NOTES AND CORRESPONDENCE

Skill Comparisons between Neural Networks and Canonical Correlation Analysis in Predicting the Equatorial Pacific Sea Surface Temperatures

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

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

15 February 1999 and 30 June 1999

ABSTRACT

Among the statistical methods used for seasonal climate prediction, canonical correlation analysis (CCA), a more sophisticated version of the linear regression (LR) method, is well established. Recently, neural networks (NN) have been applied to seasonal climate prediction. Unlike CCA and LR, NN is a nonlinear method, which leads to the question whether the nonlinearity of NN brings any extra prediction skill.

In this study, an objective comparison between the three methods (CCA, LR, and NN) in predicting the equatorial Pacific sea surface temperatures (in regions Niño1+2, Niño3, Niño3.4, and Niño4) was made. The skill of NN was found to be comparable to that of LR and CCA. A cross-validated *t* test showed that the difference between NN and LR and the difference between NN and CCA were not significant at the 5% level. The lack of significant skill difference between the nonlinear NN method and the linear methods suggests that at the seasonal timescale the equatorial Pacific dynamics is basically linear.

1. Introduction

Many forecasting models have been developed for the tropical Pacific climate variability, especially that associated with the El Niño–Southern Oscillation (ENSO) phenomenon. ENSO prediction models can be loosely categorized into two types: dynamical models and statistical models. Barnston et al. (1994) compared two dynamical models, two statistical models, and one hybrid dynamical–statistical model for their performance in ENSO prediction, and the results indicated no significant difference in prediction skills among them.

Of the statistical methods, canonical correlation analysis (CCA) is widely used (Barnett and Preisendorfer 1987; Graham et al. 1987). CCA is a linear method. However, the climate system involves many nonlinear processes, such as convection in the atmosphere and upwelling in the equatorial ocean. The question arises, Will a nonlinear statistical model improve the prediction skill? Recently, there have been studies applying neural networks (NN), a nonlinear statistical method, to seasonal climate prediction (e.g., Derr and Slutz 1994; Tang et al. 1994; Hastenrath et al. 1995; Tangang et al. 1997; Hsieh and Tang 1998). While a number of claims have been made, a rigorous comparison of linear and nonlinear statistical models has not been undertaken. The purpose of this note is to objectively compare the seasonal prediction skills in the tropical Pacific obtained by three statistical models: simple linear regression (LR), CCA, and NN.

The data and their processing are described in section 2. The three statistical models are presented in section 3. A cross-validation procedure, used to estimate the prediction skills, and an approximate significance test for the difference in skills are described in section 4. Prediction skills of the three models are presented and compared in section 5, with discussions provided in section 6.

2. Data, predictors, and predictands

The data in this study came from two datasets: the Comprehensive Ocean–Atmosphere Data Set sea level pressure (SLP) data of the tropical Pacific Ocean for latitudes from 20°S to 20°N (Woodruff et al. 1987), and the National Oceanic and Atmospheric Administration sea surface temperature (SST) data of the tropical Pacific Ocean for latitudes from 30°S to 30°N (Smith et al. 1996; Reynolds and Smith 1994). Both datasets contain monthly $2° \times 2°$ data with time coverage from January 1950 to December 1997.

Corresponding author address: Dr. Benyang Tang, M/S 300-323, Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA 91109.
E-mail: btang@pacific.jpl.nasa.gov

The gridded SLP and SST data went through several steps of processing. The SLP data were first averaged to a lower resolution of 4° lat × 10° long, and then the missing points were filled by spatial linear interpolation. Climatological monthly means, calculated from 1950 to 1997, were removed from both datasets. The SLP and SST data were then smoothed by a 3-month running mean. Empirical orthogonal function (EOF) analysis was performed on each of the two datasets for the purpose of data reduction. The first 7 EOF time series for SLP and the first 10 EOF time series for SST were retained. These retained SLP and SST EOF time series were normalized so they all had the same variance.

For a given month, we stacked the SLP EOF values of 3, 6, and 9 months before this month together with the SLP and SST EOF values of this month. Altogether, this yielded 38 time series. A second-step EOF calculation, called the extended EOF (EEOF; Graham et al. 1987), was carried out to reduce the 38 EOF time series to 12 retained EEOF time series. In this paper, all experiments with the three models used these 12 EEOF time series as predictors, denoted by $\{x_i(t), i = 1, \ldots, 12\}$.

