Climate Dynamics (1997) 13: 135–147

Climate Dynamics

Forecasting the equatorial Pacific sea surface temperatures by neural network models

F. T. Tangang, W. W. Hsieh, B. Tang

Oceanography, Dept. of Earth and Ocean Sciences, University of British Columbia, Vancouver, B.C., Canada V6T 1Z4

Received: 31 October 1995/Accepted: 25 July 1996

Abstract. We used neural network models to seasonally forecast the tropical Pacific sea surface temperature anomalies (SSTA) in the Niño 3.4 region (6 °S–6 °N, 120 °W–170 °W). The inputs to the neural networks (i.e., the predictors) were the first seven wind stress empirical orthogonal function (EOF) modes of the tropical Pacific (20 °S–20 °N, 120 °E–70 °W) for four seasons and the Niño 3.4 SSTA itself for the final season. The period of 1952–1981 was used for training the neural network models, and the period 1982–1992 for forecast validation. At 6-month lead time, neural networks attained forecast skills comparable to the other El Niño-Southern Oscillation (ENSO) models. Our results suggested that neural network models were viable for ENSO forecasting even at longer lead times of 9 to 12 months. We hypothesized that at these longer leads, the underlying relationship between the wind stress and Niño 3.4 SSTA became increasingly nonlinear. The neural network results were interpreted in light of current theories, e.g., the role of the "off-equatorial" Rossby waves in triggering the onset of an ENSO event and the delayed-oscillator theory in the development and termination of an ENSO event.

1 Introduction

On the interannual time scale, the E1 Niño-Southern Oscillation (ENSO) phenomenon is the most important short-term climatic fluctuation. Recent survey by Barnston et al. (1994) shows that there have been several ENSO prediction schemes developed and routinely used. These schemes can be divided into 4 distinct classes: (a) simple dynamical coupled models, e.g., at the Lamont-Doherty Earth Observatory (Cane et al. 1986; Cane and Zebiak 1987; Zebiak and Cane 1987), (b) coupled GCMs (general circulation models), e.g., at the National Center for Environmental Predictions (NCEP) (Ji et al. 1994a, b), (c) hybrid coupled models, where a simple empirical atmosphere is coupled to a dynamic ocean model, e.g., Barnett et al. (1993), and (d) statistical models, e.g., canonical correlation analysis (CCA) (Barnston and Ropelewski 1992), and the constructed analog method (Van den Dool 1994). Another linear statistical model that has been used to predict ENSO is the POP (principal oscillation pattern) analysis (von Storch et al. 1995).

As our physical understanding increases, dynamical models have the potential to give more skillful forecasts than purely statistical models. Currently however, there is no significant difference between the performances of these four types of models (Barnston et al. 1994). For all four types of models, the equatorial Pacific sea surface temperature (SST) forecasts at 6-month lead time have an overall correlation skill in the 0.60s for the 1982–1993 forecast period, which was considered to be much better than persistence forecasts. As lead time increases further, the dynamical models appear to have higher skills than the statistical models (Barnston and Ropelewski 1992; Barnston et al. 1994). The lower skills of these statistical models could be attributed to the absence of physics in building the models and the inability to model nonlinear relationships. In this study, the potential of nonlinear statistical modeling will be explored by using the neural network modeling technique.

Artificial neural networks are computing systems containing many simple nonlinear computing nodes (or neurons) interconnected by links. This architecture is designed to simulate to some extent the structure and function of the brain (Rumelhart et al. 1986). If trained properly, a neural network model is capable of 'learning' the linear as well as the nonlinear features in the data. Elsner and Tsonis (1992) demonstrated that neural networks could perform far better than linear statistical models when dealing with chaotic plus random noise systems, as the neural networks were able to distinguish between the seemingly random but deterministic behavior of chaos from random noise. Despite the wide use of neural networks in areas such as artificial intelligence, engineering, physics, pattern recognition, optimization and financial forecasting, the application of neural networks to climate

Correspondence to: W. W. Hsieh
e-mail: william@ocgy.ubc.ca

forecasting is still at the infancy stage. Recent studies by Tang et al. (1994), Grieger and Latif (1994) and Navone and Ceccatto (1994) have shown positive development towards their application in short-term climate prediction. In Tang et al. (1994), the ENSO forecasts involved only the wind stress field. In this study, neural networks will be applied to SST prediction in the equatorial Pacific, and compared with the ENSO forecast models described in Barnston et al. (1994).

The rest of the paper is organized as follows: Sect. 2 describes the data used, while Sect. 3 presents the neural network model. The forecast results are given in Sect. 4, the several sensitivity studies in Sect. 5, and the role of SST as a predictor in Sect. 6.

2 The data

The data used in this study, covering the period from 1952 to 1992, came from the Comprehensive Ocean-Atmosphere Data Set (COADS) (Woodruff et al. 1987; 1993). The COADS monthly data were available in 2° latitude by 2° longitude grids.

The tropical Pacific sea surface temperature (SST) is one of the major players in the development of the ENSO phenomenon. In this study, attention will be focused on the eastern–central region of the tropical Pacific (6 °S –6 °N, 120 °W–170 °W). This study region (Fig. 1) was referred to as Niño 3.4 in Barnston et al. (1994) (where the latitudinal extent is actually 5 °S – 5 °N). The monthly SST data in this region were obtained by spatially averaging the SST in the $2^\circ \times 2^\circ$ boxes. To obtain the SSTA, we subtracted the monthly SST data by the climatological monthly means which were calculated based on the entire record of the data set (1952–1992). These anomalies were then smoothed in time by the 3-point running average.

The wind stress field covering the region of 20 °S –20 °N, 120 °E–70 °W was calculated based on the drag coefficient formula given in Trenberth et al. (1989) and Large and Pond (1981). The zonal and meridional components of the wind speed also came from COADS. This region is wide enough to include the equatorial Kelvin and the off-equatorial Rossby wave activities that could be associated with the ENSO phenomenon. The wind stress data were further averaged from 2° by 2° grids to 2° latitude by 6° longitude grids. The traditional empirical orthogonal function (EOF) analysis (Preisendorfer 1988) was applied to the wind stress anomaly fields. The EOFs were calculated from the covariance matrix

$$\mathbf{C} = \mathbf{X}'\,\mathbf{X}, \text{ with } \mathbf{X} = \begin{bmatrix} \tau_u \\ \tau_v \end{bmatrix} \quad (1)$$

where τ_u and τ_v are the zonal and meridional components of the wind stress anomaly respectively. We used the dimensionally smaller covariance matrix as implemented in Lagerleof and Bernstein (1988). The resulting temporal coefficients of the EOFs were also smoothed in time using 3-point average.

Figures 1 and 2 showed the spatial pattern and the temporal coefficients of our first and second modes. Each vector in the maps corresponds to the correlation values between the amplitude (i.e., temporal coefficient) and the zonal and meridional components of the wind stress anomaly fields. The spatial and temporal patterns of the first mode were similar to those in Legler (1983). This mode accounted for 4.1% of the total variance as compared to 6.8% in the Legler analysis, where the seasonal cycle was not removed. The temporal pattern of this mode

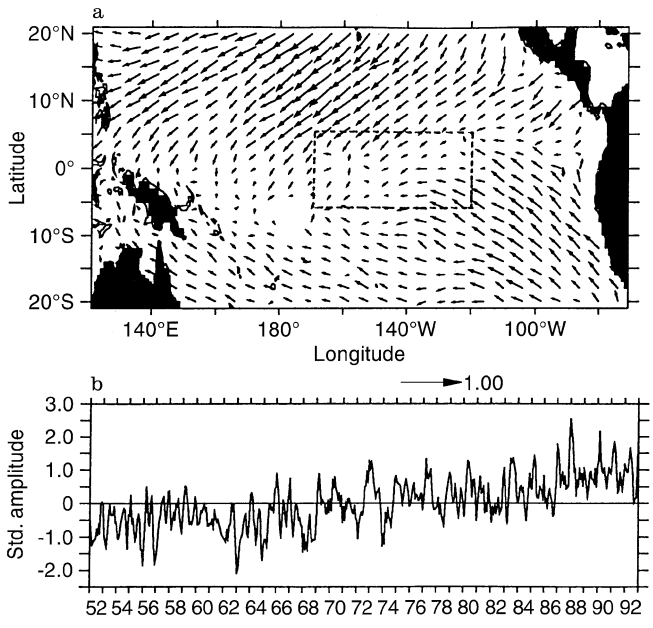

Fig. 1a, b. First EOF mode of the wind stress field. The *dashed box* centered around the equator indicates the Niño 3.4 region, for which we would make SST forecasts. The amplitude has been divided by the standard deviation

