

Forecasting ENSO Events: A Neural Network–Extended EOF Approach

FREDOLIN T. TANGANG,* BENYANG TANG, ADAM H. MONAHAN, AND WILLIAM W. HSIEH

Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, British Columbia, Canada

(Manuscript received 21 January 1997, in final form 9 May 1997)

ABSTRACT

The authors constructed neural network models to forecast the sea surface temperature anomalies (SSTA) for three regions: Niño 4, Niño 3.5, and Niño 3, representing the western-central, the central, and the eastern-central parts of the equatorial Pacific Ocean, respectively. The inputs were the extended empirical orthogonal functions (EEOF) of the sea level pressure (SLP) field that covered the tropical Indian and Pacific Oceans and evolved for a duration of 1 yr. The EEOFs greatly reduced the size of the neural networks from those of the authors' earlier papers using EOFs. The Niño 4 region appeared to be the best forecasted region, with useful skills up to a year lead time for the 1982–93 forecast period. By network pruning analysis and spectral analysis, four important inputs were identified: modes 1, 2, and 6 of the SLP EEOFs and the SSTA persistence. Mode 1 characterized the low-frequency oscillation (LFO, with 4–5-yr period), and was seen as the typical ENSO signal, while mode 2, with a period of 2–5 yr, characterized the quasi-biennial oscillation (QBO) plus the LFO. Mode 6 was dominated by decadal and interdecadal variations. Thus, forecasting ENSO required information from the QBO, and the decadal–interdecadal oscillations. The nonlinearity of the networks tended to increase with lead time and to become stronger for the eastern regions of the equatorial Pacific Ocean.

1. Introduction

Over the last 10 years, many models have been developed for forecasting the El Niño–Southern Oscillation (ENSO) phenomenon. Barnston et al. (1994) classified these models into three groups: dynamical-coupled models, hybrid-coupled models, and statistical models. When used to forecast the sea surface temperature anomaly (SSTA) in the east-central region of the equatorial Pacific for the period of 1982–93 at a lead time of 6 months, these models all performed with comparable correlation skills in the 0.60's. For longer lead time forecasts, the dynamical models appeared to have better skills than the statistical models. Data assimilation appears to offer further forecast improvements; for example, Chen et al. (1995) found skill improvement when a “nudging” technique was used to assimilate data into the Zebiak–Cane model.

Almost all statistical models used to forecast ENSO were linear (e.g., Barnett 1984; Graham et al. 1987b; Barnston and Ropelewski 1992; Tang 1995). In the last 10 years, a nonlinear empirical modeling technique has

emerged in the field of artificial intelligence; neural network models are now widely applied to numerous scientific, engineering, and financial areas. Elsner and Tsonis (1992, 1993) showed that neural networks could perform better than linear statistical models when dealing with chaotic plus random noise systems. Following the neural network forecasts of tropical Pacific wind stress by Tang et al. (1994), Tangang et al. (1997) forecasted the SSTA in the Niño 3.4 region with neural network models using wind stress as a predictor. Tangang et al. (1998) broadened the investigation by: 1) comparing the SSTA forecasts skills over various regions (Niño 3, 3.4, 3.5, 4, etc.) in the tropical Pacific; 2) comparing the sea level pressure (SLP) versus the wind stress as predictors, where the SLP emerged as the better predictor at long lead times; and 3) using an ensemble average of 20 forecasts with different initial random weights to greatly reduce the fluctuations in the forecasts, resulting in better skills.

Despite the encouraging results, the models we used in these earlier works were rather large (with many degrees of freedom), rendering an interpretation of the inner workings of the models most difficult—a common problem with neural network modeling.

In this paper, our objective is to greatly simplify and clarify the neural network models by three new initiatives: 1) We introduced extended empirical orthogonal functions (EEOF) of the SLP as input to the neural networks, which turned out to be much more effective in compressing the predictor fields than the empirical orthogonal function (EOF) method used in our earlier

* Current affiliation: Department of Marine Science, National University of Malaysia, Selangor, Malaysia.

Corresponding author address: Dr. William W. Hsieh, Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.
E-mail: william@eos.ubc.ca

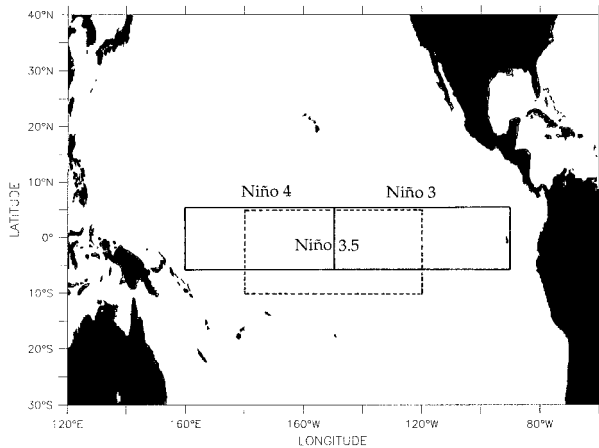


FIG. 1. The location of the three regions of interest.

papers. As the EEOF allows lag compression in addition to space compression, this enabled us to use a much smaller network. 2) We introduced a pruning technique to trim the networks of unnecessary structure, which further reduced the size of the networks. 3) By applying spectral analysis to the neural networks, we were able to understand how the network functioned, and in particular how nonlinear the network was, which enabled us to conclude that in the relation between the predictors and the predictand, nonlinearity tended to increase with lead time and to become stronger for the eastern regions of the equatorial Pacific.

The rest of this paper is organized as follows. Section 2 describes the data and the EEOF analysis of the SLP field, while in section 3, we briefly illustrate the neural network model. Section 4 presents the forecast results, section 5 performs network pruning, and section 6 applies spectral analysis on the networks.

2. Datasets and preliminary processing

The same sea surface temperature (SST) data used in Tangang et al. (1997b) were employed here. The SST data were obtained from a combination of Smith's SST and Reynolds's SST datasets. The Smith SSTs were reconstructed using the EOF method (Smith et al. 1996) and were available in 2° latitude \times 2° longitude grids over the period 1950–92. The Reynolds's SST data were derived from the in situ (ship and buoy) and satellite SST using optimum interpolation (Reynolds and Smith 1994). This dataset is available in 1° latitude \times 1° longitude grids from 1982 to the present. The combined SST used in this study comprised Smith's SSTs from 1952 to 1992 and the averaged $2^\circ \times 2^\circ$ 1993 Reynolds's SSTs. Prior to combining these two datasets, preliminary analyses were performed to determine their compatibility over the overlapping period of 1982–92. These analyses included the calculation of the climatology for the equatorial Pacific (20°S – 20°N) and the SSTA indices of the regions of interest (Fig. 1). The results confirmed

that the climatologies (not shown) were almost identical and the indices were highly correlated (0.98–0.99). A comparison between the SSTA indices calculated from this combined dataset and those calculated from the Comprehensive Ocean–Atmosphere Data Set (COADS) (Woodruff et al. 1987, 1993) showed that the indices were highly correlated (0.94–0.99).

We considered three discrete regions, that is, the Niño 4, Niño 3.5, and Niño 3 regions, which roughly represent the western-central, the central, and the eastern-central parts of the equatorial Pacific Ocean (Fig. 1). The SSTA for each region was calculated by spatially averaging the SST in the $2^\circ \times 2^\circ$ boxes in each region. The climatological means were then removed and the anomalies were smoothed with a three-point running average in time.

The SLP data came from COADS. Our results in Tangang et al. (1997b) showed that the SLP data were less noisy than the wind stress and, hence, became the better predictor at long lead times. Also, the SLP data covered a wider region of 30°S – 20°N and 50°E – 90°W (including both the tropical Indian and Pacific Oceans). We chose the region based on the following factors: 1) Several studies indicated that the surface wind and SLP anomalies, which eventually caused most of the changes in the SST associated with ENSO, originated from the equatorial Indian Ocean and propagated slowly eastward into the Pacific (Barnett 1985; Yasunari 1990; Gutzler and Harrison 1987). 2) Based on the EOF and EEOF analyses of the near-global SLP of Graham et al. (1987a), most of the variability due to ENSO was contained in the 40°S – 20°N , 50°E – 90°W region. However, COADS had too many missing values in the 40° – 30°S band, particularly in the southeastern Pacific Ocean. Hence we truncated the latitudinal extent to 30°S – 20°N .