To compare the three models, we calculated the skills of predicting four ENSO indexes: Niño1+2, Niño3, Niño3.4, and Niño4 at a lead time l months ahead. These ENSO indexes are the averaged SST in specific regions in the tropical Pacific (see Barnston and Ropelewski 1992, Fig. 2, their P3 being our Niño3.4). In each experiment, one of these four indexes was used as predictand of the models. The index data were calculated from the gridded SST dataset after smoothing by the 3-month running mean and the removal of the climatological monthly mean as described above. Since the 2° × 2° gridded SST data have grid lines at the even latitudes, we modified the latitudinal extent of Niño3, Niño3.4, and Niño4 to be 4°S–4°N, instead of the conventional definitions of 5°S–5°N.

The lead time l here is defined as the time from the center of the period of the latest predictors to the center of the predicted period. For example, if the latest predictors are the SLP and SST of February 1990 (the mean of 3-monthly data in January, February, and March 1990) and the predictand is the Niño3 of August 1990, then the lead time is 6 months. This is the same lead time definition as that of Chen et al. (1995), but different from that of Barnston et al. (1994), who defined the lead time as the time from the end of the period of the latest predictors to the center of predicted period. With the above example, their lead time would be 4.5 months.

3. LR, CCA, and NN methods

In the present study, the three models, LR, CCA, and NN, establish empirical relations between the 12 EEOF time series predictors $\{x_i(t)\}$ and the SST index predictand $z(t + l)$, where z is any one of the Niño1+2, Niño3, Niño3.4, and Niño4 indices, and the lead time l runs from 3 months to 21 months. The predictor–predictand relation in LR is linear, while NN extends LR by modeling the relation nonlinearly. CCA is linear but differs from LR in that it uses a ''global'' predictand approach, as will be shown later.

a. The LR model

The LR equation between the predictors $\{x_i(t), i = 1, \ldots, 12\}$ and predictand $z(t + l)$ is simply

$$z_{\text{model}}(t + l) = a_0 + a_1x_1(t) + a_2x_2(t) + \cdots + a_{12}x_{12}(t),$$

where $\{a_i, i = 0, \ldots, 12\}$ are calculated by the standard regression (least squares) procedure.

b. The CCA model

The CCA in this study takes the SST EOF time series $\{y_i\}$ as the predictands, instead of an SST index z. CCA first finds a linear combination u_1 of $\{x_i(t)\}$ and a linear combination v_1 of $\{y_i(t + l)\}$ so that $u_1(t)$ and $v_1(t)$ have the maximum correlation coefficient. The $u_1(t)$ and $v_1(t)$ are called the first predictor and predictand CCA time series, respectively. Higher modes of CCA time series $\{u_n, n > 1\}$ and $\{v_n, n > 1\}$ are calculated by demanding maximum correlation coefficient between $\{u_n\}$ and $\{v_n\}$, the linear combinations over the residual from the previous CCA modes. After the CCA time series are found, a multiple linear regression relation is established between the predictor and predictand CCA time series $\{u_n\}$ and $\{v_n\}$. A more detailed description can be found in meteorology literature, for example, Barnett and Preisendorfer (1987) and Graham et al. (1987), and in statistics texts, for example, Manly (1986).

In Barnett and Preisendorfer (1987), the CCA procedure was simplified assuming that the predictors and predictands are normalized and orthogonal. In the present study, due to the cross-validation procedure described in section 4, these two conditions did not hold; thus, the original CCA formula [Eq. (2) in Graham et al. (1987)] was used.

In the experiments of this note, 10 CCA time series were used. Experiments showed the CCA performance was not sensitive to the number of CCA modes as long as three or more CCA modes were used.

c. The NN model

In the neural network literature, predictors are called inputs, and predictands are called targets. The optimization process of finding the model parameters is called training. We also use the word training to denote the model construction of LR and CCA.

