly zero. If forecasting by including HCA is better, then the stan-dard deviation of the set of forecast errors should be distinctly smaller than the standard deviation of the set of errors of forecasts using SST anomaly alone. The statistical significance of the difference between standard deviations is measured using the "f statistic," which is the ratio of the variances (or squared standard deviations) of the two sample distributions (Spiegel 1988). Figure 8 represents as solid lines the f statistic for the difference in forecast skill over all temperatures (panel a) and over warm events in particular (panel b). The mean forecast error during warm events is not close to zero but is rather consistently negative, which is to be expected since the modeling technique assumes stable dynamics and consequently only decaying eigenmodes (PS), so anomalies are underestimated on average. The f test is still applicable, however, assuming that the sample distributions are normal, as they appear to be.

To determine the statistical significance level for the difference of the standard deviations of the sample sets, we require the number of independent members of the sample sets. The 1-month lagged autocorrelation of the forecast error is used to determine the number of independent samples using the commonly applied technique by Leith (1973). Based on the independent sample size for the set of prediction errors for each forecast lead time, the level at which the difference in the errors is 75% significant is determined and drawn as dashed lines in Fig. 8. As forecast lead time increases, the lag-one autocorrelation increases, the independent sample size decreases, and this "raises the bar" for a given significance level. Although we might require the f statistic to exceed the 95% significance level before concluding that the forecasting improvement is "significant," it can be seen in Fig. 8 that even the modest 75% significance level is never met, so it is apparent that the forecasting improvement is not statistically significant.