Prior to the EEOF analysis, the SLP fields were averaged to coarser grids of 4° by 10° from the $2^\circ \times 2^\circ$ COADS grids. The anomalies were calculated by removing the climatological means at each grid point, followed by smoothing with a three-point running average in time and a 1–2–1 filter in each spatial dimension. The use of the EEOF technique in the preprocessing of the SLP field allows lag as well as spatial compressions of the data field into several dominant modes (Weare and Nasstrom 1982; Graham et al. 1987a). This is particularly advantageous in relation to neural network modeling since fewer inputs are required. For example, the network requires only 7 inputs (if the first seven modes were used) for 1-yr compression of the predictor data field, as opposed to 28 inputs (i.e., 7×4 , for four 3-month lags) if ordinary EOFs were used as in Tangang et al. (1997, 1998).

Following our earlier works, we used the SLP data field, evolved over 1-yr, to be the main predictor. To characterize the evolution over this period, we calculated the EEOF at 3-month intervals over 1 yr for the period of 1952–93. This was achieved by “stacking”

MODE 1 of SLP EEOF (26.5 perc.) (LAG:9)

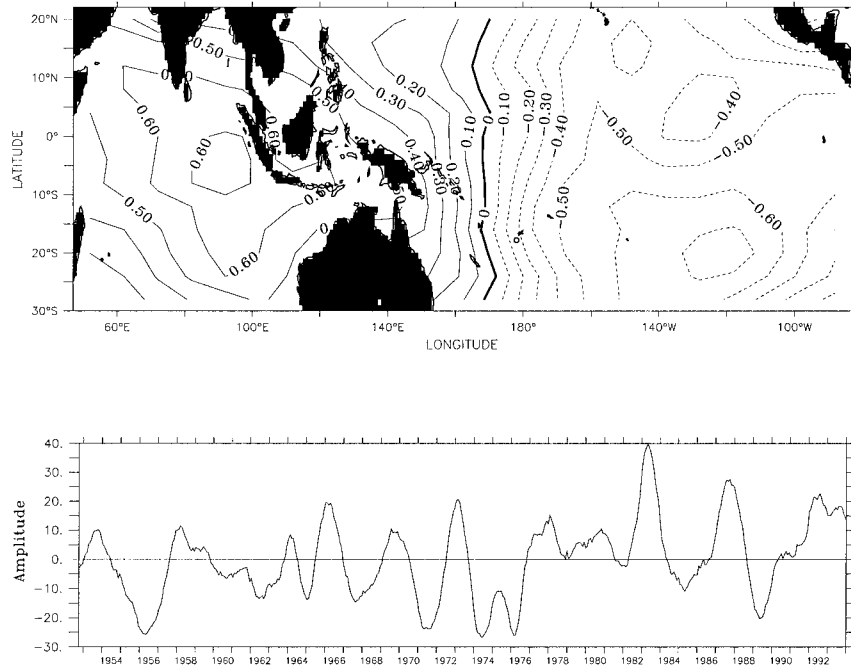


FIG. 2. Mode 1 of the SLP EEOF at lag 9 months. (a) The spatial pattern. The contours represent correlation between the EEOF amplitude and the SLP anomaly at each grid point. (b) The corresponding standardized amplitude.

the four temporal sequences of the SLP field into a matrix of the form

$$\mathbf{X} = \begin{pmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,M-9} \\ \vdots & \vdots & & \vdots \\ X_{NP,1} & X_{NP,2} & \cdots & X_{NP,M-9} \\ X_{1,4} & X_{1,5} & \cdots & X_{1,M-6} \\ \vdots & \vdots & & \vdots \\ X_{1,7} & X_{1,8} & \cdots & X_{1,M-3} \\ \vdots & \vdots & & \vdots \\ X_{1,10} & X_{1,11} & \cdots & X_{1,M} \\ \vdots & \vdots & & \vdots \\ X_{NP,10} & X_{NP,11} & \cdots & X_{NP,M} \end{pmatrix}, \quad (1)$$

where the NP ($=299$) and M ($=504$) were the number of spatial and monthly points, respectively. The EEOF was then calculated by an EOF analysis on the covariance of this stacked matrix. The eigenvectors provided a sequence of four “snapshots” (each covering a 3-month period) characterizing the evolution of a spatial pattern over a 1-yr period. As was found in Graham et al. (1987a), the ENSO signal appeared to be divided between the first two modes, each displaying different, partially overlapping portions of the SLP variations associated with ENSO. Figure 2a shows the spatial patterns (at lag 9 months) of the first mode, which accounted for 26.5% of the total variance. The contour

represents the correlation values between the amplitudes (i.e., the temporal coefficients) and the SLP anomaly at each grid point. The spatial patterns clearly indicated an ENSO association, with a positive center of action in the Indian–Asian region and the opposite center (negative) located in the southeastern tropical Pacific. The patterns for the subsequent lags (i.e., lags 6, 3, and 0) resembled the spatial pattern for lag 9, with intensification and slight eastward displacement of the Indian–Asian center (not shown). Hence these patterns characterize the seesaw oscillations between the two centers of action. The temporal coefficient shown in Fig. 2b also clearly indicated typical ENSO signatures with high values during El Niños and low ones during La Niñas (i.e., El Viejos).

The second mode (Fig. 3a), which explained 10.4% of the total variance, also displayed two centers, with, at lag 9, a high pressure cell centered in the southeastern equatorial Pacific and a low covering the Asian–Australian region. The patterns appeared to propagate eastward as they evolved to the most recent lag (i.e., lag 0), at which time a complete reversal of the pressure pattern occurred (i.e., high in the west and low in the east, with resemblance to the first mode). This evolution seems to suggest that this mode is also associated with ENSO. Mode 1 had spatial and temporal patterns similar to the SLP EOF mode 1 in Tangang et al. (1998). Both the EOF and the EEOF produced similar results since

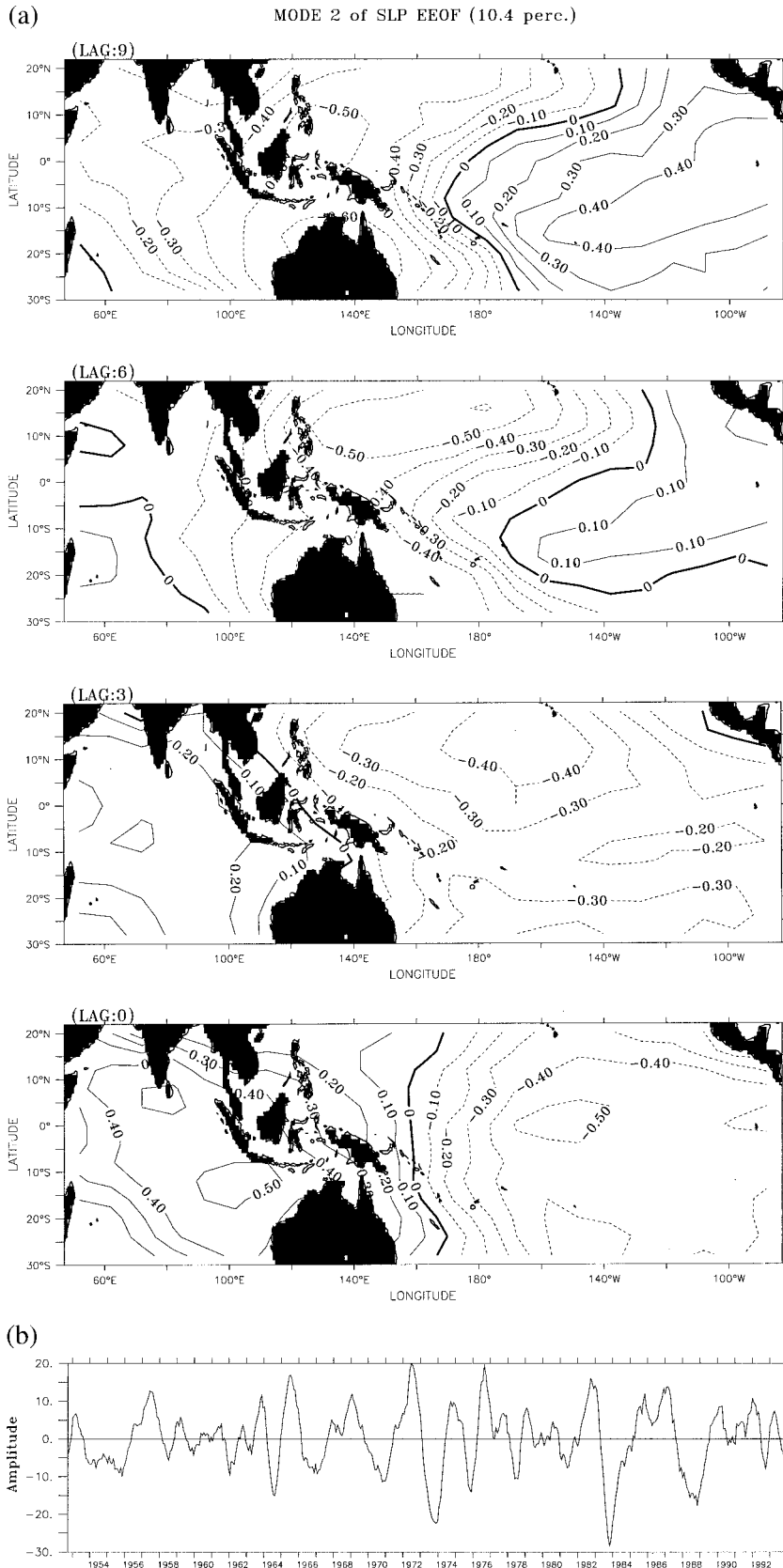


FIG. 3. As in Fig. 1 except for mode 2, with the patterns for lag 9 to lag 0.