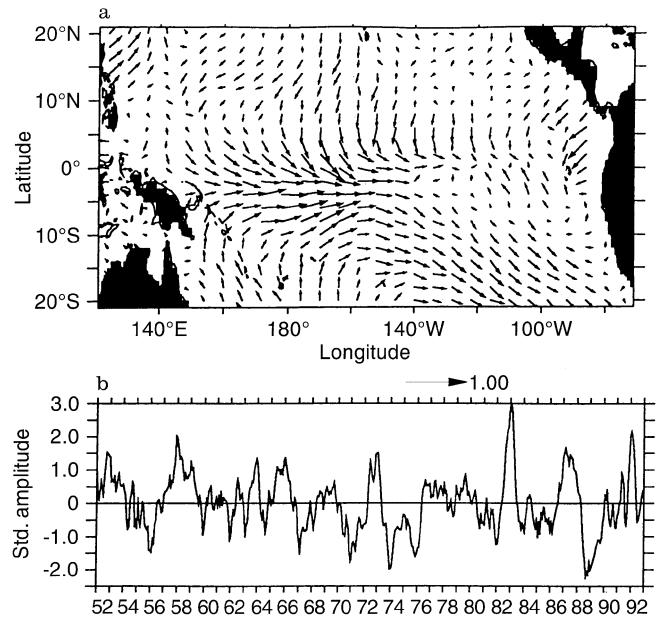

Fig. 2a, b. Second EOF mode of the wind stress field. That this mode is strongly related to ENSO can be seen by its amplitude, which rises sharply during E1 Niño years

showed an increasing trend in the post-1977 period. While there have been conflicting arguments in recent years on whether the increasing trend in the trade winds or the SST in the Pacific ocean was real (Posmentier et al. 1989; Cardone et al. 1990), recent results of Wang (1995) seemed to indicate that this trend was real and important in characterizing the pre- and post-1977 ENSO events. The second mode with 3.0% of the variance had a spatial pattern which indicated large anomalous westerly winds along the equator in the western–central region of the tropical Pacific, a feature associated with the onset of E1 Niño. The temporal coefficient also clearly indicated ENSO connection, i.e., it was high during E1 Niño and low during La Niña. These patterns resembled the third mode of Legler (1983), (which accounted for 4.8% of the variance), and also the first mode of Graham et al. (1987a), where the wind stress covered the period of 1947–1984 and the region of 30 °S – 30 °N, 30 °E – 70 °W, averaged to 2° latitude by 10° longitude grids. Our third mode with 2.4% of the variance also showed some E1 Niño connection. The fourth and higher modes had no obvious physical interpretations.

3 Neural network models

The feed-forward neural network is the most commonly used neural network model. This network has one receiving input layer, one broadcasting output layer, and one or more hidden (intermediate) layers in between. Fig. 3 shows an example of such an architecture with 4 neurons in the input layer, 3 in the hidden layer and 1 in the output layer. The number of neurons in the input and output layers depend on the number of predictors and predictands respectively. On the other hand, the number of hidden layers and the number of neurons in each hidden layer depend on the complexity of the problem. Given the right internal representation, this architecture can perform any arbitrary mapping of an n-dimensional input to an m-dimensional output. In fact, to approximate any mapping function to a given accuracy, at most two hidden layers are needed, with arbitrary accuracy attainable given enough units per layer (Cybenko 1988). If the mapping function is a continuous function, then only one hidden layer is needed (Cybenko 1989; Hornik et al. 1989; Funahashi 1989).

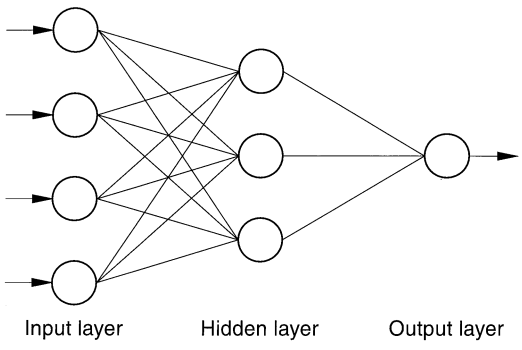

Fig. 3. An example of a neural network model, where there are 4 neurons in the input layer, 3 in the hidden layer, and 1 in the output layer

To properly relate the inputs to the desired outputs, the strength of the connection between neurons, i.e., the weights, need to be adjusted. The iterative adjustment process is often referred to as network training, and a standard algorithm is the "back-propagation" rule (Rumelhart et al. 1986; Smith 1993). In this study, we adopted an algorithm of steepest descent with momentum and adaptive learning rates (Jacobs 1988). This is a minimization technique based on a gradient-descent algorithm in which the error between the data and the network output is reduced by modifying the weights, proceeding backwards from the output layer to the input layer connections. The training process begins by loading the values of the predictors on the input layer. No calculation is performed in this layer but each neuron distributes a copy of its input to all neurons in the hidden layer. Each neuron in this hidden layer in turn performs the task of summing the weighted inputs and sending this sum through a transfer function onto the next layer of neurons.

If x_i, y_j and z denote the values of the neurons in the input, the hidden and the output layer, then we let

$$y_j = \frac{2}{1 + \exp\left[2\left(\sum_i - w_{ji}x_i - c_j\right)\right]} - 1, \tag{2}$$

where the transfer function is a tangent hyperbolic function, the shape of which is determined by the weights w and c.

For our output neuron, we let

$$z = \sum_j \tilde{w}_j\, y_j + \tilde{c}. \tag{3}$$

How closely the network outputs z match the target observations z_0 is measured by the SSE, i.e., the sum of the squared error between z and z_0 over the training data record. The SSE could serve as the cost function to be minimized. Following Chauvin (1989), the cost function of the minimization is defined as

$$C = \frac{a_1}{2} \sum (z - z_0)^2 + a_2\, a_4^2 \sum \frac{(w/a_4)^2}{1 + (w/a_4)^2} + a_3\, a_4^2 \sum \frac{(\tilde{w}/a_4)^2}{1 + (\tilde{w}/a_4)^2}, \tag{4}$$

where the first term on the RHS is basically the SSE, and the last two terms penalize excess weights (Chauvin 1989; Hanson and Part 1989; Ji and Psaltis 1990; Weigend et al. 1991). The a_i's are weighting parameters to be specified. The values of these parameters are chosen so that the networks 'excess' capability is contained and overfitting is prevented. Despite the complicated appearance of (4), as a_4 becomes large, (4) reduces to the simple formula

$$C = \frac{a_1}{2} \sum (z - z_0)^2 + a_2 \sum w^2 + a_3 \sum \tilde{w}^2, \tag{5}$$

similar to that used in Tang et al. (1994).

At the beginning of the training process, the weights w, $\tilde{w}$ and $\tilde{c}$ were initialized based on the Nguyen–Widrow (NW) scheme (Nguyen and Widrow 1990). This scheme

essentially sets the initial weights of a layer randomly so that each neuron in the layer is assigned its weight within a prescribed interval at the start of the training. With this method of assignment, the weight movement is likely to be small during the optimization search compared to the conventional method, thereby requiring less training. During the training process, these weights were iteratively adjusted until the cost function is minimized. Generally, a successful training allows the network to capture the essential relationship between the predictor and the predictand. In such cases, the trained network shows remarkable generalization capability, being able to correctly relate inputs to outputs not included in the training set. However, the size of the architecture (i.e., the number of hidden neurons) is the crucial element that needs to be considered. A smaller-than-needed network does not learn the examples well, while a bigger-than-required one usually overfits the data, i.e. learning undesirable noisy features in the data, hence the need for the penalty terms in the cost function to prune excess weights. The data set was divided into two parts; a training set and a validation set. A neural net model was trained with data from the training set, and the model forecasts were validated with the validation set.

In this study, we attempted to simulate real-time forecasting, where estimating the forecast skill would be an issue. According to Barnston et al. (1994), a purely objective and bias free method for estimating the forecast skill of any given method in a truly independent (future) forecast sitting does not exist. There are two common methods for estimating forecast skills: the retroactive method and the cross-validation method. For the cross-validation method (Michaelsen 1987), the forecast skill is estimated by withholding the data each year in turn from the model's developmental sample and making forecasts for the withheld data period. With this method, skill inflation can still occur when there are interannual autocorrelations in the data, as in ENSO-related time series (Barnston et al. 1994). This problem can be overcome by withholding groups of consecutive years for each set of forecast trial as suggested by Barnston et al. (1994). Skill deflation can also occur when the true skill is low, due to a degeneracy inherent in cross validation (Barnston and Van den Dool 1993). In the retroactive set-up, no data occurring later than the forecast time are made available to the model. According to Barnston et al. (1994), this is perhaps more realistic in simulating real-time forecasts. In this study, our approach will be based on the retroactive real-time forecast method where we trained the networks with the record of 1952–1981 and made the forecasts during 1982–1992. We varied the lead time from 0 to 15 months [our definition of lead time follows that of Barnston and Ropelewski (1992) and Barnston et al. (1994), i.e., the time between the end of the latest observed period and beginning of the predictand period].

The use of low-frequency variations in the atmosphere, e.g. the SLP or wind stress, to predict the SST in the tropical Pacific has been demonstrated in Graham et al. (1987b) and Barnston and Ropeleweski (1992). In this study, the wind stress will be nonlinearly related to the SST by the neural network model. As indicated in Graham et al. (1987b), the timing and duration of the wind stress to be included in the predictor seem to be crucial. Their time scale of a 6-month block of wind data was intended to capture the equatorial Kelvin and reflected Rossby wave activities associated with ENSO events. Barnston and Ropelewski (1992) used one year of SLP in the predictor. In an attempt to estimate the best timing, we have experimented with several different durations of the wind stress in the predictor, to be discussed later in Sect. 5. The results in the following section were based on the models using 12 months of wind stress plus the Niño3.4 SSTA itself. Using the first seven wind stress EOFs, we have 29 (i.e., $4 \times 7 + 1$) inputs in the models. We used three neurons in the hidden layer. The forecast settings, similar to that of Barnston and Ropelewski (1992), are illustrated in Fig. 4 for forecasting the onset, peak and mature phases of the 1982/83, 1986/87 and 1991/92 events. We adopted the terminology of Rasmusson and Carpenter (1982) in refering to these phases.