The NN model used here is the feed-forward NN with one hidden layer of units (or ''neurons''). The units in

the hidden layer take the model inputs and transfer the signals to the next layer (the output layer), which yields the model output. In our case, the hidden layer has seven units, each performing a simple calculation with the sigmoidal transfer function:

$$o_i = \frac{1}{1 + \exp\left(-\sum_i w_{ij}x_j - b_i\right)},$$

where o_i is the output from the hidden unit i, and x_j are the 12 inputs to the model. The parameters to be determined in the training process are the weights w_{ij} and the biases b_i. The single unit in the output layer transforms the seven values o_i to the model output z_{model} through a linear function, whose parameters (the weights and bias of the output layer) are also determined in the NN training.

The NN training minimizes the cost function

$$J = \sum_t [z_{\text{model}}(t) - z(t)]^2$$

by adjusting the NN parameters, that is, the weights and biases. The back-propagation equation, which is similar to the adjoint equation in variational data assimilation, is used to find the gradient of the cost function with respect to the NN parameters. A conjugate gradient algorithm and a cubic-interpolation line-search method are used to find the minimum of the cost function in an iterative manner (Bishop 1995).

There are two major problems in NN application to climate prediction as discussed in Hsieh and Tang (1998). The first one is overfitting. On one hand, the SLP and SST data records are noisy and short relative to the characteristic timescale of the seasonal signals of interest, and on the other hand, the NN model has 99 parameters (compared to the 13 parameters of the LR model). These two factors combined cause the NN to learn not only the underlying rules but also the noise in the data. To reduce overfitting, we stopped the training after 20 iterations. Other measures for reducing overfitting were studied and compared in Finnoff et al. (1993).

The second problem with NN is instability. When the NN parameters are initialized differently or the training procedure changes slightly, the training often leads to a different NN. The NN training is hence known to be an unstable procedure (Breiman 1996).

To alleviate the problems of overfitting and instability, we used an ensemble of 20 NNs. The parameters in each of them were randomly initialized, and the final prediction was the average of the 20 individual predictions. When compared to the individual NNs, the ensemble predictions have significantly higher skills and are less sensitive to the small changes in the training procedure. Our study failed to find any correlation between the spread of individual member predictions and the skill of the ensemble prediction. Henceforth, the term ''NN model'' will be used to denote the ensemble average of 20 NNs.

4. Estimation of the prediction skill and an approximate significance test on the skill difference

We have designed a cross-validation procedure to estimate the prediction skills. For each lead time from 3 to 21 months, data from a window of the first 7 consecutive years were withheld. A model was constructed from the remaining data. Predictions, starting from each month of the first 5 years of the 7-yr window, were made. Then the 7-yr window were moved forward by 5 yr and the procedure was repeated. This design was to make sure that there was no training data in the prediction target time. The overlap between the prediction input data and the training data is legitimate as it also happens in real-time prediction.

A separate EEOF was done for each window. However, the first stage EOF (described in section 2) was done only once for all validation windows, due to its relatively heavy computation and large number of cross-validation experiments. We assume that this introduced the same amount of artificial skill (if any) to the LR, CCA, and NN models and thus did not alter the conclusion of the comparison of the three models.

The predictions in all the 5-yr windows were then collected to form a prediction time series of the whole period from 1950 to 1997. The mean (prediction bias) over the whole period was removed. The model predictions usually had smaller variance than the observed, so the predictions over the whole period were scaled to match the variance with that of the observation time series, as was done by Barnston and Ropelewski (1992) and Smith et al. (1995). The correlation coefficient between the model predictions and the observed index was taken as a measure of the prediction skill. Following Barnston et al. (1994), when computing the correlation skill over shorter subperiods (5-yr periods, as in the cross-validated t test described later), the mean of the whole period (which was zero after removing the prediction bias) instead of the mean of the subperiod was used to calculate the correlation, so that a prediction that varies in phase but with opposite sign as the observation resulted in a negative correlation.

Besides correlation skills, root-mean-square errors (rmse), normalized by the standard deviation of the observation, can also be used as a skill measure. When the mean of the prediction is removed and the variance of the prediction is scaled to match that of the observation, as was done in all our experiments, there exists a relationship between the rmse and the correlation r: $\text{rmse} = [2(1 - r)]^{1/2}$, as also pointed out by Smith et al. (1995).

The LR and CCA predictions had smaller variances than the NN prediction; for example, for Niño3, the variance of the data, the LR, the CCA, and the NN

predictions are 0.84°, 0.27°, 0.26°, and 0.38°C^2, respectively. These differences did not affect the correlation skill and could be corrected easily by rescaling to match the observed variance.