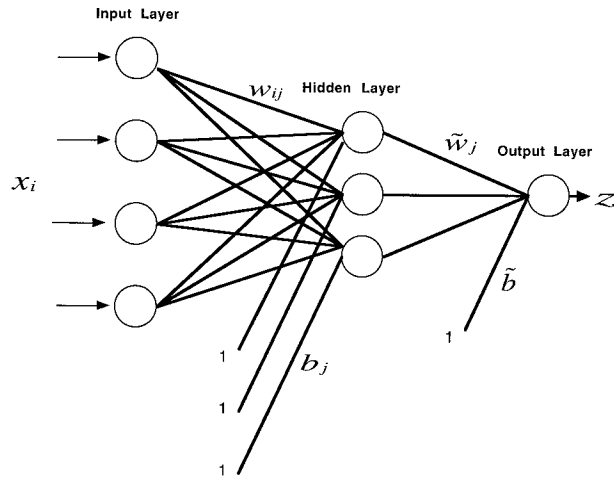


FIG. 4. An example of a feed-forward neural network model, where there are four neurons in the input layer, three in the hidden layer, and one in the output layer. The w_{ij} and \tilde{w}_j are the weights, and the b_j and \tilde{b} are the biases. The b_j and the \tilde{b} can be regarded as the weights for constant inputs of magnitude 1.

both modes characterized the standing oscillation. However, since the EOF analysis could not capture propagating features by a single mode, the SLP EOF mode 2 in Tangang et al. (1998) had different spatial and temporal patterns than the SLP EEOF mode 2. As will be discussed in section 6, we interpreted the SLP EEOF mode 1 as the low-frequency oscillation (LFO) mode and mode 2 as the LFO plus the tropospheric quasi-biennial oscillation (QBO) mode, in relation to the LFO and QBO described in Barnett (1991), Ropelewski et al. (1992), and Goswami (1995).

3. Neural network and model design

We adopted the architecture of a three-layered feed-forward network with one input layer, one hidden layer, and one output layer (see Tangang et al. 1997, 1998 for details). In a layered feed-forward network (such as shown in Fig. 4), with a linear transfer function in the output layer and a hyperbolic tangent function in the hidden layer, the output of the network is given by

$$z = \sum_j \tilde{w}_j y_j + \tilde{b}, \quad (2)$$

where the y_j is the output of j th hidden neuron,

$$y_j = \tanh\left(\sum_i w_{ij} x_i + b_j\right), \quad (3)$$

with x_i the i th input, and the parameters w_{ij} , \tilde{w}_j , b_j , and \tilde{b} the weights and biases that need to be optimized. The optimization, often referred to as network training, is an iterative procedure called the back-propagation method (Rumelhart et al. 1995; Smith 1993). We adopted an algorithm of steepest descent with momentum and adap-

tive learning rates (Jacobs 1988), in which we used a cost function of the form

$$C = \frac{a_1}{2} \sum (z - z_o)^2 + a_2 a_4^2 \sum \frac{(w/a_4)^2}{1 + (w/a_4)^2} + a_3 a_4^2 \sum \frac{(\tilde{w}/a_4)^2}{1 + (\tilde{w}/a_4)^2}, \quad (4)$$

where the first term on the rhs is basically the sum squared error (SSE), and the last two are weight decay terms that penalize excessive weights (Chauvin 1989). This form of a cost function represents a trade-off between residual error and model complexity, which have been shown to produce robust networks (Hanson and Part 1989; Ji and Psaltis 1990; Weigend et al. 1990, 1991). As a_4 becomes large, (4) simplifies to

$$C = \frac{a_1}{2} \sum (z - z_o)^2 + a_2 \sum w^2 + a_3 \sum \tilde{w}^2, \quad (5)$$

which is known as ridge regression in the statistical community. As in Tangang et al. (1998), we fixed the values of a_1 , a_2 , and a_4 to 1, 0.1, and 100, respectively. However, since we have smaller networks, we set a_3 to 2.0, a smaller value than what was used in Tangang et al. (1998). The values of a_2 and a_3 can be interpreted as ‘‘adjustment’’ toward linear or nonlinear net, as the hidden neurons have nonlinear transfer functions while the output neuron has a linear transfer function. For the case of $a_3 \gg a_2$, more weight restriction is imposed on the output layer and hence the resulting net tends to be more nonlinear.

We adopted a direct forecast setting [also known as a ‘‘jump forecast’’ (Weigend 1995)]. This is done by projecting from the present to the desired point h steps in the future, without any intermediate feedback. This approach leads to a different network for a different lead time. The model design we used here is similar to our previous models (see Fig. 7 of Tangang et al. 1998), with the exception of using the EEOFs. We used the first seven modes of SLP EEOFs, which accounted for 53.2% of the total variance, plus the SSTA (i.e., the persistence) itself as inputs. We defined the lead time as the time between the end of the latest observed period and the beginning of the predictand period [the same definition as in Barnston et al. (1994)]. For example, to forecast the December–February (DJF) SSTA at 0-month lead, we used the SLP in the 12 preceding months [i.e., DJF, March–May (MAM), June–August (JJA), September–November (SON)] plus the persistence (i.e., the SSTA at SON). We used three neurons in the hidden layer, which resulted in a much smaller network than Tangang et al. (1997, 1998), with 31 degrees of freedom [i.e., $8 \times 3 + 3 + 3 + 1$ (d.f.)]. We applied two methods in estimating the forecast skills: the retroactive (Barnston et al. 1994) and the cross-validation (Michaelsen 1987) methods. As in Tangang et al. (1997, 1998), we implemented the retroactive method by training the net-

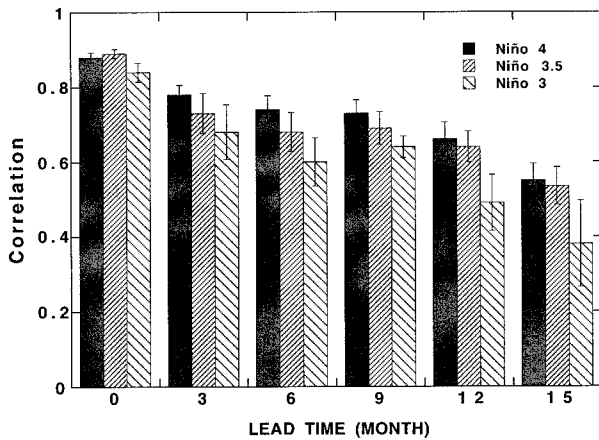


FIG. 5. The mean correlation skills for the forecast period of 1982–93, for Niño 4, Niño 3.5, and Niño 3 regions. The error bars indicate ± 1 standard deviation of the 20 forecasts.

works over the period of 1952–81 and used the subsequent period from 1982 to 1993 as forecast validation. Since the EEOF analysis was done over the entire record (1952–93), the forecast period (1982–93) does not represent a true out-of-sample period. However, our previous work, Tang (1995), showed that this approach did not significantly inflate the forecasting skills. For the cross-validation method, we withheld a group of 10 consecutive yr in turn from the record and trained the net-

works with the rest of the data. Forecasts were then made for the withheld period. Our results in Tangang et al. (1998) showed that withholding a group of 10, 5, or 3 consecutive yr produced similar skills.

4. Neural network forecasts

The mean correlation skills (averaged from 20 forecasts, each with different random initial weight initializations) of retroactive forecasts for the period from 1982 to 1993, for three regions, Niño 4, Niño 3.5, and Niño 3, are shown in Fig. 5, with the error bars indicating ± 1 standard deviation of the 20 forecasts. The western-central region (Niño 4 and Niño 3.5) appeared to be the best forecasted region, with Niño 4 having the highest skills. The eastern-central region of Niño 3 was less well forecasted, particularly for longer lead times.