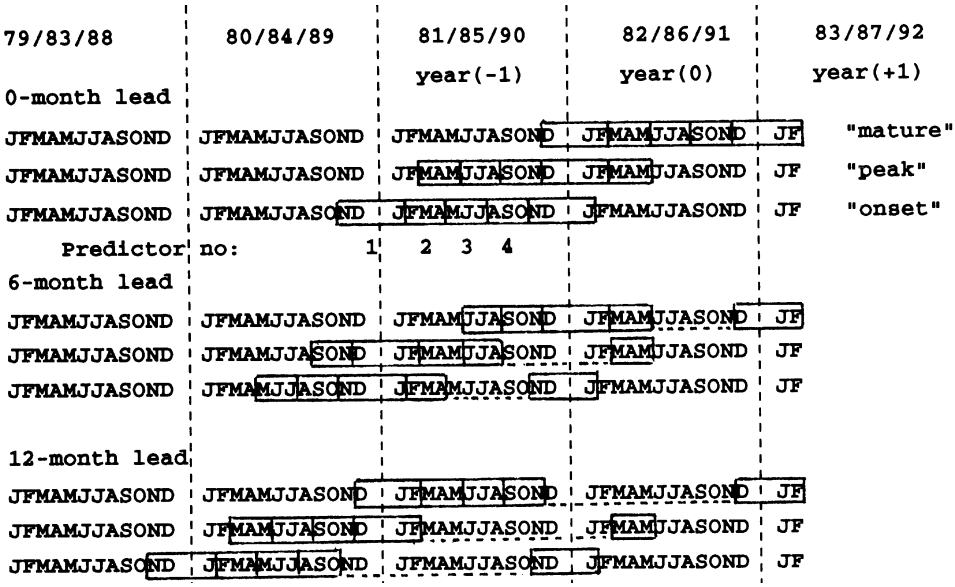

Fig. 4. Schematic diagram showing the timing of the predictor – predictand periods for the 0, 6, and 12-month leads for forecasting the onset, peak and mature phases of the 1982/83, 1986/87 and 1991/92 events

4 Results

4.1 Forecasts with neural network models

The skills of the retroactive forecasts for 1982–1992 are shown in Fig. 5. The three curves represent the control i.e., persistence forecasts, the individual forecasts and the ensemble forecasts. The ensemble forecasts were calculated by averaging six individual forecasts. For example, the ensemble forecast for the DJF period of 1982 at 3-month lead was averaged from the individual forecast of the same period at 3-month lead and those at 4, 5, 6, 7, and 8-month leads. As indicated in Fig. 5 this method of averaging tended to improve the skills in the medium range (5 to 10-month) but to degrade the skills for forecasts of very short leads and those longer than a year. At 6-month lead, the correlation was 0.66 and the root-mean-square error (rmse) was 0.83 for the individual forecasts. The ensemble forecasts appeared to be better with a correlation of 0.71 and an rmse of 0.73. These values are comparable to those of existing models described in Barnston et al. (1994), where the correlations were in the 0.60s and the rmse greater than 0.80. As lead time increased beyond 6 months, the forecast skill showed stability and minimal decay up to a lead time of a year, with the correlation skills greater than 0.6 and rsme values between 0.83–0.85. Also, the correlation at 12-month lead was 0.65, which seemed to suggest an increase of skill from that of the 9-month lead for the individual forecasts. This trend agreed with that of a hybrid coupled model shown in Fig. 6 of Latif et al. (1994). Their forecast skills, which were based on the period 1976–1990, clearly showed an increase as lead time approached a year. Beyond 12-month lead, the neural network forecast skills resumed its decline. At 15-month lead, the correlation skills for the individual forecasts was 0.57.

Figure 6 compares the correlation between the neural network outputs and the observations over the training period (1952–1981) with the correlation over the forecast validation period (1982–1992). The correlations over the training period were higher for shorter lead times, and decayed as lead time increased. Beyond 9-month the fitting seemed to improve. The correlations over the validation period appeared to be a little higher than over the training period for lead time 3 months or shorter. As lead time increased beyond 3-month lead, the training skills were higher than the forecast skills.

Figure 7 compares forecasts with observations at 3, 6, 9 and 12-month leads. The individual forecasts were somewhat noisy, similar to the forecasts by the CCA and analog models discussed in Barnston et al. (1994). By ensemble averaging the forecasts, we removed the high frequency noise, improving the skills especially for lead times of 6 and 9 months. At 3-month lead (Fig. 7a), the model picked up the main events of the 1982/83, 1986/87 and the 1991/92 El Niños and the 1988 La Niña. The ensemble forecasts gave a rather small amplitude for the 1982/83 El Niño, while both individual and ensemble forecasts were late for this warm event. Barnston and Ropelewski (1992, Fig. 11) showed similar problems for this warm event. The onset of the 1986/87 and 1991/92 El Niños were well forecasted. The model however failed to forecast the second peak of the 1986/87 El Niño. The 1988 La Niña was also well forecasted except for a shallower trough. The spin-down of the events were reasonably well forecasted for the 1982/83 and 1986/87 events, with a delay of a few months for the 1991/92 event. As for the individual forecasts there were two obvious false

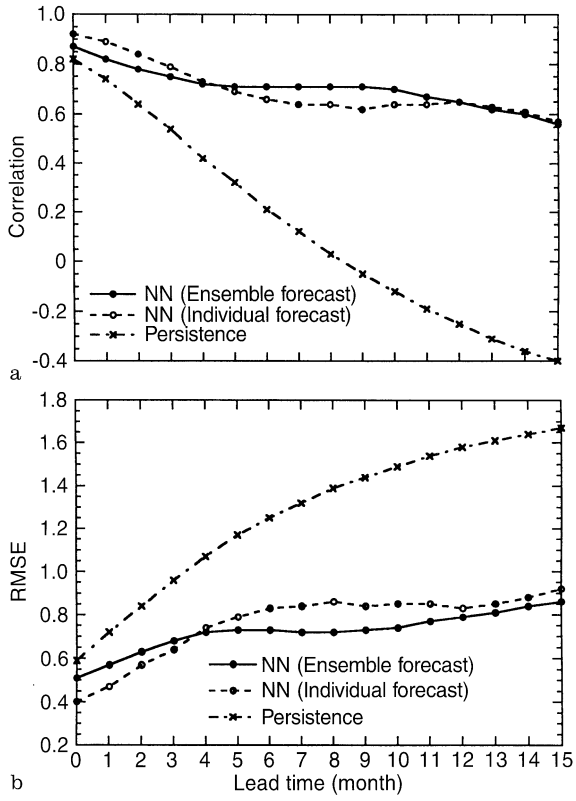

Fig. 5. **a** Correlation and **b** root-mean-square error (rmse) between neural network model forecasts (individual and ensemble) and observations for the Niño3.4 SSTA during 1982–92 for various lead times. Persistence is shown as the *dot-dash* curve. The rmse was calculated for standardized values (i.e., values inflated to zero mean and normalized to their standard deviations)

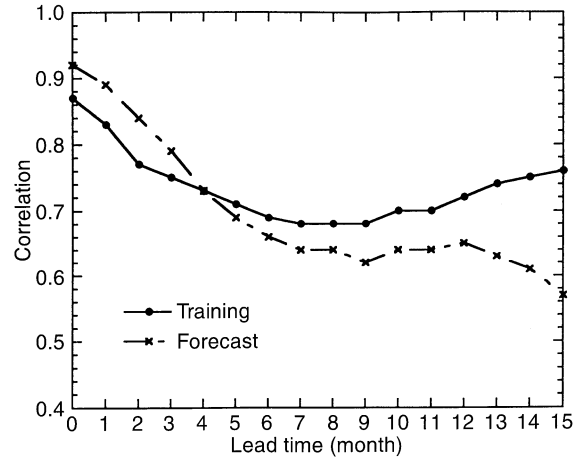

Fig. 6. A comparison of correlation skills of the neural network model for Niño3.4 SSTA during the training period (1952–1981) and the forecast validation period (1982–1992) at various lead times

alarms in the forecasts: (1) the model falsely predicted a cold event in the summer of 1984, and (2) a high peak with a duration of several months in 1990. The incorrect forecast for a cold event in the summer of 1984 is similar to that made by the CCA as noted by Barnston and Ropelewski (1992, Fig. 11), who argued that their CCA model "learned" only from the predictor-predictand relationship over the provided record and "performed" what an empirical (and linearly) minded human might do. As neural network model forecast resembled the CCA forecast, this supported our contention (to be discussed later) that for short lead times the neural network model "learned" only the linear relationship, which appeared to be the optimal solution. The high peak in 1990 was not described in Barnston and Ropelewski (1992) since their data ended in 1989. It was however clearly visible in the analog model for forecasts at 6-month lead as described in Barnston et al. (1994).

For 6-month lead (Fig. 7b), the neural network model forecasted the 1986/87 and 1991/92 El Niño rather well, but underestimated the 1982/83 El Niño, with the forecasted amplitude half that of the observed plus a delay of onset. The CCA model discussed in Barnston et al. (1994, Fig. 2d) performed better as the model forecasted the phase rather well, but the amplitude was also about half of the observed. The peak of the 1991/92 El Niño was better forecasted by the neural network model as compared to the CCA. In contrast, the amplitude of the 1988 La Niña was not as well forecasted by our neural network model, although there was a slight delay in the decay of the 1986/87 El Niño in the CCA forecast. Sometime in the summer or fall of 1984, our model also incorrectly forecasted a cold event. The CCA model also predicted a relatively high peak in 1985 which did not appear in our forecast. The differences between the CCA and the neural network forecasts could be due to: (1) the models used different predictor sets, and (2) the CCA searched for linear combinations of the predictor and the predictand data fields with the highest correlation, while the neural network model searched for the best linear or non-linear relationship between the predictors and the predictand.