The LR, CCA, and NN models shared the same computer codes for data processing and cross validation. Only the codes for model building were different. This guaranteed that the differences in skills among the models were not caused by any subtle differences in data processing and cross validation, but by the models themselves.

Significance tests on the skill difference from two models are difficult to carry out. Correlations are rather weak for deciding whether the skills from two models are significantly different. For example, using Fisher's z transform (Press et al. 1986) and assuming 77 independent events (data of 47 yr and a 7.5 months of autocorrelation e-folding time), a model with a correlation skill of 0.60 is different from another model at the 5% significance level if the correlation skill of the other model is below 0.36 or above 0.77. A similar example was also given in Barnston and Ropelewski (1992).

Lacking a better alternative, we adopted an approximate statistical test called cross-validated t test from Dietterich (1998). The cross-validated t test goes well with our cross-validation procedure in skill estimate described above. To compare model A with model B, the cross-validated t test proceeds as follows: For each 5-yr prediction subperiod in our cross-validation procedure, the correlation skills r's for model A and model B are calculated and are transformed by Fisher's z transform $z = 0.5 \ln[(1 + r)/(1 - r)]$ to remove the skewness in the correlation distribution. The difference of the two transformed skills is calculated. There are 10 subperiods, resulting in 10 skill differences. These 10 differences are considered to be independent events and subject to a two-tailed t test to see whether their mean is significantly different from zero.

Dietterich (1998) studied five approximate statistical tests and found that the cross-validated t test is the most powerful one (meaning that it has the largest probability in detecting skill difference when the difference exists). For example, for a 4% difference in misclassification rate, the probabilities for the cross-validated t test to detect the difference to be significant are 28%, 22%, and 17% for three different datasets, respectively, compared to 13%, 9%, and 7% of the next most powerful method (the 5×2 cv paired t test). The drawback of the cross-validated t test is a slightly inflated rate of type I error (the error of detecting a skill difference when there exists no difference): When testing on two models of the same performance, the type I error rates at 5% significance level were 7%, 8%, and 10% for three datasets, respectively. (The correct rate should be 5%.) This inflated rate of type I error is due to the violation of the assumption that each skill difference in a subperiod is independent of the skill differences in other subperiods. Although the data in two subperiods are independent (neglecting the serial correlation over a 5-yr time), the models in two of the subperiods are trained with 90% common training data (when the whole period is divided into 10 subperiods) and thus are somewhat dependent on each other. Because of this violation and the resulting inflated rate of type I error, the test is called an approximate significance test.

5. Skills of the LR, CCA, and NN models

Comparisons of the overall prediction correlation skills of the three models as well as the persistence, for the period of 1950–97, are shown in Fig. 1 for the four indexes. The skills for the lead times of 18 and 21 months are low, generally less than 0.3 in correlation; henceforth we will limit our discussion to the lead times from 3 to 15 months.

The three models perform much better than persistence for lead times of 6 months and longer. However, the skill differences among the three models are small, usually less than 0.05 in correlation skill. The cross-validated t test, described in section 4, was applied to test the differences among the three models. Figure 2 shows the significance levels by which the three pairs of models (NN vs CCA, NN vs LR, and CCA vs LR) are different. The figure shows the significance level of the two-tailed test, and a separate one-tailed test was carried out (not shown) to reveal which model is better than the others when the difference is significant. Figure 2a shows that, at the 5% level, the skill differences between NN and CCA are not significant. This is not surprising given the small difference in correlation skill. The NN is not significantly better than LR either, except for four cases (in Niño4 for 3-month and 6-month lead times, in Niño3 and Niño3.4 for a 12-month lead time). However, CCA was detected to be significantly better than LR for 13 cases out of 28.

This may create an impression of contradiction: NN is comparable to CCA, NN comparable to LR, but CCA better than LR. We offer two explanations. First, it should be remembered that failure to find significant difference between skills of two models does not prove the null hypothesis that the skills of the two model are the same. It can well be that NN is better than LR, or CCA is better than NN, but we just do not have sufficient data to be certain.