Figure 6 shows the forecasts versus observations at 3-, 6-, 9-, and 12-month lead times for the Niño 4 region. These were ensemble forecasts (averaged from the 20 forecasts), and hence the correlation skills of the ensemble forecasts shown were not identical to the ensemble average of individual forecast skills shown in Fig. 5, but were 0.01–0.04 higher. These skills were significant at the 95% level, assuming 19 independent realizations of forecast skills over the 12-yr period. The assumption of 19 independent realizations was a rough approximation based on the decorrelation (autocorre-

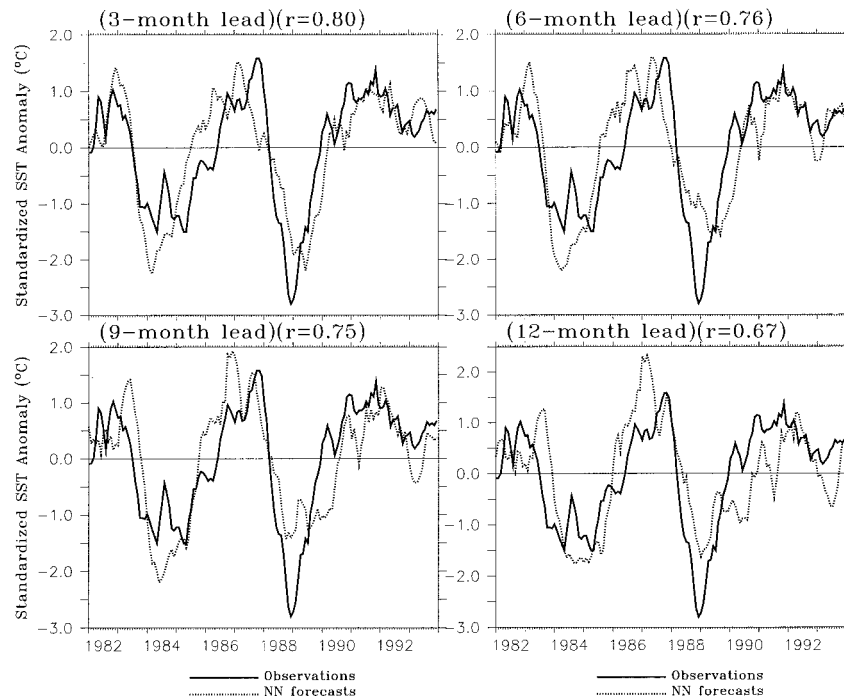


FIG. 6. Standardized SST anomalies of the neural net forecasts (dashed curves, averaged from 20 individual forecast) of Niño 4 region compared with the observed anomalies (solid curves) for the forecast period of 1982–93 at lead times of 3, 6, 9, and 12 months with r the correlation coefficient.

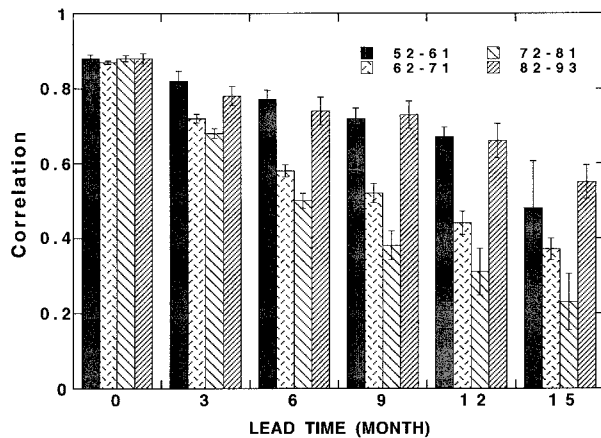


FIG. 7. The forecast correlation skills for various decades, corresponding roughly to the 1950s, 1960s, 1970s, and 1980s for the Niño 4 SSTA. The error bars indicate ± 1 standard deviation of the 20 forecasts.

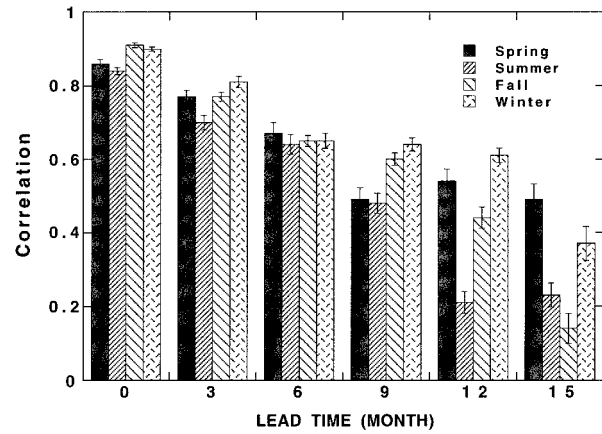


FIG. 9. The seasonal cross-validated correlation skills for the Niño 4 SSTA. The error bars indicate ± 1 standard deviation of the 20 forecasts.

lation e -folding) time of approximately 7.5 months of the observed SSTA. These forecasts resembled those presented in Tangang et al. (1998), with improvements over the period of 1990–92.

Figure 7 shows the correlation skills for forecasting various decades for the region of Niño 4. The skills showed strong decadal dependence, with higher skills in the 1950s and 1980s and lower in the 1960s and 1970s, similar to those in Tangang et al. (1998). Figure 8 shows the cross-validated correlation skills calculated over all decades for the three regions. The cross-validated skills were higher in the western-central region than in the eastern-central region. The cross-validated seasonal skills for the region of Niño 4 are shown in Fig. 9. The seasonality in the skills was apparent, with the winter season having the highest skills, which appeared useful up to 12-month lead. In general, for all regions, the smaller networks we adopted here had similar skills as the larger networks in Tangang et al. (1998).

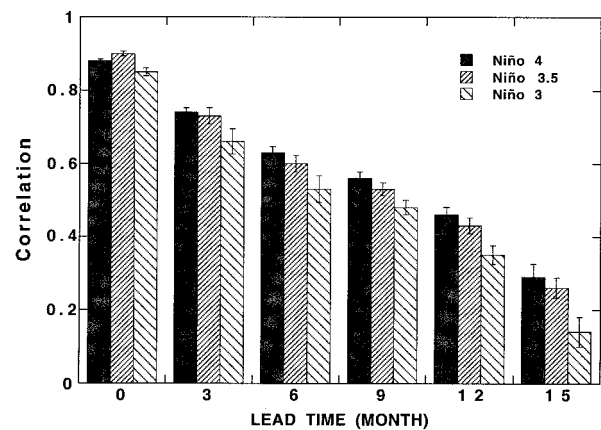


FIG. 8. The cross-validated correlation skills calculated over all decades for the Niño 4, Niño 3.5 and Niño 3 SSTA. The error bars indicate ± 1 standard deviation of the 20 forecasts.

5. Network pruning analysis

In Tangang et al. (1997, 1998), we used large networks (with 94 d.f.) with the EOFs as inputs. With such large networks, the weight interpretation can be very difficult. The smaller networks here (with 31 d.f.) produced comparable forecast skills and very consistent results. This shows that the lag compression via the EEOF analysis of the predictor data provides an effective way to reduce the size of the networks. With smaller networks, the interpretations of the weights are more lucid.

In addition to lag compression by EEOF, we also introduced a pruning scheme that further trimmed the networks to minimal size. To demonstrate this, we present the weights and biases of the hidden layer in Table 1, for the network forecasting Niño 4 SSTA at 6-month lead. This forecast was one of the 20 forecasts used to calculate the mean forecasts presented in the last section. The correlation skills for the training and forecast periods were 0.69 and 0.76, respectively, comparable to the mean skills. The corresponding weights, \bar{w}_j , for the output layer, were -0.3474 , -0.6110 , and -0.7401 for first, second, and third neurons, respectively. The bias, \bar{b} , for the neuron in the output layer was -0.0725 . If

TABLE 1. The weights, w , and biases b , of the hidden layer, for the network forecasting Niño 4 SSTA at 6-month lead.