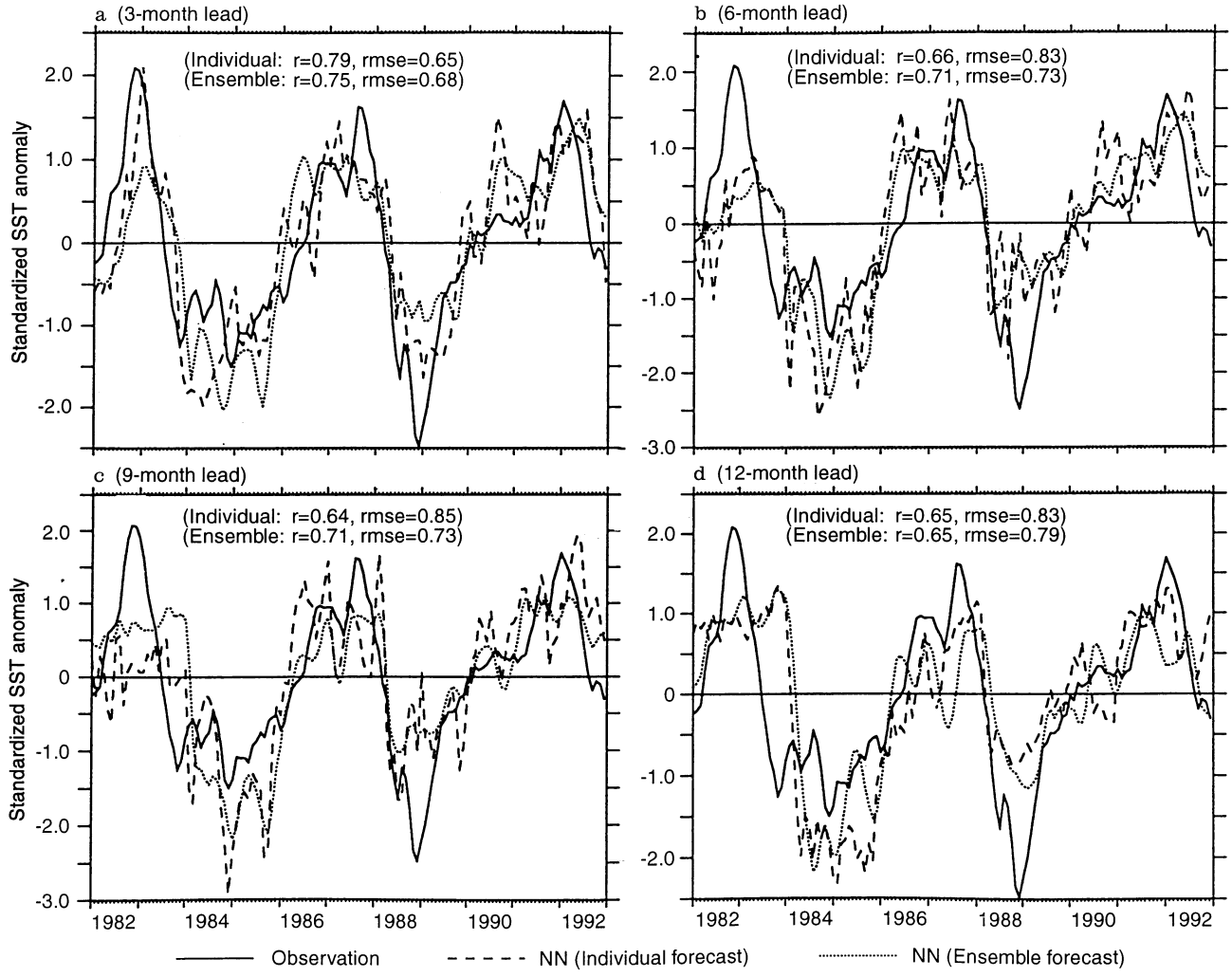

Fig. 7a–d. Standardized SST anomalies of the neural net forecasts (*dashed curves* for the individual forecasts, *dotted curves* for the ensemble forecasts) of the Niño 3.4 SSTA compared with the observed anomalies (*solid curves*) during 1982–1992 at lead times of **a** 3, **b** 6, **c** 9 and **d** 12 months

At 9-month lead (Fig. 7c), the individual forecast skills were marginally worse than those at 6-month lead, while the ensemble forecasts had correlation and rmse values identical to those at 6-month lead. A direct comparison cannot be made between the CCA and our neural network model as no details of the CCA forecasts at 9-month lead and beyond were given in Barnston et al. (1994). The results of Barnston and Ropelewski (1992) and Graham et al. (1987b) seemed to indicate low skills for the CCA method at 9-month lead or beyond.

At 12-month lead (Fig. 7d), the ensemble forecasts did not improve on the individual ones since the individual forecasts degraded rapidly beyond 15-month lead. Overall, our results seemed to indicate that a neural network model is viable for forecasting the ENSO phenomenon with reliable forecasts up to a year ahead. At 6-month lead, the neural networks models performed comparably to the existing models discussed in Barnston et al. (1994). The model performances at 9 and 12-month leads were rather encouraging since previous studies of Barnston and Ropelewski (1992) and Graham et al. (1987b) seemed to suggest lower skills for the CCA at these leads. At these longer leads we would expect the skills from the emprical models to be lower than those from the dynamical models (e.g., Chen et al. 1995). The relatively high skills attained by neural network models at longer lead time could be attributed to two factors: (1) The predictors (i.e., the wind stress and the SSTA) and the predictand (i.e., the SSTA) could be increasingly nonlinearly related as the lead time increased, with neural networks having the ability to model nonlinear relations. (2) The wind stress fields we used covered the latitudes from 20 °S to 20 °N, which was wide enough to cover the activities of the off-equatorial oceanic Rossby waves that may be associated with ENSO. Graham et al. (1987b) used the wind fields covering only from 9 °S to 9 °N, which did not reach far enough off the equator, as noted by Graham and White (1988).

To test (1), we used a stepwise multiple linear regression model to fit the data. The linear model was set up exactly the same as the neural network except that the predictors were related to the predictand by linear regression. This simple method was not intended to be representative of an optimal linear model but merely a way to see whether the neural network model learns linear or nonlinear relationships between the predictor and the predictand. Nonetheless, we expect the Nino3.4 SSTA forecasts of this regression model to be comparable to those from CCA models. Figure 8 compares the neural network model and linear regression model in fitting the training data and forecasting the validation data. For 0-month lead, there was no difference between the two models. As lead time increased, the fitting by the simple regression method became poorer, whereas the neural networks fittings showed very slow decay (Fig. 8a). In fact the fit improved as the lead time approached 12 months, where the correlation was above 0.70. The failure of this linear regression model to fit the training data at longer lead of 9 to 12 months was consistent with the results presented by Barnett (1984), who showed that his linear statistical model produced no skillful hindcasts beyond 5-month leads.

What did the neural networks learn from the training data? It did not just learn the linear relationship in the data as it did much better than the regression model at 9 and 12-month leads. That leaves two possibilities: either the neural network learned nonlinear relationships, or it fitted to the noise in the training data. With a very flexible method like neural networks, any degree of fitting can be achieved by adding more neurons in the hidden layer. Better fitting with the training data does not guarantee better forecasting with the validation data. Figure 8b shows that over the validation data, the neural net model was able to produce better skills than the regression model, particularly at longer leads of 9 to 12 months.

Figure 9 compares the fittings in the training data at 12-month lead. The three curves show the corresponding observations (solid line), the neural network hindcasts (dashed) and those of the regression model (dotted). Overall, neural networks hindcasts were better than those of the regression model with a correlation of 0.72 versus 0.38. One particularly interesting feature in Fig. 9 is the period of 1974/75, known as 'aborted El Niño'. Latif et al. (1994) noted that most prediction schemes failed to forecast the anomalously cold year of 1975 as the atmosphere and ocean conditions were favorable for a warm episode to occur. This was the case with our regression model which hindcasted a warm event in 1975. The neural network model was able to hindcast the cold event, though the

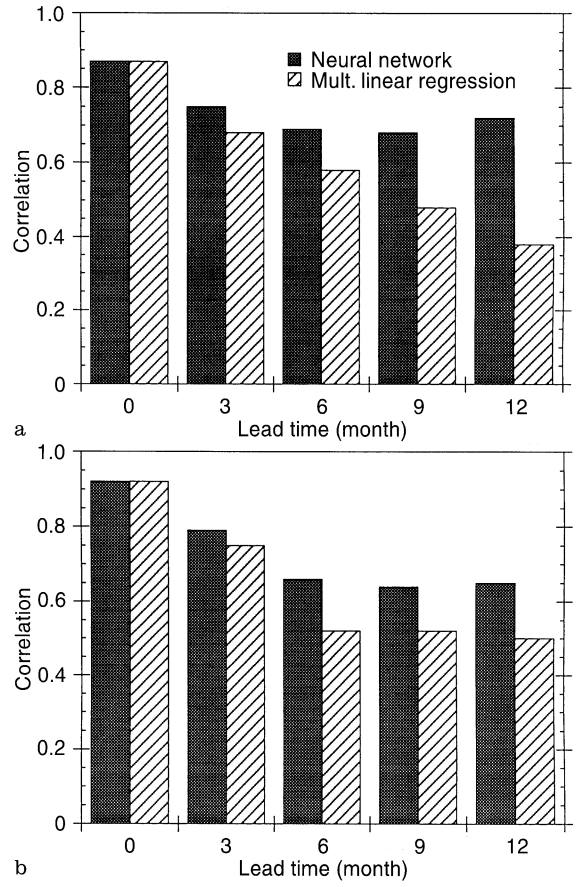

Fig. 8a, b. A comparison of correlation skills between the neural network model and the multiple linear regression model during **a** training period (1952–1981) and **b** forecasting period (1982–1992), for various lead times

