Neural Network forecasts of the tropical Pacific sea surface temperatures

Aiming Wu, William W. Hsieh
Dept. of Earth and Ocean Sciences, University of British Columbia
Vancouver, BC, Canada

Benyang Tang
Jet Propulsion Laboratory, Pasadena, CA, USA

Neural Networks (in press)
December 11, 2005

Acknowledgement: This work was supported by a strategic grant from the Natural Sciences and Engineering Research Council of Canada. We are grateful for the computational resources from Westgrid.

Corresponding author: William W. Hsieh, Dept. of Earth and Ocean Sciences, University of British Columbia, 6339 Stores Road, Vancouver, BC V6T 1Z4, Canada; Phone: (604) 822-2821, Fax: (604) 822-6088; Email: whsieh@eos.ubc.ca

Running title: Forecast of sea surface temperature

Neural Network forecasts of the tropical Pacific sea surface temperatures

Abstract

A nonlinear forecast system for the sea surface temperature (SST) anomalies over the whole tropical Pacific has been developed using a multi-layer perceptron neural network approach, where sea level pressure and SST anomalies were used as predictors to predict the 5 leading SST principal components at lead times from 3 to 15 months. Relative to the linear regression (LR) models, the nonlinear (NL) models showed higher correlation skills and lower root mean square errors over most areas of the domain, especially over the far western Pacific (west of 155°E) and the eastern equatorial Pacific off Peru at lead times longer than 3 months, with correlation skills enhanced by 0.10-0.14. Seasonal and decadal changes in the prediction skills in the NL and LR models were also studied.

Keywords: neural network, El Niño, ENSO, nonlinear, forecast, sea surface temperature, tropical Pacific

1 Introduction

The El Niño-Southern Oscillation (ENSO) phenomena, the strongest climate fluctuation on time scales ranging from a few months to several years, is characterized by interannual variations of the tropical Pacific sea surface temperatures (SST), with warm episodes called El Niño, and cold episodes, La Niña. ENSO affects not only regional but also global climate (Wallace et al., 1998; Trenberth et al., 1998). Thus successful prediction of ENSO is of both scientific and practical interest.

Since the early 1980s, much effort has been devoted to forecasting the tropical Pacific SST anomalies. ENSO forecast models can be categorized into three types: coupled physical models, statistical models and hybrid models (Latif et al., 1994, 1998). Despite their relative simplicity, Barnston et al. (1994) found that the statistical models have comparable skills (to the physical models) in forecasting the SST anomalies in the central-eastern equatorial Pacific at up to 6-month lead time. The statistical models are still playing an important role in ENSO forecasts.

Most of the statistical models used for ENSO forecast have been linear models, e.g., multivariate linear regression (LR) and canonical correlation analysis (CCA, Barnston and Ropelewski, 1992). However, there is an increasing body of evidence indicating that ENSO has nonlinear features. The SST (and surface wind) spatial anomaly patterns during the warm and cold episodes are clearly not mirror images of each other (Monahan, 2001; Wu and Hsieh, 2002), and this asymmetry has been attributed to nonlinear physics (An and Jin, 2004).

A series of nonlinear regression models using neural network (NN) approaches has been developed by our group to predict the SST anomalies over several key regions in the equatorial Pacific (Tangang et al., 1997, 1998a, 1998b, Tang et al., 2000), where in the latest version the predictors were the sea level pressure (SLP) and the prior SST anomalies over the tropical Pacific, and the predictand was the SST anomaly spatially averaged over a specific region in the equatorial Pacific, e.g. the Niño3.4 region (5°S-5°N, 170°W-120°W). In this paper, we built NN models to predict each of the five leading principal components (PCs) separately, and then the predicted PCs were combined with the corresponding eigenvectors (also called empirical orthogonal functions, EOFs) to yield the forecast of SST anomalies over the whole tropical Pacific. Bayesian regularization (Mackay, 1992) was used during the NN training, which effectively reduced the danger of overfitting and dramatically shortened the process of the model searching.

The data and model building are introduced in Section 2. Forecast skills of the NN models and comparison with the LR models are presented in Section 3. Seasonal and interdecadal changes of forecast skills are discussed in Section 4, and Section 5 offers a summary and discussion.