5. Empirical modeling issues

Given the importance of equatorially confined wave propagation to the ENSO cycle, it would be worthwhile to investigate whether or not a finer meridional resolution of the subsurface temperature data than the 4° in our data is required to see the true contribution by subsurface information to SST in predicting ENSO, if there is any. To answer this question, a version of the high-resolution (1°) reanalysis dataset of the National Centers for Environmental Prediction referred to as RA6 will be utilized to develop Markov models to test forecast skill as meridional resolution is varied from 1° to 2° to 4°. The Markov models are constructed using as predictors the reanalysis SST and heat content, and the predictand is the reanalysis SST Niño 3.4 index. The zonal grid box size in all cases involving RA6 data was 6°. The dataset was developed by assimilating XBT and Tropical Atmosphere Ocean data into the Geophysical Fluid Dynamics Laboratory modular ocean model to
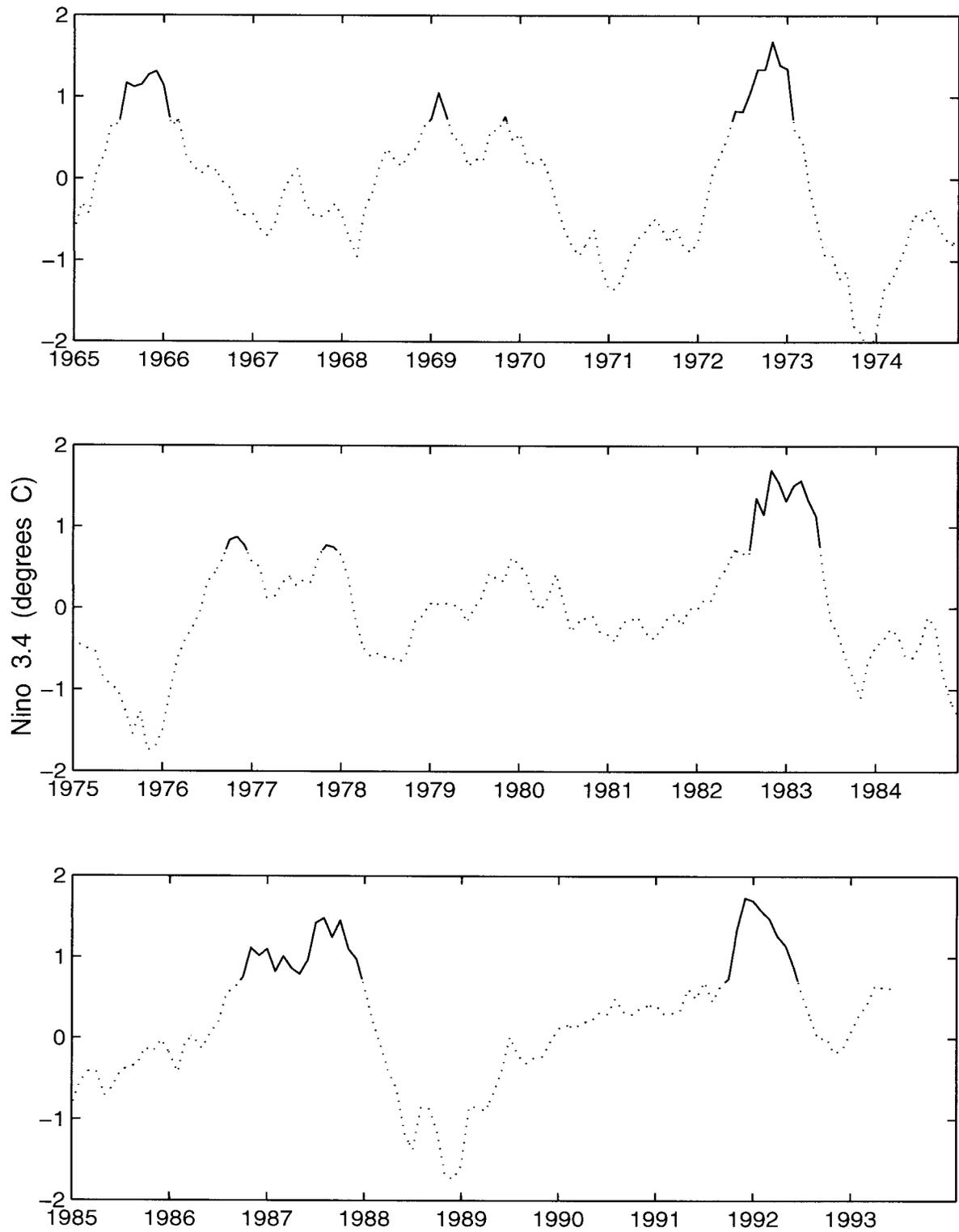
FIG. 6. Observed Niño 3.4 index (dotted line) highlighting the segments during which the index exceeds 0.7°C.
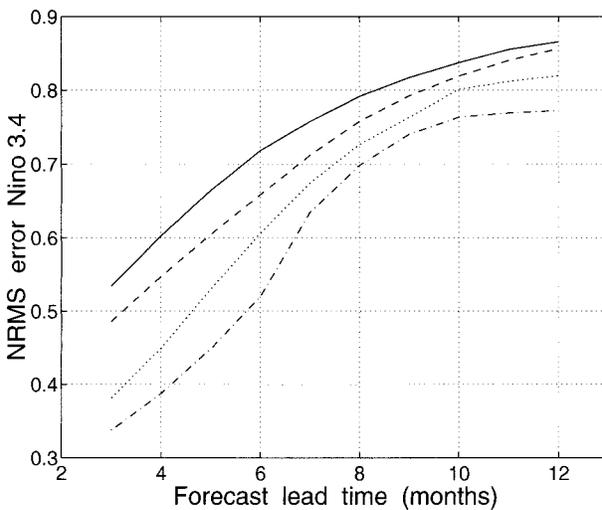
FIG. 7. Normalized rms error of forecasts of Niño 3.4 index at all times during Jan 1965–Jun 1993 for 14 PC SST-only EDM and 14 PC SST plus 2 PC HCA EDM (solid and dashed lines, respectively, the same as in Fig. 4b) and for forecasts of warm events (defined as when Niño 3.4 index exceeds 0.7°C) for 14 PC SST-only EDM and 14 PC SST plus 2 PC HCA EDM (dotted and dot–dashed lines, respectively).

produce a dataset including surface and subsurface temperature in the Pacific over the period 1980–95 (Xue and Leetmaa 2000; Behringer et al. 1997).