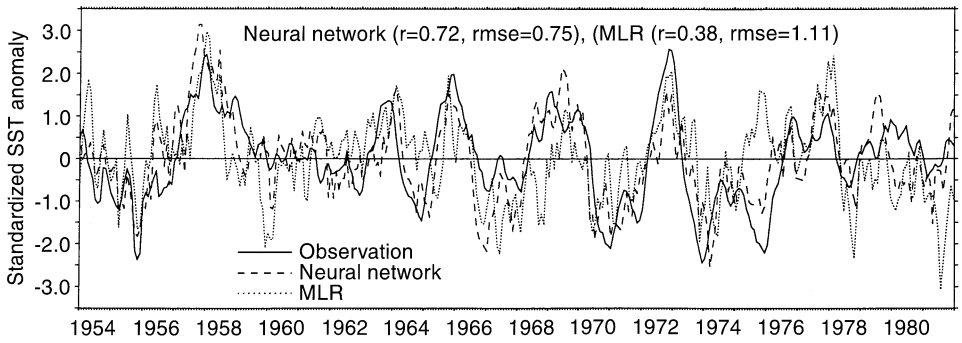

Fig. 9. Standardized SST anomalies of the neural network hindcasts (*dashed curve*) and the multiple linear regression hindcasts (*dotted curve*) for the training data at 12-month lead

trough was somewhat underestimated. These results seemed to suggest that nonlinearity could be important in forecasting at longer leads of 9 to 12 months, even though at shorter leads nonlinearity was negligible. How much nonlinearity is there in ENSO? Munnich et al. (1991) have used a nonlinear atmosphere/ocean model to show that irregular interannual variability can result from the coupling of the atmosphere and the ocean. Mantua (1994) and Mantua and Battisti (1995) analyzed the ZC numerical model and demonstrated that the irregularity in the model ENSO cycle was due to the interaction between two coupled unstable atmosphere/ocean modes. They also concluded that the irregularities in the model ENSO events were caused by deterministic, nonlinear processes in the model associated with the interactions between the unstable modes. These modeling results seemed to suggest that nonlinearity plays an important role in the irregularities of ENSO events, and appeared consistent with the notion that nonlinearity could be important in long lead forecasting.

To test (2), (i.e., the effect of having wind stress fields of different latitudinal extents), we repeated the process by re-calculating the wind stress EOFs for the two smaller bands of 10 °S–10 °N and 6 °S–6 °N. We again trained the neural networks with the first seven wind stress EOFs for these regions plus the SSTA itself, and to the point where all performed comparably in the training data. The forecast skills were then assessed on the validation data. The forecast skills of models using three different wind stress domains i.e., 20 °S–20 °N, 10 °S–10°N and 6 °S–6 °N, showed that for forecasting at 3-month lead or shorter, no distinction can be made from using the three domains (Fig. 10). As lead time increased, the skills of the model using the wind stress of 6 °S–6 °N decayed rather rapidly. There were no differences between the skills of the model using wind stress from 20 °S–20 °N and that from 10 °S–10 °N up to 9-month lead. Overall, the model using wind stress from 20 °S–20 °N seemed to be the best.

While these experiments gave some insight on how the latitudinal width of the wind stress field affects the forecasting skills, they did not provide a definitive answer on whether the off-equatorial Rossby waves are important for the ENSO oscillation. The relative importance of these off-equatorial Rossby waves in driving ENSO has been a topic of debate in recent years (Graham and White 1988; Battisti 1989; Kessler 1991; Graham and White 1991; Battisti 1991; White and Tai 1992). The recent study of

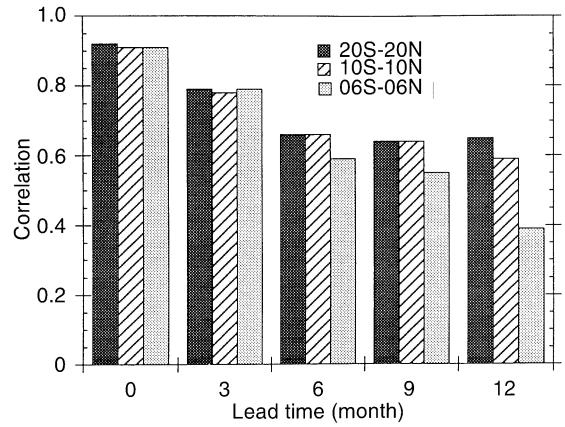

Fig. 10. A comparison of forecast correlation skills between neural network models using wind stress of 20 °S –20 °N, 10 °S –10 °N and 6 °S – 6 °N for the forecasting period (1982–1992), at various lead times

Schneider et al. (1995) identified the low-order freely propagating Rossby waves, excited by wind stress forcing near the date line, as being important. Battisti and Sarachik (1995) concluded that presently, the theory for ENSO events which best fits the observations is the 'delayed-oscillator theory' of ENSO, put forward by Suarez and Schopf (1988) and Battisti and Hirst (1989). According to this theory, the easterly (westerly) wind anomalies over the western Pacific prevailing during the La Niña (El Niño) event force upwelling (downwelling) Kelvin waves which propagate eastward along the equator, causing cooling (warming) at the sea surface in the eastern Pacific, where the thermocline is shallow. The oceanic response to the easterly (westerly) winds in the west, also consists of downwelling (upwelling) Rossby waves, which propagate westward. As the thermocline is deep in the western Pacific, the SST is not strongly influenced by these waves. The Rossby waves then reflect at the western boundary into a packet of downwelling (upwelling) Kelvin waves which propagate eastward. Once the Kelvin waves have propagated far enough eastward, they are able to affect the SST, and a positive (negative) SSTA develops which can grow further by unstable air-sea interactions into an El Niño (La Niña). This theory describes a perfectly periodic oscillation which supports a never-ending succession of alternating warm and cold events. In nature, the occurrences of ENSO events were observed to be irregular, a

characteristic not explained by the delayed oscillator theory. Nonetheless, Mantua and Battisti (1994) stressed that the evolution of wind and SST anomalies in each (warm or cold) event remains consistent with the major features of the theory. They further argued that the only missing ingredient is the initial perturbation for the instability in the central and eastern Pacific to act on. White et al. (1989) presented evidence of positive thermocline depth anomalies, associated with wind-driven baroclinic Rossby waves, along the western boundary during the onset phase of an ENSO event. Graham and White (1990) found that most of the magnitude in this wave activity occurred in response to the wind stress curl forcing in the central and western Pacific, amplified by resonance. They further argued that these anomalies actually triggered the ENSO event. These effects are likely to be dominant only during the onset of an event. The equatorial oceanic response to increasingly anomalous westerly wind activities is responsible for the subsequent development in the peak and mature phases. The wind stress forcing in the off-equatorial domain occurs 1 to 2 years prior to the onset phase, whereas that in the equatorial domain occurs during the later part of the onset phase and continues onto the peak and mature phases, lasting for 9 months or less.

We interpret our model forecasts in light of the theory and evidence presented in these earlier studies. Figure 4 illustrates the predictor-predictand relation for forecasting the warm events of 1982/83, 1986/87 and 1991/92. At all lead times, to forecast the onset phase, the model utilized anomalous wind stress 1–2 years prior to the onset phase. For forecasting the peak and mature phases, the model could have used the anomalous winds responsible for the off-equatorial signals and/or the anomalous winds resulted from the ongoing event. The model skills in forecasting the onset, peak and mature phases reflected the model's ability to detect the off-equatorial wind signals 1–2 years prior to the onset phase and the equatorial wind effects as the event developed. This tendency is consistent not only with the evidence of baroclinic Rossby waves presented by White et al. (1989), but also with the delayed oscillator theory in which Kelvin wave effects dominated the progression of the event from the peak through the mature phase until terminated by the delayed Rossby wave effect.

We were puzzled by the model failure or underestimation for the 1982/83 El Niño at 6 to 12-month leads. This could be due to the unusual evolution of the 1982/83 El Niño (Philander 1990), which did not resemble any of the El Niños in the period 1952–1981 used to train our model. Another plausible explanation is that the winds used as predictors at 9 to 12-month leads did not contain the anomalous off-equatorial winds, which could have occurred late for this El Niño. White et al. (1987) showed the evolution of the heat content propagating as baroclinic Rossby waves from the central tropical North Pacific in mid-1981 to the maritime coast of Asia in late 1981 and early 1982, and generating the equatorial Kelvin wave in middle-to-late 1982 (see their Fig. 3). Also, their comparison of the observed and model anomalous dynamic heights showed that the observed generation of the Rossby waves at the source region (near the date line) came much later (mid to late in 1981) than the model result (mid 1980). These off-equatorial anomalous wind effects were therefore delayed by about a year before the 1982/83 El Niño occurred. With this delay, there would be little or no precursor patterns in the wind stress data for forecasting at longer leads of 9 to 12 months (Fig. 4) and hence the model failed to forecast the event. In comparison to the 1982/83 El Niño, White et al. (1989) noted that the generation of the Rossby waves responsible for the onset of 1986/87 El Niño took place sometime in 1984–1985. In Sect. 6, we ran our model with the SST EOFs as predictors, which contained little off-equatorial signal, and the model tended to give late forecasts for the onset phase even at a lead of 0 month.