2 Data and model

2.1 Data, predictors and predictands

The data used in this study came from two datasets: (a) the monthly SLP on 2.5°×2.5° grids from the NCEP/National Center for Atmospheric Research (NCAR) reanalysis, (Kalnay et al., 1996; downloadable from ftp.cdc.noaa.gov/Datasets/ncep.reanalysis.derived/surface); and (b) the monthly extended reconstructed SST on 2°×2° grids (ERSST version 2; Smith and Reynolds, 2004; downloadable from ftp.ncdc.noaa.gov/pub/data/ersst-v2). Although the ERSST data were available back to 1854, only data after January 1948 were used here. Anomalies were calculated by subtracting the monthly climatology based on the 1948-2004 period.

The predictand is one of the five leading PCs of the SST anomalies over the tropical Pacific, i.e., the SST PCs were predicted separately. The spatial patterns (or EOFs) of the 5 leading SST modes are shown in Fig. 1, which account for 85.2% of the total variance of the SST anomalies.

For the predictors, after applying a 3-month running mean to the gridded anomaly data, principal component analysis (PCA) was performed separately on the two anomaly datasets over the tropical Pacific (SLP: 120°E-70°W, 20°S-20°N; and SST: 124°E-70°W, 20°S-20°N) with 7 SLP PCs and 9 SST PCs retained. These retained SLP and SST PCs were then separately normalized by dividing the standard deviation of their first PC. For a given month, we stacked the SLP PCs of 3, 6 and 9 months before this month together with the SLP and SST PCs of this month, altogether yielding 37 time series. A second PCA, also known as extended EOF (EEOF) analysis or singular spectrum analysis (SSA) (von Storch and Zwiers, 1999) was carried out to compress this 37-PC time series to 12 SSA PCs, which served as the final predictors.

Similar to Tang et al. (2000) and Chen et al (1995), the lead time is defined as the time from the center of the period of the latest predictors to the center of the predicted period.

A cross-validation scheme was used to evaluate the forecast skills. The data record (January 1948 – October 2005) was divided into ten equal segments. Data from one segment were withheld as validation data, while data from the other nine segments were used to train the models. Thus, independent forecast was made for the period of the validation data using the models based on the training data. This procedure was repeated until all ten segments were predicted, and the correlation and root mean square error (RMSE) between the predicted SST anomalies and the corresponding observations could be calculated over the whole record.

2.2 The NN model

The NN model used here is the standard feed-forward multi-layer perceptron NN model with Bayesian regularization (Bishop, 1995). The NN has multiple inputs (i.e., the predictors) $(x_i, i = 1, \cdots, l)$, which are nonlinearly mapped to m intermediate variables called hidden neurons $(h_j, j = 1, \cdots, m)$, which are then linearly mapped to a output variable (y, the

predictand). Here y is one of the five leading PCs of SST anomalies over the tropical Pacific. Mathematically, the relationship linking input to output is given by

$$h_j = \tanh(\sum_{i=1}^{l} w_{ij} x_i + b_j), \quad j = 1, \cdots, m, \tag{1}$$

$$y = \sum_{j=1}^{m} \tilde{w}_j h_j + \tilde{b}. \tag{2}$$

The characteristics of this empirical model are determined by $m \times (l+1)$ weight parameters (w_{ij} and $\tilde{w}_j$) and $m+1$ bias parameters (b_j and $\tilde{b}$). Given enough hidden neurons, this NN model is capable of representing any nonlinear continuous function $y = f(\mathbf{x})$ to arbitrary accuracy (Bishop, 1995). Starting from random initial values, the model parameters are optimized so that the cost function, which is basically the mean square error (MSE) between the output (y) and the target SST PC, is minimized.