Input	Hidden neuron		
	1	2	3
Mode 1	-0.0776	0.4946	-0.0363
Mode 2	-0.2230	0.0117	-0.2511
Mode 3	0.0527	0.1630	0.0244
Mode 4	0.2738	0.2255	-0.3460
Mode 5	-0.1287	-0.0057	-0.0894
Mode 6	-0.1426	0.1986	-0.4514
Mode 7	-0.1078	0.0200	0.0663
SSTA	-0.2085	-0.0776	-0.5923
Bias	0.6380	-0.4411	0.0188

TABLE 2. The pruning statistics for forecasting Niño 4 SSTA for the 1982–93 period at 6-month lead. The values shown are the correlation and rms between the outputs of the hidden neurons (N1, N2, and N3) of the pruned network and the corresponding outputs of the unpruned network. The correlation and rms values between the pruned network outputs and the corresponding unpruned outputs are given in the last two columns. Bolding in the output column indicates the important inputs.

Input pruned	Corr			Rms			Output	
	N1	N2	N3	N1	N2	N3	Corr	Rms
Mode 1	0.98	0.50	1.0	0.0697	0.4584	0.0292	0.91	0.2387
Mode 2	0.84	1.0	0.95	0.1942	0.0111	0.1943	0.92	0.2008
Mode 3	0.99	0.96	1.0	0.0428	0.1438	0.0157	0.98	0.1115
Mode 4	0.80	0.92	0.93	0.2187	0.1745	0.2434	0.99	0.0759
Mode 5	0.96	1.0	1.0	0.1034	0.0051	0.0585	0.99	0.0793
Mode 6	0.93	0.92	0.87	0.1353	0.1904	0.3260	0.93	0.1820
Mode 7	0.97	1.0	1.0	0.0895	0.0166	0.0449	1.0	0.0201
SSTA	0.98	1.0	0.96	0.0854	0.0322	0.1929	0.94	0.1890
Modes 3–7	0.72	0.90	0.92	0.2567	0.2224	0.2654	0.95	0.1589
Modes 1, 2, 6, and SSTA	0.50	0.62	0.66	0.3275	0.4634	0.5383	0.35	0.4596

one considers the magnitude of the weights, one may conclude that some of the inputs, for example, modes 5 and 7 of the SLP EEOF, may not contribute significantly. These weights, however, need to be interpreted with caution since the modeling strategy learned by the network cannot be readily understood by visually inspecting weights and biases. Different signal patterns may interact with the weight patterns in subtle ways; hence, the signals themselves must be included in the analysis. In particular, it is important to determine how the signals activate the hidden neurons. It is also crucial to consider how these neurons interact in the output layer, since the response for a particular input may be large for several neurons in the hidden layer, but the effects cancel out in the output layer.

To show this effect, we display the pruning statistics in Table 2 for the forecast period of 1982–93. Similar statistics were obtained for the training period. These statistics were obtained by pruning each input in turn (replacing it with zeros) and feeding the pruned inputs into the trained network. Equivalently, this can be achieved by setting the weights associated with that particular input to zero. We then calculated the correlation and rms values between outputs of each hidden neuron in the unpruned network and those of the pruned ones. We also calculated the same statistics for the final output neuron. These statistics provide some clues on the importance of each input in activating each neuron and how it affects the final outputs.

Both the correlation and rms values suggest that modes 2, 4, and 6 of the SLP EEOF gave relatively larger responses than the other inputs for the first hidden neuron (denoted as N1 in Table 2), as their removal led to a drop in the correlation and a rise in the rms value. The SSTA, however, produced a smaller response despite its larger weight than mode 6 (see Table 1) since the magnitude of the SSTA input was smaller than mode 6, and hence the overall response on the neuron activation was smaller for the SSTA than mode 6. The second neuron was dominated by mode 1, with modes 4 and 6 also showing relatively important contributions.

Modes 2, 4, and 6 of the SLP EEOF and the SSTA had impacts on the third neuron. It appeared that only five inputs (i.e., modes 1, 2, 4, 6, and the SSTA) made important contributions in activating the hidden neurons. The statistics for the output neuron (both the correlation and rms), however, showed that only four inputs,—that is, modes 1, 2, 6, and the SSTA—appeared to have relatively important contributions. The mode 4 input, despite larger responses on all three hidden neurons, produced negligible net effect on the final output. To explain this, we note that the weights for mode 4 (see Table 1) were 0.2738, 0.2255, and -0.3460 for the first, second, and third hidden neurons, respectively, while all the weights for the output layer (-0.3474 , -0.6110 , and -0.7401) were of the same sign. Such combination of weight patterns allowed the mode 4 signal to essentially cancel out in the final output layer. We further simultaneously omitted modes 3, 4, 5, and 7, which, when removed individually, produced no net effect on the network output. The net effect of removing these four inputs simultaneously appeared to be minimal, smaller than the effect of removing mode 1 alone. The network, however, completely failed when modes 1, 2, 6, and the SSTA were simultaneously removed from the inputs. We also calculated the means (based on 20 forecasts) for these pruning statistics and found similar results.

These results suggested that only four of the eight original inputs were important, and as modes 3, 4, 5, and 7 made no significant contribution, they could be ignored. To further investigate this, we retrained two new networks, with the first using modes 1, 2, 6, and the SSTA as inputs, and the other using modes 3, 4, 5, and 7. Figures 10a, b show the correlation skills for the training period (1952–81) and the forecast period (1982–93) for these two networks. For comparison, we also plotted the skills for the full model (i.e., with eight inputs). The skills of the first pruned model were slightly lower than the full model during the training period but were slightly higher than the full model during the forecast period. The slightly higher skills for the full model

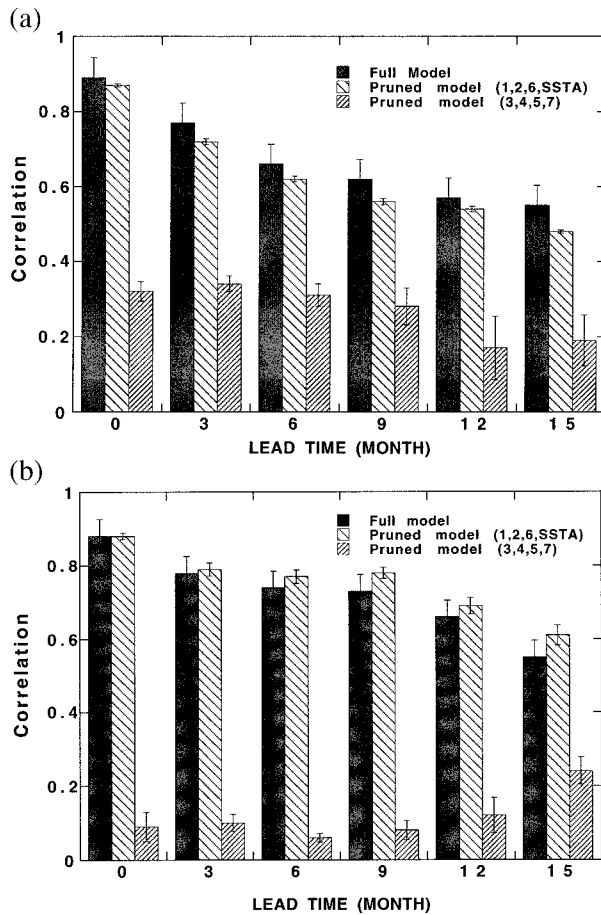


FIG. 10. The correlation skills for three different networks, the full model network trained with all the eight inputs; the network trained with modes 1, 2, 6 of the SLP EEOF and the SSTA; and the network trained with modes 3, 4, 5, and 7, for (a) the training period (1952–81) and (b) the forecast period (1982–93).

in the training period were expected since the full model had more free parameters to fit the data. The pruned model also showed less sensitivity to initial weights by having smaller standard deviations with ensemble forecasts. On the other hand, the model with modes 3, 4, 5, and 7 as inputs showed no skills in both the forecast and training periods. These further substantiate that modes 1, 2, 6, and the SSTA were the only relevant inputs, while the other inputs had minimal or no important contributions.