Figure 9a shows the normalized rms error of the Niño 3.4 forecast over the 16-yr span of the dataset versus the number of PCs of SST anomaly that were included in the construction of the Markov model. In this run, only SST was used to determine whether or not increasing the meridional resolution of SST would improve the performance of the EDM. As before, cross validation was employed to minimize artificial skill. The solid line corresponds to 4° meridional resolution and the dashed line and dotted lines correspond to 2° and 1° meridional resolution, respectively. It is apparent that changing the meridional resolution in the SST-only model does not significantly alter forecast skill.

Figure 9b shows the results of the model runs in which forecasts were made using EDMs of SST having 4° meridional resolution plus two PCs of heat content anomaly having 4° meridional resolution (dashed line), 2° resolution (dot–dashed line), 1° resolution (dotted line), as well as the case for SST alone (solid line, same as solid line in Fig 9a). The heat content anomaly field was derived, as before, by vertically integrating subsurface temperature anomalies to a depth of 300 m. The second principal component is again the part of the HCA data that improves forecast skill. Varying the meridional resolution does not significantly vary the forecast skill; in fact the uniformity of the forecast skill is striking. It is clear that changing the resolution of the HCA dataset does not change the second EOF by much; the second PC remains very much the same. This should not be surprising, however, since the second EOF is essentially
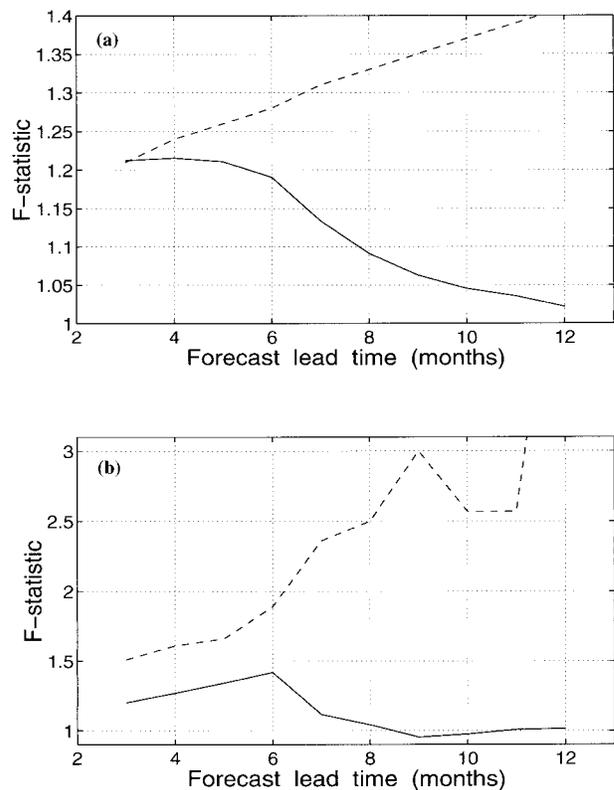

FIG. 8. (a) The f statistic, defined as the ratio of squared standard deviations of sampling distributions of Niño 3.4 forecast errors by EDM using SST alone to EDM using SST plus two PCs of HCA, vs lead time of forecast (solid line). Dashed line is the 75% significance level, which varies according to the number of independent samples. (b) The f statistic when forecasts are limited to months during which Niño 3.4 index exceeds 0.7°C. Dashed line is the 75% significance level.

a large-scale pattern. It is reasonable to conclude from the tests using the RA6 data that the coarseness of the observational subsurface data during 1965–93 does not degrade forecasting in our linear inverse modeling scheme. It may be that there is skill to be gained by increasing the meridional resolution but this improvement cannot be captured by linear inverse modeling as it is employed in this study.