Second, it can also be that the cross-validated t test made a type I error for the case of CCA versus LR, that is, the difference between CCA and LR was actually not significant. We found that the variance of the skill differences for the case of CCA versus LR is about 10% of that for the case of NN versus LR, an indication that the cross-validated t test might underestimate the variance for the case of CCA versus LR due to the statistical dependence among the skill differences. Dietterich (1998) examined the cross-validated t test on various NN models, and the robustness of the test on linear models has not been established.

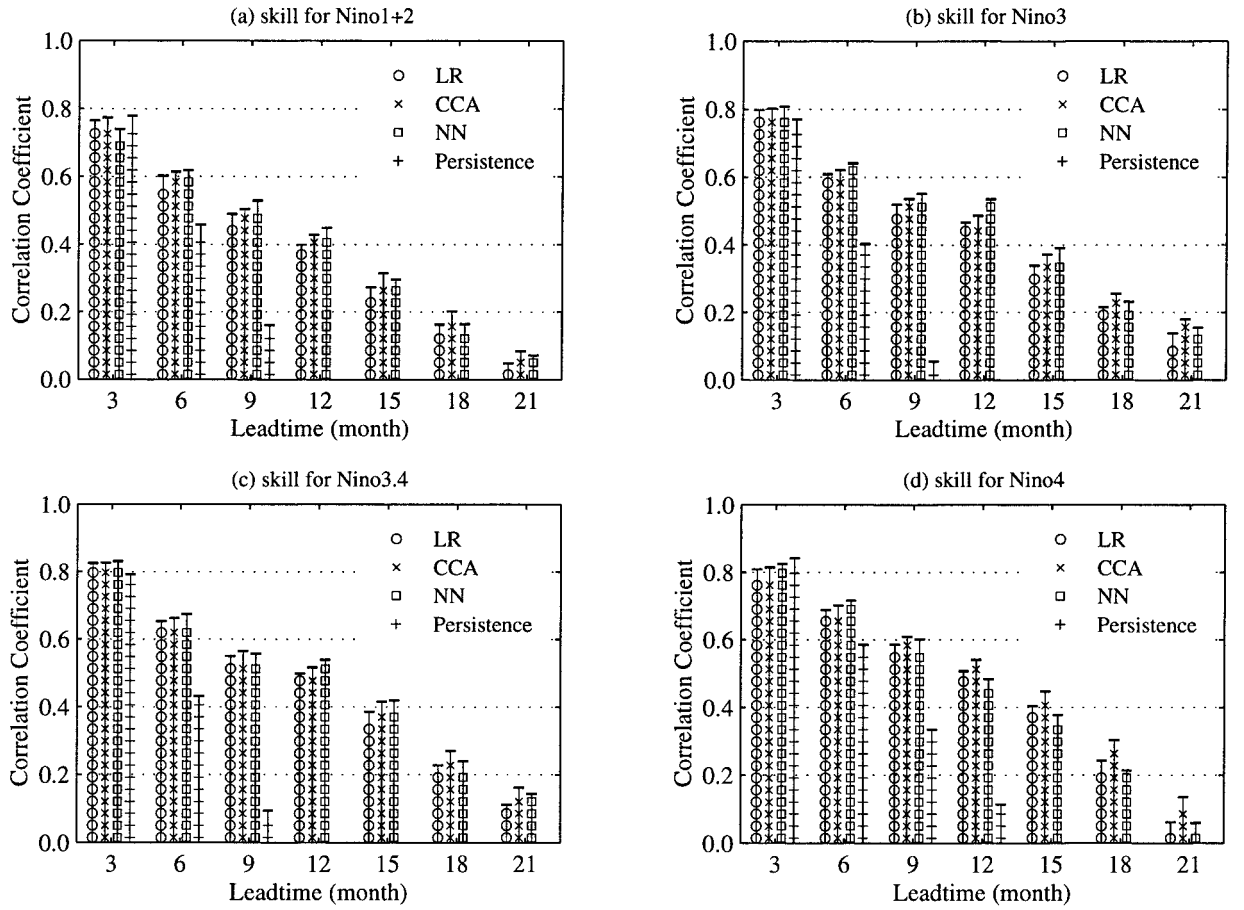

FIG. 1. The prediction correlation skills of the LR, CCA, and NN models and persistence for the four indices (a) Niño1+2, (b) Niño3, (c) Niño3.4, and (d) Niño4, respectively, at various lead times.