4.2 Seasonality of the skills

The seasonality of the forecasting skills was first noted by Hasselmann and Barnett (1981) and subsequently observed in many studies. Barnston et al. (1994) discussed the seasonality of the skills of the five models based on the retroactive forecasts of the period 1982–1992. Their forecasts at 6-month lead were also seasonally dependent, with the forecasts made in the fall (or winter) for the following summer (or fall) having lower skills than those made at other times. This is related to the 'spring barrier', which is difficult to traverse in a forecast (Ji et al. 1994b; Latif et al. 1994). Chen et al. (1995) however found no spring barrier when they used a "nudging" procedure with the ZC model. To examine the seasonality of the forecast skills, we displayed the forecast correlation skills at 0, 3, 6, 9, 12 and 15-month leads for various seasons in Fig. 11. Due to the shortness of the record here (1982–1992), a detailed discussion is not warranted, except to note that the forecast skills of our models are also seasonally dependent, with forecasts for spring and winter having higher skills than those for summer and fall.

5 Sensitivity analyses

We have used the steepest descent with momentum and adaptive learning rates as the optimization technique in

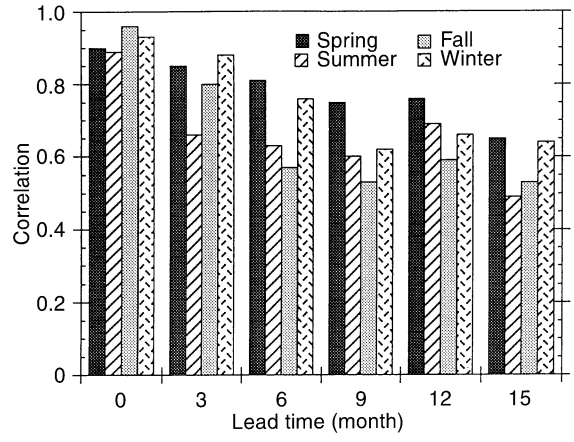

Fig. 11. The forecast correlation skills for various predictant seasons at 0, 3, 6, 9, 12 and 15-month leads

training the networks (Jacobs 1988). The use of momentum decreases the backpropagation's sensitivity to small details in the error surface. This would minimize the possibility of the networks getting stuck in shallow minima. The use of adaptive learning rate reduces the training time where the learning step size is kept as large as possible while keeping learning stable. This method was also adopted in the studies of Grieger and Latif (1994) and Navone and Ceccato (1994). We conducted extra experiments, with different initial learning rate and momentum, which showed the forecast results to be rather insensitive to these parameters. However, we found that this optimization scheme was somewhat sensitive to the initial weights, as was also found by Kolen and Pollack (1991). Hence the optimized solutions may not be the global ones as noted by Grieger and Latif (1994).

We also investigated the effects of the weighting parameters in Eq. (4). With a_1 set to 1, we fixed the value of a_4 to 100, as this allows Eq. (4) to approach Eq. (5). We then varied the other two parameters, a_2 and a_3, and we found the values of 0.1 and 2.5 for a_2 and a_3 respectively to be reasonable with no obvious overfitting. These chosen parameters put more restriction on the neurons in the output layer than those in the hidden layer in trying to contain the networks' excessive capability for fitting the training data.

Choosing the right internal representation in the network is a crucial element in neural network modeling. We varied the number of neurons in the hidden layer from 2 to 5 for all lead times. For all models we used 7 EOF modes and 12 months duration of the wind stress plus the Niño3.4 SSTA itself as predictors. We trained all the networks to the point where they have comparable skills in fitting the training data (1952–1981), then the forecast skills were estimated on the validation data (1982–1992). Figure 12 compares the forecast skills of these models for 0, 3, 6, 9 and 12-month leads. For short leads, three months or less, varying the number of hidden neurons had minor effects on the forecast skills. For 6 and 9-month leads, there was little difference between using 3 and 4 neurons. For 12-month lead, the model with 3 neurons performed best. Overall, the models with 3 neurons in the hidden layer seemed to be the best choice.

We next examined the results from using different duration of the wind stress in the predictor set. Graham et al. (1987b) used 6 months of wind stress to predict the SST whereas Barnston and Ropelewski (1992) used 12 months of SLP as predictors. Here we conducted an experiment where we compared the models using 6, 9 or 12 months of wind stress. In all models, 7 wind stress EOFs and the Niño3.4 SSTA were used as predictors, with 3 neurons in the hidden layer. We again trained all the models to the point where all attained comparable fitting skills in the training data, then assessed the forecast skills on the validation data (Fig. 13). For all lead times, the model using 12 months wind stress generally had the best forecast skills, especially at longer lead times.

We tested how many wind stress EOF modes were needed as predictors, by varying the number of EOFs from 4 to 7. We used 12 months of wind stress EOFs plus the Niño 3.4 SSTA as predictors, with 3 neurons in the hidden layer. We again trained the models to the point where all achieved comparable fitting skills with the training data. The forecast skills of these models (Fig. 14) showed that at lead times up to 6 months, there were only minor differences between the models. For longer leads, the model with 4 EOFs did not produce usable skills. Overall, using 7 EOFs in the model seemed to give better forecast skills at all lead times.

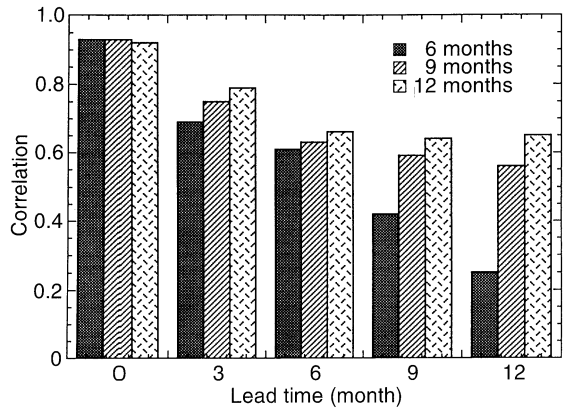

Fig. 13. The forecast correlation skills at 0, 3, 6, 9, 12-month leads for various duration of wind stress data used as predictors

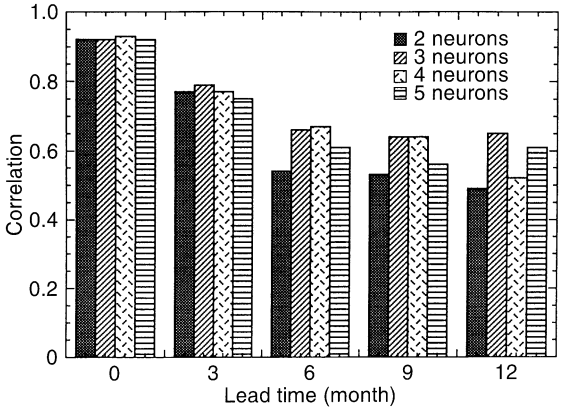

Fig. 12. The forecast correlation skills at 0, 3, 6, 9, 12-month leads, as the number of neurons in the hidden layer varied from 2 to 5

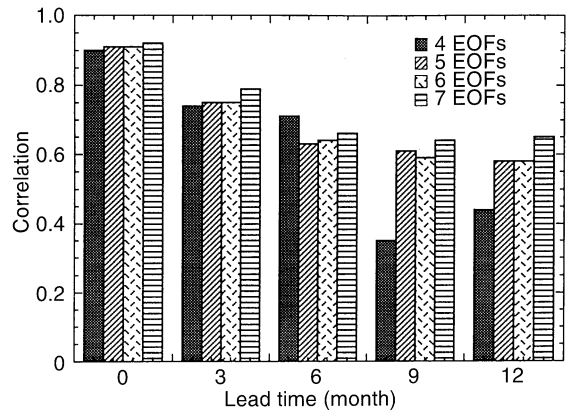

Fig. 14. The forecast correlation skills at 0, 3, 6, 9, 12-month leads, as the number of input wind stress EOFs varied from 4 to 7

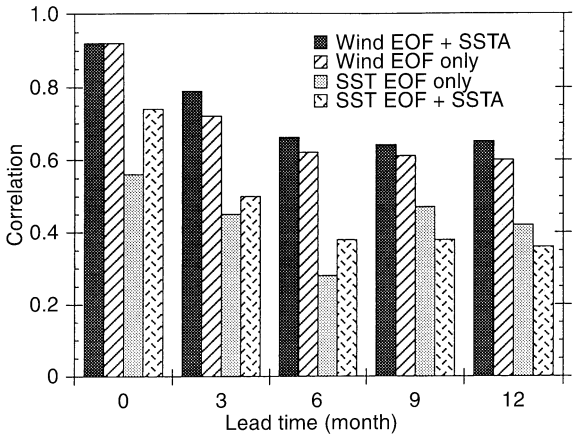

Fig. 15. Niño3.4 SST forecast correlation skills at 0, 3, 6, 9, 12-month leads for neural network models with different sets of predictors