Figure 10 is presented with two points in mind: the first is to illustrate the importance of cross validation when applying empirical dynamical modeling to forecasting the same field from which the predictive model is developed, and the second is to demonstrate that skill is diminished when the Markov model is built to include an annual cycle. All four curves represent the normalized rms error of the forecasts averaged over all lead times (from 3 to 12 months) of Niño 3.4 over the time period from 1965 to June 1993. The models are built using SST only. The solid line is the same as the solid line in Fig. 3 and the dashed line represents skill when cross validation is not employed, all other things being
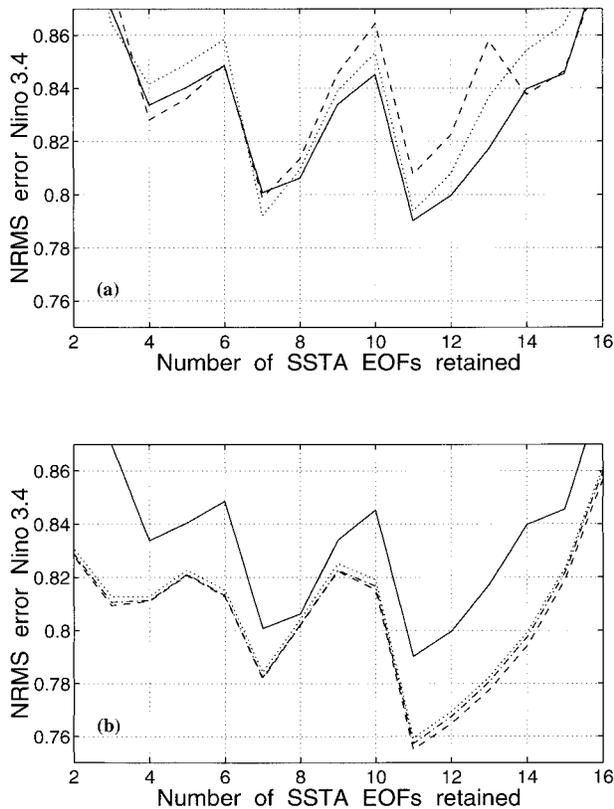
FIG. 9. (a) Normalized rms error of Niño 3.4 forecasts over the period 1980–95 in which the Markov model is constructed using SST anomaly data from the NCEP reanalysis set that spans 1980–95. The solid line, dashed line, and dotted line refer to runs in which the meridional resolution of the data is 4°, 2°, and 1°, respectively. (b) As in (a) but the dashed, dot-dashed, and dotted lines correspond to runs in which the PCs of SST anomaly data with 4° meridional resolution are augmented by the top two PCs of heat content anomaly data with 4°, 2°, and 1° resolution, respectively.
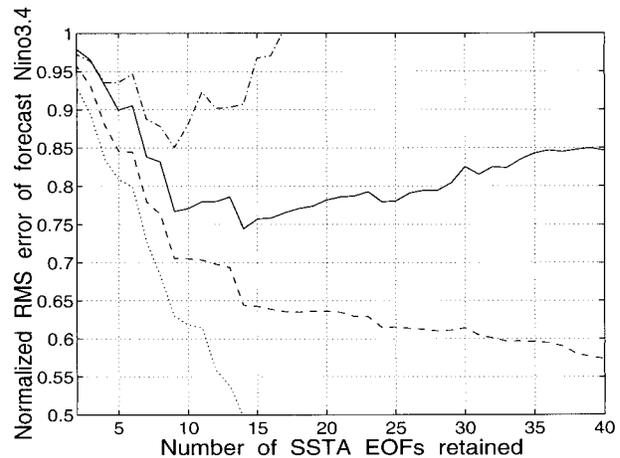

FIG. 10. Normalized rms error of Niño 3.4 forecasts over the period 1965–93 for cases in which the Markov model is 1) seasonally invariant and cross validated (solid line, same as in Fig. 3), 2) seasonally invariant and not cross validated (dashed line), 3) seasonally varying and cross validated (dot–dashed line), and 4) seasonally varying and not cross validated (dotted line).