Figure 3 shows the prediction skills of NN for Niño3 and Niño4, as functions of lead times and target months. As has been found by many other studies (e.g., Barnston and Ropelewski 1992), skills are high during the winter months and are low during the spring and summer months. All three models (LR and CCA not shown) have similar seasonal dependence.

The lowest skill for Niño3 occurs in April or May, while the lowest skill for Niño4 occurs in July, about 2 or 3 months later. We also calculated the variance of both indices by the calendar months and found that the lowest variance for Niño3 occurs in March and that for Niño4 occurs in June. Thus the month of the lowest skill roughly lags the month of the lowest variance by one month for both Niño3 and Niño4.

6. Discussions

Under cross validation, we have calculated and compared the prediction skills of two linear statistical models (LR and CCA) and a nonlinear statistical model (NN) in predicting the four tropical Pacific SST indices: Niño1+2, Niño3, Niño3.4, and Niño4. All three models had better skills than persistence. However, the three models themselves had similar skills, the differences in correlation skill being less than 0.05. An approximate significance test revealed that at 5% significance level, the skill differences between NN and CCA, and between NN and LR were generally not significant.

Despite its nonlinear capability, NN failed to improved upon the linear methods LR and CCA. We have obtained far better prediction with NN than LR for other datasets, for example, the three-variable Lorenz system (Lorenz 1963), and the laser dataset in the Sante Fe time-series Prediction Competition (Weigend and Gershenfeld 1994). The question is, then, why for the monthly data of the tropical Pacific, NN failed to show improvement over LR and CCA.

There are three possible answers to this question. The most likely one is that over the seasonal timescale, the tropical Pacific is basically linear; nonlinear processes play only minor roles in the system. Constructing linear models from the output of the Cane–Zebiak dynamical model, Xue et al. (1994) found that the skills