We further examined the pruned networks (with modes 1, 2, 6, and the SSTA) and determined how the three hidden neurons were being activated. We found that the outputs of one of the neurons were highly correlated with the networks outputs, whereas the other two neurons tended to cancel out each other. This seems to suggest that one neuron in the hidden layer would be sufficient. We then retrained another network with only a single neuron in the hidden layer. This represents a minimal network with four inputs, one hidden neuron, and one output neuron with a total of 7 degrees of free-

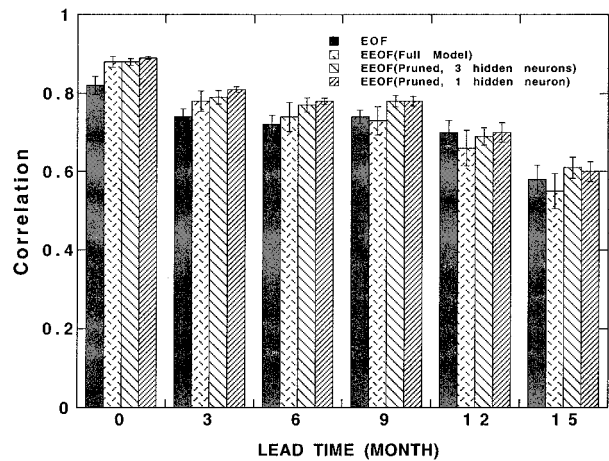


FIG. 11. The comparison between forecast skills of various neural network models for the 1982–93 period for the Niño 4 region. The pruned models have only SLP EEOF modes 1, 2, 6, and the SSTA as inputs, with either three or one hidden neurons.

dom. Figure 11 shows the forecast skills for the 1982–93 forecast period, for the Niño 4 SSTA. For comparison purposes, we also showed the forecast skills of the pruned model with three hidden neurons, the full model (all eight inputs), and those of larger networks using SLP EEOF as inputs presented in Tangang et al. (1998). The results showed that this minimal network gave a slight improvement of the forecast skills. The minimal networks also showed less sensitivity to the selection of the initial random weights. Such improvements were also seen for other decades as well as for other equatorial regions (not shown).

6. Spectral analysis of the trained network

The results in the previous section showed that only modes 1, 2, 6, and the SSTA were important. The SSTA (i.e., persistence) has been shown to improve the skills (Tangang et al. 1997; Barnston and Ropelewski 1992). In this section, we present spectral analysis, which provides supporting evidence on the importance of modes 1, 2, and 6. Figure 12 shows the spectra of modes 1, 2, and 6 of the SLP EEOF. The spectrum for mode 1 showed a peak at a 4–5-yr period. This agreed with our earlier interpretation that mode 1 was due to ENSO. Mode 2 had a broad spectrum with several peaks around 2–5 yr. These two modes were consistent with the LFO and the tropospheric QBO described in Barnett (1991), Ropelewski et al. (1992), and Goswami (1995). Mode 1, with a 4–5-yr period and a spatial pattern indicating a standing wavenumber 1 oscillation, fitted the description of the LFO. Mode 2 appeared to be a mixture of the QBO and the LFO, with an eastward propagating spatial pattern. Barnett (1991) concluded that the ENSO process is partially due to the nonlinear interaction between these two longer timescale oscillations. Mode 6 showed high energy peaks at low frequencies (10-yr

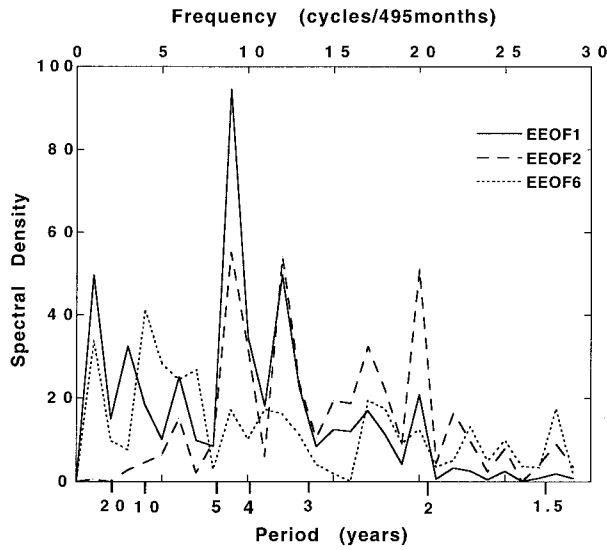


FIG. 12. The spectra for various modes of the SLP EEOF: mode 1 (solid), mode 2 (dashed), and mode 6 (dotted).

period or longer). Figures 13a,b show the spatial pattern (at lag 9) and the temporal pattern of mode 6, respectively. The spatial pattern indicated two opposite centers, a relatively strong center located in the central equatorial Pacific about 170°W, and a much weaker one in the Australian–Indonesia region. Also shown is the 6-yr low-pass-filtered time series (generated by succes-

sive applications of centered 25- and 37-month running means), clearly revealing the decadal oscillation. This filtered time series resembled the 6-yr low-pass-filtered time series of the SST over the Pacific Ocean shown in Zhang et al. (1997, Fig. 3).

While both modes 1 and 2 were expected to be important since both explained relatively large fractions of the total variance and both were associated with ENSO, it was surprising to see that mode 6, explaining only 2.7% of the total variance, played a significant role in forecasting the SSTA. Wang (1995) hypothesized that the interdecadal SSTA are coupled with the interdecadal atmospheric circulation anomalies, where the changes in the atmospheric circulation may in turn feed back into the ocean, effectively influencing the ENSO cycle. Wang (1995) also argued that the conditions during the warm state of the interdecadal mode favor the formation of positive SST anomalies stretching south-westward from Baja California to the central equatorial Pacific, weakening the trades in the central Pacific. Wang (1995, Fig. 12) summarized that there are two ways this condition may influence the ENSO onset: 1) The warming can offset the cool condition prior to an El Niño, resulting in a near-normal Walker circulation over the western and central Pacific and enhanced trades in the south-eastern Pacific. The strengthening of the southeast trades in the onset phase could prevent the South American coastal warming in the subsequent phase, resulting in an absence of coastal warming. Also, the warming

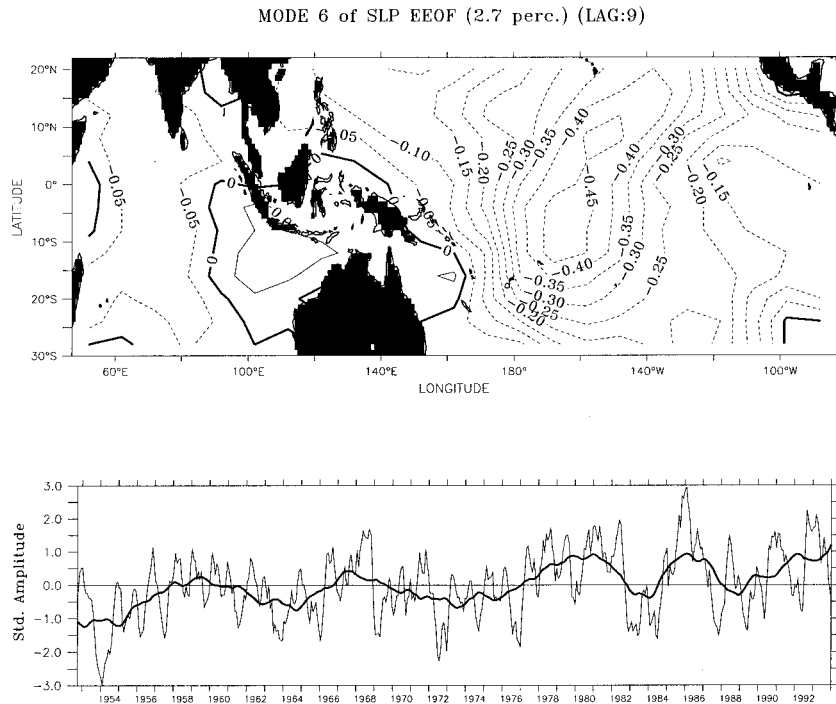


FIG. 13. (a) The spatial pattern for the SLP EEOF mode 6 at lag 9 months. The contours represent correlation between the amplitude and the SLP anomaly at each grid point. (b) The corresponding standardized amplitude. The thick line indicates the 6-yr low-passed time series.

TABLE 3. The weights, w , and biases, b , in several neural network models for forecasting the SSTA in various regions at 0-, 6-, and 12-month leads (the weights w_1 , w_2 , w_3 , and w_4 correspond to modes 1, 2, and 6 of the SLP EEOF and the SSTA persistence, respectively). The models were all trained with data from 1952 to 1981. Region P4 is at the eastern boundary (0° – 10° N, 80° – 100° W).