the same. When more than 14 principal components are employed to build the model, the cross-validated skill steadily decreases while the unvalidated prediction skill steadily increases. This appears to be an example of what was reported by Davis (1976) who related intrinsic predictive skill to artificial (apparent) skill in the cases in which 1) the predictand is among the data used to construct the predictor and 2) the time frame of the predictand and the time frame of the data used to construct the predictor do not overlap. The same behavior is discussed and demonstrated by Lorenz (1977). When no cross validation is used, the apparent skill exaggerates the true skill. When cross validation is employed, on the other hand, the apparent skill *underestimates* the skill that would be theoretically be possible if the training period were unbounded in length. Cross-validated skill therefore represents a lower bound on the intrinsic predictive skill. Davis (1978) suggests that the intrinsic predictive skill can be estimated by averaging the experimental forecast and hindcast skills (the solid and dashed lines, respectively, in the nonseasonal case),

which would appear to give a normalized rms error (averaged from 3 to 12 months) of about 0.7 if the training period were much longer than 26.5 yr as used.

JBS showed that seasonality is present in the internal dynamics (B or G) but found that building a seasonally varying Markov model did not improve predictions of Niño 3.4. To demonstrate that this is again the case using the approximately Markovian scheme employed in this study, a seasonally varying Markov model was built by deriving 12 different versions of each of $\mathbf{G}(3)$, $\mathbf{G}(4)$, $\mathbf{G}(5)$, and $\mathbf{G}(6)$, such that a 3-, 4-, 5-, and 6-month Markov model is built for making predictions starting in each of the 12 calendar months. The 10-month Markov model for states beginning in January, for example, is then built by left-multiplying the January to June 5 month) matrix operator by the June to November (5 month) operator. This combined operator is right-multiplied by the column vector representing the January state to predict the November state. The dot–dashed line in Fig. 10 shows the forecast error of the cross-validated monthly varying Markov model and the dotted line shows the apparent forecast error when cross validation is not applied. The apparent increase (decrease) in forecast error for the cross-validated (not cross validated) case versus the number of PCs of SST used is much more rapid than in the seasonally invariant cases (solid and dashed lines) since the sample size of the data used to make the monthly varying Markov model is 1/12th the size, or about 28 (since about 28 yr of data are used). When the number of EOFs of SST reaches 28, the error in the unvalidated case will reach zero, and the Markov model will predict with apparently perfect skill. Considering the question of whether or not building a monthly varying Markov model will improve forecasts, while it has been shown that there is clearly a seasonal

cycle in the empirically derived internal dynamics (JBS), comparing the solid and dot–dashed lines in Fig. 10 indicates that cross-validated forecasts by the seasonal Markov model of this study are less skillful than the forecasts by the nonseasonal one. The reason for this appears to be that dividing the already limited dataset (by 12) leaves training samples that are too small to build an accurate seasonally varying Markov model.

Our results are at odds with those of XL in two ways. First, we find that (a) our seasonal Markov model has a lower forecast skill than our nonseasonal Markov model and (b) weighing subsurface temperature data more heavily than SST degrades, rather than improves, forecasting of SST anomalies. Difference a may arise if XL do not use cross validation in their study as we do, as was illustrated in Fig. 10. It is not clear to us whether or not XL use cross validation. The differences must otherwise be due to other differences in the methods and data employed, which are as follows. First, our datasets are purely observational and span the time period 1965 to June 1993, whereas their data is the product of assimilating observations into an ocean GCM and spans the time period 1980–97. Probably more importantly, however, are the differences in how the Markov models are constructed. Our Markov models are built using **G**(3, 4, 5, and 6 months), whereas their Markov models are built using **G**(1 month). In addition, the state variables used in our formulation are the values of the first n principal component time series (temporal coefficients of EOFs of the data) at a given time, whereas XL splice together the principal components of an EOF analysis of the fields and then perform a second EOF analysis of those spliced PCs. Xue and Leetma (1997) then use the principal components of the second EOF analysis to compose the state vectors. Although the differences in our methods are substantial, it is not apparent to us why these differences lead to the different findings in our respective studies.