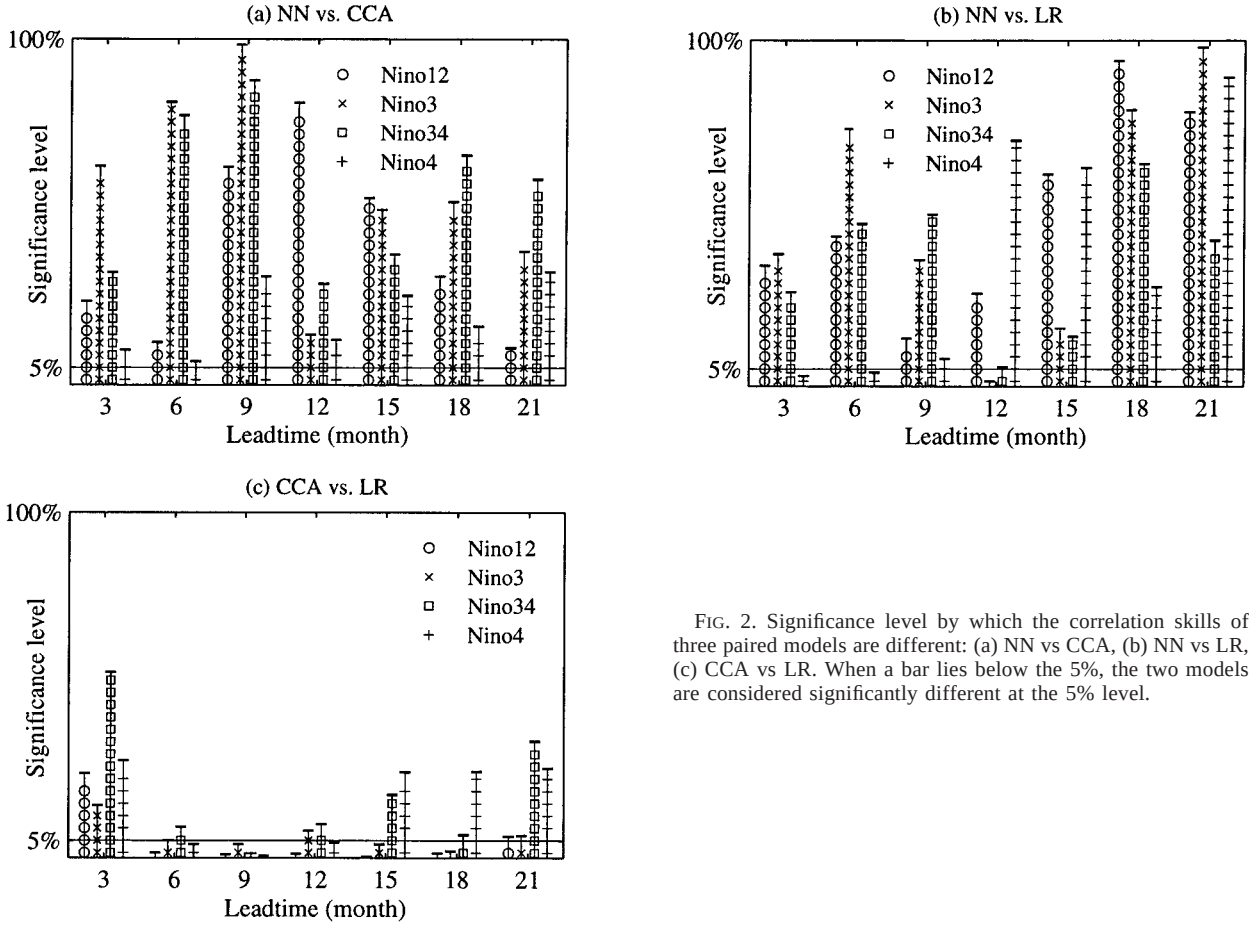

FIG. 2. Significance level by which the correlation skills of three paired models are different: (a) NN vs CCA, (b) NN vs LR, (c) CCA vs LR. When a bar lies below the 5%, the two models are considered significantly different at the 5% level.

of the linear reconstructed model and the original nonlinear model were comparable. In addition, through their ''tau-test,'' Penland and Sardeshmukh (1995) have argued that tropical Pacific SST data are consistent with a linear dynamical model for lead times of up to 10 months.

The second answer is that the climate data records are perhaps not long enough. As a forecasting method,

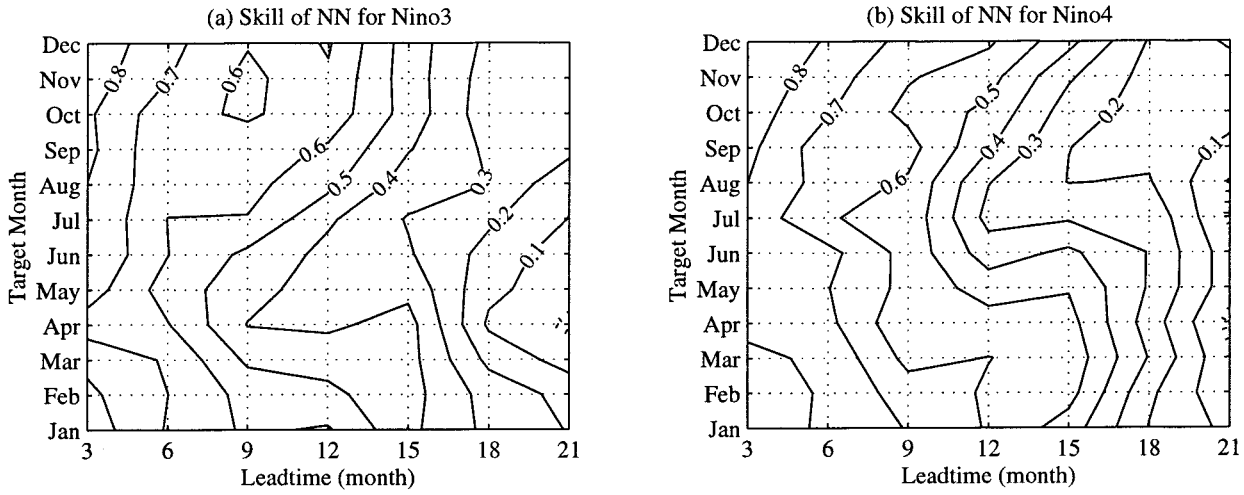

FIG. 3. Prediction skills of NN as a function of lead times and target months for (a) Niño3 and (b) Niño4.

NN is in principle more capable than LR and CCA, but the data requirement is also higher. To extract more than the linear rules from the data, longer records of better quality data are needed.