We also tested the consistency of the forecast skills based on the forecast period of 1982 to 1992 with other forecast periods. We partitioned the time series into 4 parts with roughly 10-year duration for each. These 4 periods are 1952–1962, 1963–1972, 1973–1982 and 1983–1992, approximately representing the 1950s, 1960s, 1970s and 1980s data. We withheld one part of the data in turn as forecast validation data and trained the neural networks with the remaining three parts. The forecast skills based the 1980s period were evidently similar to those presented earlier in Sect. 4. The model performances during the other periods were not as good. At 6-month lead, the forecast correlation skills were 0.02, 0.34, 0.25 and 0.70 for the four periods respectively. The fitting skills with the training data were comparable, with a correlation of 0.76, 0.79, 0.78 and 0.63 respectively. That the higher skills for the 1980s were not duplicated for the other decades could well be caused by the following factors: (1) The 1980s period experienced rather strong ENSO activities. In particular, the ENSOs of the 1980s had stronger warming in the central equatorial Pacific (in contrast to the stronger warming in the eastern equatorial Pacific during the earlier ENSOs; see Wang 1995). The quality of the data in the 1980s was also better than the other decades. These factors could make forecasting the Niño3.4 SST (in the central equatorial Pacific) in the 1980s more accurate. (2) Compared to the other decades, the 1960s had the weakest ENSO fluctuations. In general, ENSO forecast schemes do not perform well when there are no strong El Niño or La Niña events. (3) The qualities of the data in the 50s and 60s were relatively poor, and would decrease the forecast skills. Despite the poor data quality in the earlier decades, the networks were apparently able to filter out the noise when using these data for training, thus yielding high forecast skills for the 1980s period. The results of Chen et al. (1995) showed similar trend in which their skills tended to be lower in 1970s but higher in 1980s. One caveat is that there is clearly decadal dependence in the skills, but stationarity is implicitly assumed in our neural network models.

6 The SST as predictor

Barnston and Ropelewski (1992) showed that the inclusion of the SST as a predictor would increase the forecast skills for SST, as the thermal inertia of the SST would help stabilize the forecasts at short leads. Here we also compared the performance between the model using wind stress over four seasons plus the Niño3.4 SSTA in the final season as predictors and the model using only the wind stress. Both models were trained to the point where they have comparable fitting skills on the training data. The forecast skills in Fig. 15 indicated only minor improvement by including the SSTA as predictor.

As ENSO is a coupled phenomenon, and we know it can be forecast using wind data, an interesting question is whether it can be forecast using only ocean data, e.g., SST. To test this, we reran the models with the SST EOFs replacing the wind stress EOFs. The spatial coverage of the SST EOFs was the same as those of the wind i.e., 20 °S–20 °N, 120 °E–70 °W. We also used 7 SST EOFs for the four predictor seasons plus the Niño3.4 SSTA in the final predictor season. We also trained another model with just the 7 SST EOFs as predictors. We trained both models to the point where they had fitting skills comparable to those models fed using wind stress or wind stress plus Niño3.4 SSTA. The forecast skills of these two models (Fig. 15) were much lower than the models using wind stress, with no usable skills for forecasting at 6 months or longer. These results seemed to suggest that Niño3.4 evolution cannot be tracked well based on the basinwide evolution of the SST. This supported our earlier contention that the models with wind stress as the predictor, utilized the effects of the off-equatorial and/or

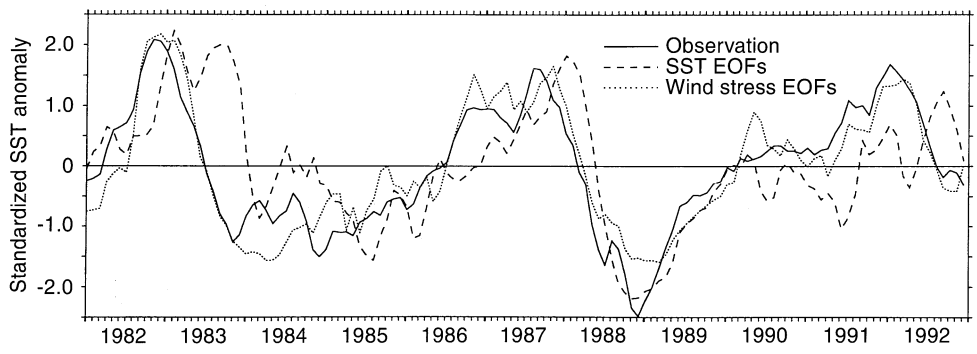

Fig. 16. The standardized SST anomalies of observations (*solid curve*) versus forecasts at 0-month lead for a model using SST EOFs (*dashed curve*) and a model using wind stress EOFs (*dotted curve*) as predictors

the equatorial anomalous winds. It is unlikely that the model used SST EOFs as predictors captured the off-equatorial signals since the signatures of the Rossby waves were less likely to be visible in the SST. This was particularly true in the western Pacific ocean, where the thermocline is deep. White et al. (1987, Fig. 3) showed that these waves were visible in vertically average temperature but not in the SST. In contrast, since it was the anomalous wind which forced these waves, the signatures should appear in the wind stress EOFs. Figure 16 displays the observations versus the forecasts at 0-month lead for models using only SST EOFs (dashed curve) or wind stress EOFs (dotted curve) as predictors. The correlation skills were 0.91 and 0.56 for the wind stress model and the SST model respectively. With the absence of the off-equatorial Rossby waves signals in the SST EOF predictors, the model tended to give late forecasts for the onset of all the events.

7 Summary and conclusion

We have demonstrated the use of neural network models to seasonally forecast the tropical Pacific SST in the Niño 3.4 region (6 °S–6 °N, 120 °W–170 °W). The inputs to the neural networks were the first seven wind stress EOFs of the tropical Pacific (20 °S–20 °N, 120 °E–70 °W) for four seasons and the SSTA in the Niño3.4 region itself for the final season. The period of 1952–1981 was used for training the neural network models, and the period 1982–1992 for forecast validation. At 6-month lead time, neural network models attained comparable skill to the models surveyed by Barnston et al. (1994) (where the forecasts were validated over the period 1982–1993). The results also suggested that the neural network models were viable for ENSO forecasting even at longer leads of 9 to 12 months leads.

The advantage neural network models have over linear statistical models (e.g., the CCA) appeared to lie in the longer lead time regime of 9 to 12 month leads. We found that the ability of a simple linear model in relating the wind stress and the Niño3.4 SSTA declined as lead time increased. At 9 to 12-month leads, the simple linear regression model could not fit the Niño 3.4 SSTA in the training data well. On the other hand, neural network models explained a significant fraction of the variability in Niño 3.4 SSTA at these longer leads and performed well in forecasting the validation data. Based on this comparison, we hypothesized that the underlying relationship between the wind stress and Niño3.4 SSTA becomes increasingly nonlinear as the lead time increases. This nonlinearity becomes possible since many processes could affect the relation at longer lead times. Zebiak and Cane (1987) argued that many processes come into play in relating the thermocline depth to the surface wind stress, and others act somewhat independently, e.g., zonal temperature advection in the oceanic surface layer and moisture convergence effects in the atmosphere. Despite the ability of neural network models in simulating nonlinear relationships, there is no easy way of interpreting the physical meaning of the weights in the hidden and output layers of the network. In the CCA model however, this can be done rather easily by examining its canonical loading patterns. We also did not address the issue of confidence intervals for our forecasts, as there is no easy way of estimating them. A recent study by Chryssolouris et al. (1996) indicated the possibility of estimating the intervals.

We also interpreted our model forecasts in light of the current theories, e.g., the role of off-equatorial baroclinic Rossby waves in triggering the onset of an ENSO event, as proposed by White et al. (1989), and the delayed oscillator theory with respect to the development of the event from the peak through the mature phase (positive Kelvin wave effect) and the termination of the event (delayed negative Rossby wave effect). In the immediate future, we will extend the SST forecasts to beyond the Niño 3.4 region. We will also try forecasting the entire tropical Pacific SST field by forecasting the amplitude of the first few SST EOF modes. We plan to apply the neural network method to the extratropics, where the relationship between the wind stress and the SST is clearly no longer linear.

Acknowledgements. We are grateful to Roy Hourston for providing the COADS data tape, and the PMEL staff for the Ferret plotting package. F. Tangang was supported by the National University of Malaysia, and W. Hsieh was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Atmospheric Environment Service, Environment Canada. We are grateful to two anonymous reviewers for their valuable comments.