6. Summary

In this study, empirical dynamical models are constructed based on the linear inverse modeling approach of Penland (Penland and Magorian 1993). The EDMs are Markov models derived from principal component time series of the observed monthly sea surface temperature anomaly data and the observed heat content anomaly data (subsurface temperature anomalies integrated to a depth of 300 m). The SST data are from COADS and span the region 32°N–32°S, 120°E–70°W. The HCA data are from NODC (and supplied to us by CDC) and extend from 30°S–30°N to 120°E–72°W.

Forecasts of the Niño 3.4 index (SSTA average over the region 4°N–4°S, 170°W–120°W) are made over the ~2 yr span of each of 14 segments of the data that extends from January 1965 to June 1993 using an EDM derived from the other 13 segments of data to avoid artificial skill arising from overlapping data periods used

for training and for prediction. Prediction lead time ranged from 3 to 12 months. The forecasts by SST-only EDMs were compared to forecasts by EDMs built from principal components (time coefficients of EOFs) of SSTA plus principal components of HCA to determine if subsurface temperature information could improve prediction by adding relevant dynamical information not already contained in the SSTA data.

Forecast skill is measured by correlation between the true and predicted Niño 3.4 index time series and by rms difference between the true and predicted indices for each of 342-τ forecast months, where τ is the lead time in months. The results, shown in Fig. 3, suggest greater skill when subsurface data is included. The second EOF of HCA is identified as the key element of the HCA data, which, when included, enhances forecast skill. However, a side by side comparison of 6-month forecasts, shown in Fig. 5, indicates that adding subsurface data does not induce obvious systematic changes in skill. Figure 7 shows that forecasting of positive SST anomalies, defined as when Niño 3.4 exceeds 0.7°C, appears to have above average skill, and again, the second EOF of HCA is found to be the key to the apparent enhancement of forecast skill.

To determine whether or not the skill enhancement associated with adding heat content data to the EDMs is statistically significant, the standard deviation of the error in the set of forecasts made with the SST-only Markov model is compared to that made with the Markov model using both SST and heat content anomaly data. Figure 8 shows that the level of significance never exceeds 75%. The high autocorrelation of the forecast errors reduces the independent sample size of the sets of errors tremendously, rendering the difference between the sets of errors statistically indistinguishable.

While the information of the heat content does not significantly improve the skill of the 14-EOF SSTA-only EDM, if 6 or less EOFs of SSTA are used for prediction, it is seen that the skill of the EDM is greatly improved by adding the second EOF of HCA (see Fig. 3). Our results indicate that the 7th through 14th EOFs of SST contain as much valuable information for predicting the Niño 3.4 index as is contained in the second EOF of heat content. The first EOF of heat content is not a useful addition to the SST-only Markov model since it is very similar to the first EOF of SST and is therefore largely redundant.

Although it could not be concluded that adding heat content anomaly data to SSTA data in the derivation of the Markov model can produce improvement in forecast skill that is statistically significant, there is at least an *apparent* improvement in prediction at all lead times, as shown in Fig. 4 and Fig. 7. We expect that limitations in the heat content data are a limiting factor in the determination of the extent to which subsurface temperature anomaly information can contribute in efforts to predict ENSO events.

REFERENCES

Barnett, T. P., M. Latif, N. Graham, M. Fluegel, S. Pazan, and W. White, 1993: ENSO and ENSO-related predictability. Part I: Prediction of equatorial Pacific sea surface temperature with a hybrid coupled ocean–atmosphere model. *J. Climate,* **6,** 1545–1566.

Barnston, A. G., and C. F. Ropelewski, 1992: Prediction of ENSO episodes using canonical correlation analysis. *J. Climate,* **5,** 1316–1345.

Battisti, D. S., 1989: On the role of off-equatorial oceanic Rossby waves during ENSO. *J. Phys. Oceanogr.,* **19,** 551–559.

Behringer, D. W., M. Ji, and A. Leetmaa, 1998: An improved coupled model for ENSO prediction and implications for ocean initialization. Part I: The Ocean Data Assimilation System. *Mon. Wea. Rev.,* **126,** 1013–1021.