The third possibility is that we have not yet found the technique to build a good NN model. The process of building an NN involves more decisions than that of LR or CCA. The NN is a relatively new method, and novel techniques appear every year. Several projects are under way to develop improved NN models, for example, NN models that perform nonlinear CCA (given that CCA appeared to outperform LR).

If the main reason why neural networks fail to improve over the linear methods is that the tropical Pacific is basically governed by linear dynamics of seasonal timescales, the next question is, Will neural networks yield better skills in the extratropics? Many studies (e.g., Hoerling et al. 1997; Livezey et al. 1997; Shabbar et al. 1997) have shown that responses of the extratropics climate to the tropical SST are quite nonlinear. This implies that linear statistical methods probably impose substantial model biases in predicting the extratropical variables. We are currently investigating the applications of neural networks to the seasonal predictions of extratropical variables. However, the short climate data records may still be a limitation.

Acknowledgments. We benefited from discussions with Anthony Barnston. Critical comments from two reviewers have helped to improve the manuscript. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada, and from Environment Canada.

REFERENCES

Barnett, T. P., and R. Preisendorfer, 1987: Origins and levels of monthly and seasonal forecast skill for united states surface air temperatures determined by canonical correlation analysis. *Mon. Wea. Rev.,* **115,** 1825–1850.

Barnston, A. G., 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.

Bishop, C. M., 1995: *Neural Networks for Pattern Recognition.* Clarendon Press, 482 pp.

Breiman, L., 1996: Bagging predictions. *Mach. Learning,* **24,** 123–140.

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.

Derr, V. E., and R. J. Slutz, 1994: Prediction of El Niño event in the Pacific by means of neural networks. *AI Interact.,* **8,** 51–63.

Dietterich, T. G., 1998: Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comput.,* **10,** 1895–1923.

Finnoff, W. F., F. Hergert, and H. G. Zimmermann, 1993: Improving model selection by nonconvergent methods. *Neural Networks,* **137,** 771–783.

Graham, N. E., J. Michaelsen, and T. P. Barnett, 1987: 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.

Hastenrath, S., L. Greischar, and J. Heerden, 1995: Prediction of the summer rainfall over South Africa. *J. Climate,* **8,** 1511–1518.

Hoerling, M. P., A. Kumar, and M. Zhong, 1997: El Niño, La Niña, and the nonlinearity of their teleconnections. *J. Climate,* **10,** 1769–1786.

Hsieh, W. W., and B. Tang, 1998: Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bull. Amer. Meteor. Soc.,* **79,** 1855–1870.

Livezey, R. E., M. Masutani, L. A. H. Rui, M. Ji, and A. Kumar, 1997: Teleconnective response of the Pacific-North American region atmosphere to large central equatorial Pacific SST anomalies. *J. Climate,* **10,** 1787–1820.

Lorenz, 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 130–141.

Manly, B. F. J., 1986: *Multivariate Statistical Method: A Primer.* Chapman and Hall, 159 pp.

Penland, C., and P. D. Sardeshmukh, 1995: The optimal growth of tropical see surface temperature anomalies. *J. Climate,* **8,** 1999–2024.

Press, W. H., B. P. Flannery, S. Teukolsky, and W. Vetterling, 1986: *Numerical Recipes.* Cambridge University Press, 818 pp.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimum interpolation. *J. Climate,* **7,** 929–948.

Shabbar, A., B. Bonsal, and M. Khandekar, 1997: Canadian precipitation patterns associated with the Southern Oscillation. *J. Climate,* **10,** 3016–3027.

Smith, T. M., A. G. Barnston, M. Ji, and M. Chelliah, 1995: The impact of pacific Ocean subsurface data on operational prediction of tropical Pacific SST at the NCEP. *Wea. Forecasting,* **10,** 708–714.

——, 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., 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.

Weigend, A. S., and N. A. Gershenfeld, 1994: *Time Series Prediction: Forecasting the Future and Understanding the Past.* Addison-Wesley, 643 pp.

Woodruff, S. D., R. L. J. R. J. Slutz, and P. M. Steurer, 1987: A comprehensive ocean-atmosphere data set. *Bull. Amer. Meteor. Soc.,* **68,** 1239–1250.

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