Region	Lead time														
	0-month					6-month					12-month				
	w_1	w_2	w_3	w_4	b	w_1	w_2	w_3	w_4	b	w_1	w_2	w_3	w_4	b
Niño 4	-0.06	0.18	0.08	0.98	0.15	-0.35	-0.25	0.31	0.89	-0.27	-0.15	0.08	0.22	0.10	-0.32
Niño 3.5	-0.07	0.15	0.06	0.63	-0.05	-0.28	0.27	0.27	0.40	-0.62	-0.22	0.08	0.27	0.05	-0.61
Niño 3	-0.06	-0.19	0.09	0.39	-0.47	-0.24	-0.40	0.30	0.00	-0.78	-0.38	0.07	0.46	-0.06	-1.13
P4	-0.07	0.26	0.11	0.76	-0.49	-0.46	-0.25	0.41	0.77	-0.83	-0.40	0.10	0.39	-0.02	-1.17

in the central equatorial Pacific (near the date line) triggers the eastward movement of convection and associated westerly anomalies from the western to the central equatorial Pacific, leading to a warming there. 2) The condition favors the formation of the Philippine sea onset cyclone, believed to be associated with the equatorial westerly anomalies, which eventually cause the Pacific warming. Despite the profound impacts of the 1976–77 “regime shift,” Zhang et al. (1997) speculated that the small shifts, like those in 1957–58 and 1990–91, are no less important in terms of climate assessment, diagnosis, and prediction. Wang (1995) also suggested that the anomaly models that are used to forecast ENSO should incorporate the secular changes in the background “mean” state. The results presented here, which showed the importance of the low frequency of the interdecadal oscillation in forecasting the SSTA, support this contention.

To further demonstrate the relative importance of each of the contributing inputs (i.e., modes 1, 2, 6, and the SSTA persistence) for forecasting the SSTA in different regions at different lead times, we present the weights and biases in Table 3. The weights and biases were obtained from training the minimal networks without the weight decay terms in (4). The networks converged to a unique solution for different random initializations. Since there is only one neuron in the hidden layer, the magnitude of the weight reflects its contribution. At 0-month lead, the SSTA and mode 2 had the most important contributions. The importance of persistence (as measured by w_4) declined eastward from Niño 4 to Niño 3 but then increased at P4 (0° – 10° N, 80° – 100° W). At 6-month lead, all three EEOF modes had roughly equal contributions, while the persistence (w_4) declined (except at P4). For a longer lead time of 12 months, only mode 1 and mode 6 had significant contributions. Also, there appeared to be a trend in which the biases increased in magnitude as the regions shifted eastward and lead time increased.

For the Taylor series expansion of the hyperbolic tangent function in (3) about $x = 0$, the linear, quadratic, and cubic parts of the expansion series are $\text{sech}^2(b)(\sum_i w_i x_i)$, $\tanh(b) \text{sech}^2(b)(\sum_i \sum_j w_i w_j x_i x_j)$, and $\text{sech}^2(b)(1 - 3 \tanh^2(b))(\sum_i \sum_j \sum_k w_i w_j w_k x_i x_j x_k)$, respectively. Clearly, the contribution of each term depends on the weights, the bias, and the inputs. Since the

weights in Table 3 are small, the quadratic and cubic parts are likely to be smaller than the linear ones. Also, as the bias approaches zero, there would be no quadratic component as $\tanh(b)$ vanishes. On the other hand, the cubic component will be suppressed at b around 0.61 for which the expression $[1 - 3 \tanh^2(b)]$ vanishes. For large values of b , both the quadratic and cubic components will contribute.

What kind of nonlinearity is embedded in the weights and biases of Table 3? We fed the networks with four pure sinusoidal functions, each with a different frequency. We used sine waves with frequencies of 40, 55, and 75 ω_o and amplitude 3 to substitute for mode 1 mode 2, and mode 6, respectively, where ω_o represents the lowest frequency, that is, $1024^{-1} \text{ month}^{-1}$. The amplitude 3 was chosen to match the maximum and minimum of the three original inputs that lie between +3 and -3. This is crucial since the networks would produce much stronger nonlinear responses if fed with inputs with much larger amplitude than those of the original ones. For the SSTA input, we used a frequency at 100 ω_o , and amplitude 2 to approximately match the magnitude of the SSTA input. These four frequencies were chosen so that the frequencies arising from quadratic interactions would not overlap the original input frequencies. We then input these four sine waves and ran spectral analysis on the outputs. If the network is linear, then only four spectral peaks will appear representing the four original frequencies. If the network is nonlinear, there will be additional peaks for the nonlinear interactions. As an example, Figs. 14a–c show the spectra of the neural network outputs for the Niño 3 region at 0-, 6-, and 12-month leads. At 0 months, most of the energy peaked at 100 and 55 ω_o corresponding to the SSTA and mode 2 inputs, with the other two inputs having relatively smaller peaks, in agreement with the weights in Table 3. Also seen was the interaction between the two dominant inputs, that is, at $45(=100 - 55)\omega_o$ and $155(=100 + 55)\omega_o$. At 6-month lead, all inputs but the SSTA had important contributions. Also, there appeared to be nonlinear interactions between these modes. The interaction between modes 1 and 2 may reflect the interaction between the QBO and LFO envisioned by Barnett (1991). At a longer lead of 12 months, only modes 1 and 6 appeared to be important, with strong nonlinear interaction among them mani-

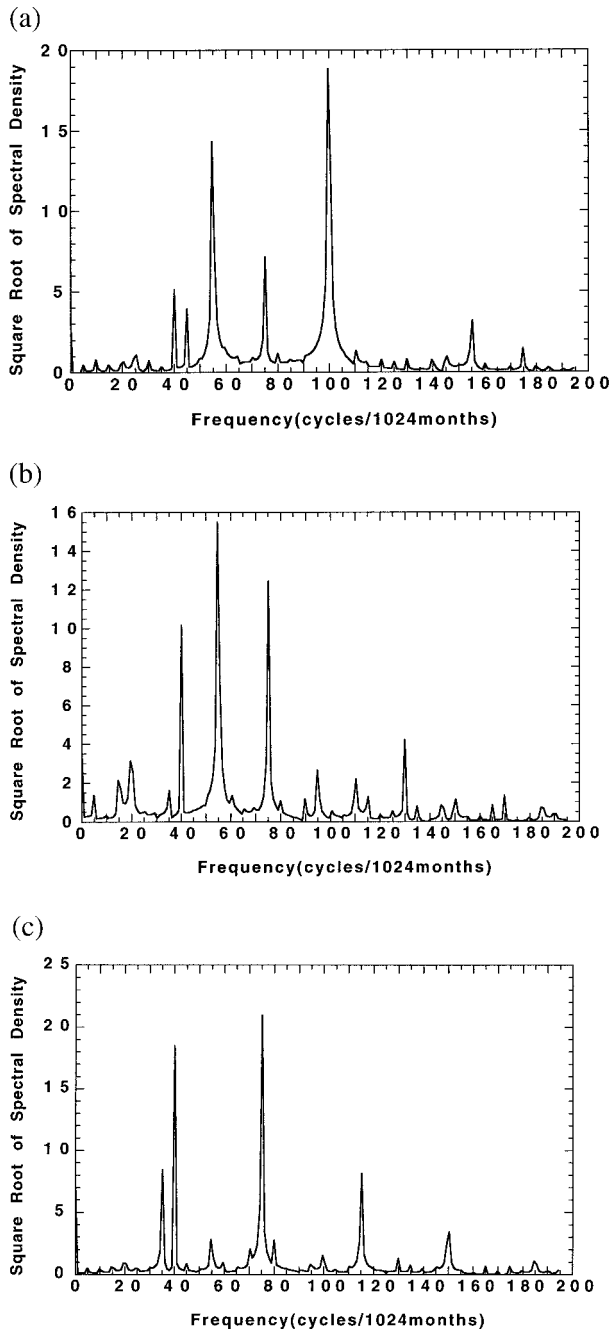


FIG. 14. The spectra for the neural network outputs for the Niño 3 region when fed with four pure sinusoidal signals at frequencies of 40, 55, 75, and 100 ω_o at lead times of (a) 0, (b) 6, and (c) 12 months. The square root of the spectral density was plotted to render the weaker signals more visible.

fed by the peaks at 35 and 115 ω_o . These results showed that the networks, fed with the original inputs, produced nonlinear responses. To roughly estimate the nonlinearity, we calculated the ratio between the power due to nonlinear interactions (i.e., total area underneath the quadratic and cubic peaks) and the total power (i.e.,

TABLE 4. The ratio (in percentage) of the power from nonlinear interactions to the total power for various regions at 0-, 6-, and 12-month leads. The models were all trained with data from 1952 to 1981.

Region	Lead time		
	0-month	6-month	12-month
Niño 4	9.4	12.2	3.9
Niño 3.5	4.5	9.0	10.4
Niño 3	9.4	14.9	22.4
P4	11.2	19.8	22.6

total area under all peaks). The results shown in Table 4 (in percentages) are for the Niño 4, Niño 3.5, Niño 3, and P4 regions at 0-, 6-, and 12-month leads, with region P4 at the eastern boundary (0° – 10° N, 80° – 100° W). East of Niño 4, there appeared a trend where the ratio increased eastward and with lead time. These results supported our earlier contention that the relationship between the predictor and the predictand became more nonlinear as lead time increased. Also, the nonlinearity appeared to be more prominent in the eastern region. The trend for the western-central region of Niño 4 was somewhat surprising, in that the ratio at 12-month lead was smaller than the ratio at 6-month lead.