Bigg, G. R., and J. R. Blundell, 1989: Equatorial Pacific Ocean prior to and during El Niño of 1982/83: A normal mode view. *Quart. J. Roy. Meteor. Soc.,* **115,** 1039–1069.

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

Cane, M. A., S. E. Zebiak, and S. C. Dolan, 1986: Experimental forecasts of El Niño. *Nature,* **321,** 827–832.

Chen, D., S. E. Zebiak, A. J. Busalacchi, and M. A. Cane, 1995: An improved procedure for El Niño forecasting: Implications for predictability. *Science,* **269,** 1699–1702.

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.

——, 1978: Predictability of sea level pressure anomalies over the North Pacific Ocean. *J. Phys. Oceanogr.,* **8,** 233–246.

DelSole, T., 1998: A fundamental limitation of Markov models. COLA Rep. 55, 32 pp. [Available from Center for Ocean–Land–Atmosphere Studies, Institute of Global Environment and So-

ciety, Inc., 4041 Powder Mill Road, Suite 302, Calverton, MD 20705.]

Fluegel, M., and P. Chang, 1998: Stochastically induced decadal modulation of El Niño–Southern Oscillation. *Geophys. Res. Lett.,* **26,** 2473–2476.

Goddard, L., and N. E. Graham, 1997: El Niño in the 1990's. *J. Geophys. Res.,* **102,** 10 423–10 436.

Hasselmann, K., 1988: PIPs and POPs—A general formalism for the reduction of dynamical systems in terms of Principal Interaction Patterns and Principal Oscillation Patterns. *J. Geophys. Res.,* **93,** 11 105–11 021.

Ji, M., A. Leetmaa, and V. Kousky, 1996: Coupled model predictions of ENSO during the 1980s and the 1990s at the National Centers for Environmental Prediction. *J. Climate,* **9,** 3105–3120.

Kessler, W. S., 1990: Observations of long Rossby waves in the northern tropical Pacific. *J. Geophys. Res.,* **95,** 5183–5217.

Latif, M., and N. E. Graham, 1992: How much predictive skill is contained in the thermal structure of an oceanic GCM? *J. Phys. Oceanogr.,* **22,** 951–962.

——, R. Kleeman, and C. Eckert, 1997: Greenhouse warming, decadal variability, or El Niño? An attempt to understand the anomalous 1990s. *J. Climate,* **10,** 2221–2239.

Lorenz, E. N., 1977: An experiment in nonlinear statistical weather forecasting. *Mon. Wea. Rev.,* **105,** 590–602.

Penland, C., 1989: Random forcing and forecasting using Principal Oscillation Pattern analysis. *Mon. Wea. Rev.,* **117,** 2165–2185.

——, 1997: A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D,* **98,** 534–558.

——, and T. Magorian, 1993: Prediction of Niño 3 sea surface temperatures using linear inverse modeling. *J. Climate,* **6,** 1067–1076.

——, 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.

Tourre, Y. M., and W. B. White, 1995: ENSO signals in upper-ocean global temperature. *J. Phys. Oceanogr.,* **25,** 1317–1332.

Xue, Y., and A. Leetmaa, 1997: Predictability of ENSO: A study with Markov models. *Proc. 22d Annual Climate Diagnostic Workshop,* Berkeley, CA, 42–45.

——, and ——, 2000: ENSO prediction with Markov models: The impact of sea level. *J. Climate,* in press.

——, M. A. Cane, S. E. Zebiak, and M. B. Blumenthal, 1994: On the prediction of ENSO: A study with a low order Markov model, *Tellus,* **46A,** 512–528.

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

Zhang, R. H., and S. Levitus, 1996: Structure and evolution of interannual variability of the tropical Pacific upper ocean temperature. *J. Geophys. Res.,* **101,** 20 501–20 524.

——, and ——, 1997: Interannual variability of the coupled tropical Pacific ocean–atmosphere system associated with the El Niño–Southern Oscillation. *J. Climate,* **10,** 1312–1330.