There are two potential problems in nonlinear optimization: (a) local minima, i.e., the optimization often terminates at a local (non-global) minimum of the cost function; (b) overfitting, i.e., the trained model fits to the noise in the training data. To avoid these problems, we adopted the following two strategies: (a) Bayesian regularization was used in the NN training, where the optimal weight penalty parameter was estimated by a Bayesian approach (Mackay, 1992), as coded by the program "trainbr" in the MATLAB Neural Network toolbox. (b) Only 85% of the training data (denoted by D85, randomly chosen from the 9 segments of training data) were used to train the NN model and the remaining 15% (D15) were reserved for an overfitting test. For each set of D85, we trained an NN model, and a corresponding LR model, then independent forecasts were made for the D15 data. If the correlation skill of the NN model was higher than that of the LR model, and MSE of the NN model less than that of the LR model, the NN model was accepted; if otherwise, it was rejected. This entire procedure was repeated until 30 NN models had been accepted. The average of the predictions from these 30 models was used as the final forecast model. Since there are 10 segments for cross-validation and each segment has 30 models, a total of 300 models are used for forecasting over the whole record.

The above procedure was repeated until all SST PCs at all lead times (3, 6, 9, 12 and 15 months) were predicted. In each case, eight experimental forecasts were carried out separately with the number of hidden neurons (m) varying from 1 to 8. Table 1 shows the number of hidden neurons used in the NN, which gives the best correlation skill among the 8 NN models. Apparently, in different cases, the m may be very different, which is understandable as the nonlinear relationship between the predictors and predictand may change when different lead time or target SST PC is applied.

3 Forecast results

A comparison of forecasts for individual SST PCs between the NN models and LR models is given in Fig. 2, where the cross-validated correlation skills of the NN models are generally

higher than those of the LR models except the forecasts of PC_5 at 12 and 15-month lead (Fig. 2e), and of PC_1 at 3 and 6-month lead, where the NN skill is marginally lower than the LR skill (Fig. 2a). The NN models show considerable improvement over the LR models for PC_2 and PC_4 with the correlation skill enhanced by 0.1-0.2 at most lead times (Figs. 2b,d), and moderate improvement for PC_3 by 0.05-0.1 (Fig. 2c). Note that the SST EOF_2 shows a pattern with the strongest variability over the eastern equatorial Pacific off Peru, centered roughly at the Niño1+2 area (90°W-80°W, 10°S-0°) (Fig. 1b), while EOF_3 shows strongest variability over the western Pacific (Fig. 1c), and EOF_4 over central-eastern equatorial Pacific, roughly covering the Niño3 area (150°W-90°W, 5°S-5°N) (Fig. 1d). Therefore, it can be expected that main improvements of the NN forecasts over the LR forecasts will occur in the eastern equatorial Pacific and western Pacific.

Now we combine these predicted SST PCs with the corresponding EOFs (Fig. 1) to reconstruct the SST anomalies over the whole tropical Pacific. Since the NN forecasts of PC_5 at 12 and 15-month lead and PC_1 at 3 and 6-month lead are not as good as the LR forecasts, we used the LR forecasts to substitute for the NN forecasts in these four cases. The NN model results (with the LR substitution in the four special cases) will be referred to as the "nonlinear" (NL) model forecasts. At each spatial grid over the tropical Pacific, the correlation between the predicted SST anomalies and observations was computed. Figs. 3a-e show the correlation skills of the NL models at lead times from 3 to 15 months, where high correlations (over 0.85, 0.75, 0.65, 0.5 and 0.45 at 3, 6, 9, 12 and 15-month lead, respectively) appear over the central equatorial Pacific near the dateline, which drop slowly eastward, but rapidly westward. The difference in correlation skills between the NL models and LR models (NL minus LR) is shown in Figs. 3f-j. We see positive difference covering almost the whole domain, indicating that the NL models generally perform better than the LR models. Despite the slight improvement (by less than 0.05) over the central equatorial Pacific, the correlation skills are considerably enhanced (by 0.1-0.14) in far western Pacific (west of 155°E) and in eastern equatorial Pacific off Peru, especially, at 6, 9 and 12-month lead (Figs. 3g-i). At 3-month lead, the NL and LR models give almost the same skills over the central-eastern Pacific, but the NL models are still slightly better than the LR models in the western Pacific with correlation increased by 0.04 (Fig. 3f). At 15-month lead, a moderate improvement (by about 0.08) can be seen over the eastern tropical Pacific (Fig. 3j).