7. Summary and conclusions

We applied the neural network technique to forecast the SSTA of the three regions, Niño 4, Niño 3.5, and Niño 3, representing the western-central, the central, and the eastern-central parts of the equatorial Pacific Ocean, respectively. Niño 4 appeared the best forecasted, with useful forecasting skills of up to 1-yr lead time. By using the EEOF technique to compress the SLP predictor field, we showed that much smaller neural networks produced similar or even slightly higher forecast skills to those of larger models using the SLP EOFs as inputs (as found in Tangang et al. 1998). In our earlier models, there were 94 degrees of freedom; by introducing the EEOF, we reduced the size to 31 d.f. and applying a pruning technique, we trimmed the models to only 7 d.f.

We discovered that only four of the eight original inputs (i.e., modes 1, 2, and 6 of the SLP EEOF, and the SSTA persistence) made important contributions to forecasting the SSTA, so the four insignificant inputs were eliminated. At a 0-month lead, mode 2 and persistence made the important contributions. At a 6-month lead, all four inputs made roughly comparable contributions, while at a 12-month lead, modes 1 and 6 were most important. Mode 1, with a 4–5-yr period and a spatial pattern indicating east–west seesaw oscillation, fit the description of the low-frequency oscillation (LFO). This mode is usually seen as the familiar ENSO signal with high amplitudes during El Niños and low ones during La Niñas. Mode 2 displayed an eastward propagating pattern with spectral peaks between 2–5 yr, indicating a mixture between the QBO and the LFO.

The importance of mode 6, characterized by decadal and interdecadal oscillations, showed that very long period oscillations affect ENSO forecasting.

By applying spectral analysis to neural networks, we were able to identify the important inputs and the nonlinear responses. We calculated the spectral power due to nonlinear interactions relative to the total power, which showed that east of Niño 4, nonlinearity appeared to increase with lead time, thus supporting our earlier contention that the relationship between the atmospheric field and the SST became increasingly nonlinear as the lead time increased. Nonlinearity also tended to grow stronger as one moved eastward from Niño 3.5.

Acknowledgments. This work was supported by the National University of Malaysia (F. Tangang) and by grants from the Natural Sciences and Engineering Research Council of Canada, and Environment Canada (W. Hsieh, B. Tang, and A. Monahan).

REFERENCES

- Barnett, T. P., 1984: Prediction of the El Niño of 1982–83. *Mon. Wea. Rev.*, **112**, 1043–1047.
- , 1985: Variations in near global sea level pressure. *J. Atmos. Sci.*, **42**, 478–501.
- , 1991: The interaction of multiple time series in the tropical climate system. *J. Climate*, **4**, 269–285.
- Barnston, A., and C. F. Ropelewski, 1992: Prediction of ENSO episodes using canonical correlation analysis. *J. Climate*, **5**, 1316–1345.
- , and Coauthors, 1994: Long-lead seasonal forecasts—Where do we stand? *Bull. Amer. Meteor. Soc.*, **75**, 2097–2114.
- Chauvin, Y., 1989: Generalization performance of overtrained back-propagation. *Proc. Eurasp Workshop on Neural Networks*. L. B. Almedia and C. J. Wellekens, Eds., Springer-Verlag, 46–55.
- 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.
- Elsner, J. B., and A. A. Tsonis, 1992: Nonlinear prediction, chaos, and noise. *Bull. Amer. Meteor. Soc.*, **73**, 49–60.
- , and —, 1993: Corrigendum to “Nonlinear prediction, chaos and noise.” *Bull. Amer. Meteor. Soc.*, **74**, 243.
- Goswami, B. N., 1995: A multiscale interaction model for the origin of the tropospheric QBO. *J. Climate*, **8**, 524–534.
- Graham, N. E., J. Michaelsen, and T. P. Barnett, 1987a: An investigation of the El Niño–Southern Oscillation cycle with statistical models, 1, predictor field characteristics. *J. Geophys. Res.*, **92** (C13), 14 251–14 270.
- , —, and —, 1987b: An investigation of the El Niño–Southern Oscillation cycle with statistical models, 2, model results. *J. Geophys. Res.*, **92** (C13), 14 271–14 290.
- Gutzler, D. S., and D. E. Harrison, 1987: The structure and evolution of seasonal wind anomalies over the near-equatorial eastern Indian and western Pacific Oceans. *Mon. Wea. Rev.*, **115**, 169–192.
- Hanson, S. J., and L. Y. Part, 1989: Comparing biases for minimal network construction with back-propagation. *Advances in Neural Information Processing Systems 1*, D. S. Touretzky, Ed., Morgan Kaufmann Publishers, 177–185.
- Jacobs, R. A., 1988: Increased rates of convergence through learning rate adaptation. *Neural Networks*, **1**, 295–308.
- Ji, C., and D. Psaltis, 1990: Generalizing smoothness constraints from discrete samples. *Neural Comput.*, **2** (2), 188–197.
- Michaelsen, J., 1987: Cross-validation in statistical climate forecast models. *J. Climate Appl. Meteor.*, **26**, 1589–1600.
- Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate*, **7**, 929–948.
- Ropelewski, C. F., M. S. Halpert, and X. Wang, 1992: Observed tropospheric biennial variability and its relationship to the Southern Oscillation. *J. Climate*, **5**, 594–614.
- Rumelhart, D. E., R. Durbin, R. Golden, and Y. Chauvin, 1995: Back-propagation: The basic theory. *Back Propagation: Theory, Architectures, and Applications*. Y. Chauvin and D. E. Rumelhart, Eds., Lawrence Erlbaum, 1–34.
- Smith, M., 1993: *Neural Networks for Statistical Modeling*. Van Nostrand Reinhold, 235 pp.
- Smith, T. M., R. W. Reynolds, R. E. Livezey, and D. C. Stokes, 1996: Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Climate*, **9**, 1403–1420.
- Tang, B., 1995: Periods of linear development of the ENSO cycle and POP forecast experiment. *J. Climate*, **8**, 582–691.
- , G. M. Flato, and G. Holloway, 1994: A study of Arctic sea ice and sea level pressure using POP and neural network methods. *Atmos.–Ocean*, **32**, 507–529.
- Tangang, F. T., W. W. Hsieh, and B. Tang, 1997: Forecasting the equatorial sea surface temperatures by neural network models. *Climate Dyn.*, **13**, 135–147.
- , —, and —, 1998: Forecasting regional sea surface temperatures of the tropical Pacific by neural network models, with wind stress and sea level pressure as predictors. *J. Geophys. Res.*, in press.
- Wang, B., 1995: Interdecadal changes in EL Niño onset in the last four decades. *J. Climate*, **8**, 267–285.
- Weare, B. C., and J. S. Nasstrom, 1982: Examples of extended empirical orthogonal functions. *Mon. Wea. Rev.*, **110**, 481–485.
- Weigend, A. S., 1996: Time series analysis and prediction. *Mathematical Perspectives on Neural Networks*, P. Smolensky, M. C. Mozer, and D. E. Rumelhart, Eds., Lawrence Erlbaum, 395–449.
- , D. E. Rumelhart, and B. A. Hubermann, 1990: Predicting the future: A connectionist approach. *Int. J. Neur. Sys.*, **1**, 193–209.
- , —, and —, 1991: Generalization by weight-elimination with application to forecasting. *Advances in Neural Information Processing System 3*, R. Lippmann, J. E. Moody, and D. S. Touretzky, Eds., Morgan Kaufmann Publishers, 875–882.
- Woodruff, S. D., R. J. Slutz, R. L. Jenne, and P. M. Steurer, 1987: A comprehensive ocean–atmosphere data set. *Bull. Amer. Meteor. Soc.*, **68**, 1239–1250.
- , S. J. Lubker, K. Wolter, S. J. Worley, and J. D. Elms, 1993: Comprehensive ocean-atmosphere dataset (COADS) release 1a: 1980–92. *Earth Syst. Monitor*, **4** (1), 1–8.
- Yasunari, T., 1990: Impact of Indian monsoon on the coupled atmosphere/ocean system in the tropical Pacific. *Meteor. Atmos. Phys.*, **44**, 29–41.
- Zhang, Y., J. M. Wallace, and D. S. Battisti, 1997: ENSO-like interdecadal variability: 1900–93. *J. Climate*, **10**, 1004–1020.