References

Barnett TP (1984) Prediction of the EL Niño of 1982–83. Mon Weather Rev 112 : 1043–1047

Barnett TP, Latif M, Graham N, Flugel M, Pazan S, White W (1993) ENSO and ENSO-related predictability. Part I. Prediction of Equatorial Pacific sea surface temperature with a hybrid coupled ocean-atmosphere model. J Clim 6(8) : 1545 –1566

Barnston AG, Ropelewski CF (1992) Prediction of ENSO episodes using canonical correlation analysis. J Clim 5(11) : 1316–1345

Barnston AG, van den Dool HM (1993) A degeneracy in cross-validated skill in regression-based forecasts. J Clim 6 : 963–977

Barnston AG, van den Dool HM, Zebiak SE, Barnett TP, Ji M, Rodenhuis DR, Cane MA, Leetmaa A, Graham NE, Ropelewski CR, Kousky VE, O'Lenic EA, Livezey RE (1994) Long-lead seasonal forecasts–where do we stand? Bull Am Meteorol Soc 75 : 2097–2114

Battisti DS (1989) On the role of off-Equatorial oceanic Rossby waves during ENSO. J Phys Oceanogr 19 : 551–559

Battisti DS (1991) Reply. J Phys Oceanogr 21 : 461– 465

Battisti DS, Hirst AC (1989) Interannual variability in a tropical atmosphere-ocean model: Influence of the basic state, ocean geometry and nonlinearity. J Atmos Sci 46 : 1687–1712

Battisti DS, Sarachik ES (1995) Understanding and predicting ENSO. Rev Geophys Suppl : 1367–1376

Cane MA, Zebiak SE (eds) (1987) Prediction of El Niño events using a physical model. Atmospheric and Oceanic Variability, Royal Meteorological Society

Cane MA, Zebiak SE, Dolan SC (1986) Experimental forecasts of E1 Niño. Nature 321 : 827–832

Cardone VJ, Greenwood JG, Cane MA (1990) On trends in historical marine wind data. J Clim 3 : 113–127

Chauvin Y (1989) Generalization performance of overtrained back-propagation. In: Almedia LB, Wellekens CJ (eds) Proc Eurasip Workshop of Neural Networks. Springer, Berlin Heidelberg New York, pp 46–55

Chen D, Zebiak SE, Busalacchi AJ, Cane MA (1995) An improved procedure for E1 Niño forecasting: implications for predictability. Science 269 : 1699 –1702

Chryssoloris G, Lee M, Ramsey A (1996) Confidence interval prediction for neural network models. IEEE Trans. Neural Network 7(N1): 229–232
Cybenko G (1988) Continous valued neural network with two hidden layers are sufficient. Medford, Department of Computer Science, Tufts University
Cybenko G (1989) Approximation by superpositions of a sigmoidal function. Math Control Signal Sys 2: 303–314
Elsner JB, Tsonis AA (1992) Nonlinear prediction, chaos, and noise. Bull Am Meteorol Soc 73(1): 49–60
Funahasi KI (1989) On the approximate realization of continuous mappings by neural networks. Neural Networks 2: 183–192
Graham NE, White WB (1988) The E1 Niño cycle: a natural oscillator of the Pacific ocean-atmosphere system. Science 240: 1293– 1302
Graham NE, White WB (1990) The role of the western boundary in the ENSO cycle: experiments with coupled models. J Phys Oceanogr 20: 1935–1948
Graham NE, White WB (1991) Comments on "on the role of off-Equatorial Rossby waves during ENSO". J. Phys Oceanogr 21: 453–460
Graham NE, Michaelsen J, Barnett TP (1987a) An investigation of the E1 Niño-southern oscillation cycle with statistical models, 1, predictor field characteristics. J Geophys Res 92(C13): 14251–14270
Graham NE, Michaelsen J, Barnett TP (1987b) An investigation of the E1 Niño-southern oscillation cycle with statistical models, 2, model results. J Geophys Res 92(C13): 14 271–14 290
Grieger B, Latif M (1994) Reconstruction of the E1-Niño attractor with neural networks. Clim Dyn 10(6): 267–276
Hanson SJ, Part LY (1989) Comparing biases for minimal network construction with back-propagation. In: Touretzky DS (ed) Advances in neural information processing systems 1. Morgan Kaufmann, San Mateo, California pp 177–185
Hasselmann K, Barnett TP (1981) Techniques for linear prediction for systems with periodic statistics. J Atmos Sci 38: 2275–2283
Hornik K, Stinchcomb M, White H (1989) Multilayer feedforward networks are universal approximator. Neural Networks 2: 359–366
Jacobs RA (1988) Increased rates of convergence through learning rate adaptation. Neural Networks 1: 295–308
Ji C, Psaltis D (1990) Generalizing smoothness constraints from discrete samples. Neural Comput 2(2): 188–197
Ji M, Kumar A, Leetmaa A (1994a) A multiseason climate forecast system at the National Meteorological Center. Bull Am Meteorol Soc 75(4): 569 –577
Ji M, Kumar A, Leetmaa A (1994b) An experimental coupled forecast system at the National Meteorological Center: some early results. Tellus 46A: 398– 418
Kessler WS (1991) Can reflected extra-equatorial Rossby waves drive ENSO? J Phys Oceanogr 21: 444–452
Kolen JI, Pollack JB (1991) Backpropagation is sensitive to initial conditions. In: Lippmann R, Moody JE, Touretzky DS (eds) Advances in neural information processing system 3. Morgan Kaufmann, San Mateo, California pp 860–867
Lagerloef GSE, Bernstein RL (1988) Empirical orthogonal function analysis of advanced very high resolution radiometer surface temperature patterns in Santa Barbara channel. J Geophys Res 93(C6): 6863–6874
Large WG, Pond S (1981) Open ocean momentum flux measurements in moderate to strong winds. J Phys Oceanogr 11(3): 324 –336
Latif M, Barnett TP, Cane MA, Flugel M, Graham NE, von Storch H, Xu JS, Zebiak SE (1994) A review of ENSO prediction studies. Clim Dyn 9(4): 167–179
Legler DM (1983) Empirical orthogonal function analysis of wind vectors over the tropical Pacific region. Bull Am Meteorol Soc 64(3): 234–241
Mantua NJ (1994) Numerical modeling studies of the E1 Niño/Southern Oscillation. PhD Thesis, University of Washington, Seattle
Mantua NJ, Battisti DS (1995) Aperiodic variability in the Zebiak-Cane coupled ocean-atmosphere model: ocean-atmosphere interaction in the western equatorial Pacific, revised, J Clim 8: 2897–2927
Mantua NJ, Battisti DS (1994) Evidence for the delayed oscillator mechanism for ENSO: the "observed" oceanic Kelvin mode in the far western Pacific. J Phys Oceanogr 24: 691–699
Michaelsen J (1987) Cross-validation in statistical climate forecast models. J Clim Appl Meteorol 26: 1589–1600
Münnich M, Cane MA, Zebiak SE (1991) A study of self-excited oscillations of the tropical ocean-atmosphere system. Part II: nonlinear cases. J Atmos Sci 48: 1238–1248
Navone HD, Ceccatto HA (1994) Predicting Indian Monsoon rainfall – a neural network approach. Clim Dyn 10(6): 305–312
Nguyen D, Widrow B (1990) Improving the learning speed of 2-layer neural networks by choosing initial values of adaptive weights. Int Joint Conf Neural Networks 3: 21–26
Philander SGH (1990) El Niño, La Niña, and the Southern Oscillation. Academic, San Diego, California
Philander SGH, Yamagata T, Pacanowski RC (1984) Unstable air-sea interactions in the tropics. J Atmos Sci 41: 604–613
Posmentier ES, Cane MA, Zebiak SE (1989) Tropical Pacific climate trends since 1960. J Clim 2: 731–736
Preisendorfer RW (1988) Principal component analysis in meteorology and oceanography. Elsevier, New York
Rasmusson EM, Carpenter TH (1982) Variations in tropical sea surface temperature and surface wind fields associated with the Southern Oscillation/E1 Niño. Mon Weather Rev 110: 354–384
Rumelhart DE, Hinton GE, Williams, RJ (1986) Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, Group PR, eds., Parallel distributed processing, Cambridge, MA, MIT Press. 1: 318–362
Schneider EK, Huang B, Shukla J (1995) Ocean wave dynamics and E1 Niño. J Clim 8: 2415–2439
Smith M (1993) Neural networks for statistical modeling. Van Nostrand Reinhold, New York, pp 235
Suarez MJ, Schopf PS (1988) A delayed action oscillator of ENSO. J Atmos Sci 45: 3283–3287
Tang B, Flato GM, Holloway G (1994) A study of Arctic sea ice and sea-level pressure using POP and neural network methods. Atmosphere-Ocean 32: 507–529
Trenberth K, Large W, Olson J (1989) The effective drag coefficient for evaluating wind stress over the ocean. J Clim 2: 1507–1516
Van den Dool HM (1994) Searching for analogues, how long must we wait? Tellus 46A: 314–324
von Storch H, Burger G, Schnur R, von Storch JS (1995) Principal oscillation patterns: a review. J Clim 8(3): 377–400
Wang B (1995) Interdecadal changes in E1 Niño onset in the last four decades. J Clim 8: 267–285
Weigend AS, Rumelhart DE, Hubermann BA (1991) Generalization by weight-elimination with application to forecasting. In: Lippmann R, Moody JE, Touretzky DS (eds) Advances in neural information processing system 3. Morgan Kaufmann San Mateo, California pp 875–882
White WB, Tai CK (1992) Reflection of interannual Rossby waves at the maritime western boundary of the tropical Pacific. J Geophys Res 97: 14305–14322
White WB, Pazan SE, Inoue M (1987) Hindcast/forecast of ENSO events based upon the redistribuation of observed and model heat content in the western tropical Pacific, 1964–86. J Phys Oceanogr 17: 264–280
White WB, He Y, Pazan SE (1989) Off-equatorial westward propagating Rossby waves in the tropical Pacific during the 1982–83 and 1986–87 ENSO events. J Phys Oceanogr 19: 1397 –1406
Woodruff SD, Slutz RJ, Jenne RL, Steurer PM (1987) A comprehensive ocean-atmosphere data set. Bull Am Meteorol Soc 68: 1239–1250
Woodruff SD, Lubker SJ, Wolter K, Worley SJ, Elms JD (1993) Comprehensive ocean atmosphere data set (COADS) Release 1a: 1980–92. Earth Syst Monitor 4(1): 1–8
Zebiak SE, Cane MA (1987) A model El Niño - Southern oscillation. Mon Weather Rev 115(10): 2262–2278