Fig. 4 shows the time evolution of SST anomalies averaged within the Niño3 region, predicted by the NL models and the LR models at lead times from 3 to 15 months. Both the NL models and LR models gave successful forecasts for the major El Niño episodes (1957/58, 1965/66, 1968/69, 1972/73, 1982/83, 1986/87, 1990/92 and 1997/98) and La Niña episodes (1954/55, 1964/65, 1973/74, 1975/76, 1984/85, 1988/89 and 1999/2000) at up to 9-month lead. At 3 and 6-month lead, the NL forecasts are visually very similar to the LR forecasts in both amplitude and phase of the SST variations (Figs. 4a,b), but at longer lead time, the LR models underestimate the positive SST anomalies and overestimate negative SST anomalies, esp. after 1980. The NL and LR skills (as listed in the upper-right corner in each panel of Fig. 4) indicate that the NL models performed slightly better than the LR models in forecasting the Niño3 SST anomalies. In the Niño1+2 area, the NL models have greater advantage over the LR models (Fig. 5), where the NL forecasts (the solid lines) agree better with the observations (shading) than the LR forecast (dashed lines), especially for the

1982/83 and 1997/98 strong warm episodes (Figs. 5a-c). The difference in correlation skill (NL minus LR) exceeds 0.12 at 9 and 12-month lead.

A comparison of the forecast skills (of both correlation and RMSE) between the NL models and LR models (Fig. 6) over the Niño4 [160°E-150°W, 5°S-5°N], Niño3.4, Niño3 and Niño1+2 regions (stretching from the western equatorial Pacific to eastern equatorial Pacific) shows that the NL models have generally higher correlation skills and lower RMSE than the LR models. It is interesting that, at a given lead time, as the SST region is shifted from west to east, both the NL and LR skills drop, but the improvement in the NL skills over the LR skills increases, consistent with Fig. 3. Note that the Niño4 area is not west enough to cover the gray areas (Figs. 3f-j) in the far western Pacific, so the improvement of the NL models is not remarkable in Fig. 6a.

4 Seasonal and interdecadal changes of predictability

The correlation skills for Niño3 SST index, as functions of lead time and target month are shown in Figs. 7a and 7b, respectively, for the LR and NL forecasts. For each month, 55 years (1950-2004) data were used to calculate the correlations. As has been found by other studies (Latif et al., 1998), both models have the highest skills during winter months. For LR models, the lowest skills occur during the boreal spring at 3 and 6-month lead, and during summer at 9 and 12-month lead; while for NL models, the lowest skills are mainly limited in spring months (April – June), which is usually called the spring prediction barrier. The difference between the NL and LR correlation skills (Fig. 7c) indicates that the NL skills are up to 0.16 higher in the summer and fall seasons at 9 and 12-month lead, eliminating the "summer prediction barrier" found in the LR forecasts.

We also calculated correlation skills of Niño3 index during different decades since 1950. Both models show higher skills during the 1980s at all lead times (Figs. 7d,e). When the lead time is longer than 9 months, both models exhibit lower skills during the 1950s, and the LR models experienced a second low skill period during the mid 1970s (Fig. 7d). Fig. 7f displays the difference between the decadal NL skills and LR skills, indicating that NL models perform generally better than LR models since 1960, especially during the 1970s with correlation skills enhanced by up to 0.18 at 12-month lead.

5 Summary and Discussion

In summary, a nonlinear ENSO forecast system was developed via a neural network (NN) modeling approach. Four consecutive seasons of SLP anomalies plus the current season of SST anomalies were used as predictors to predict the 5 leading SST PCs to give a forecast of the SST anomalies over the whole tropical Pacific. Compared to the linear regression (LR) models, the nonlinear (NL) models showed advantage in forecasting the SST anomalies over the far western Pacific (west of 155°E) and the eastern equatorial Pacific off Peru with

correlation skills enhanced by 0.10-0.14. The advantage of the NL models over the LR models was found to occur (a) at longer lead time (> 6 months), (b) during the boreal summer and fall seasons, and (c) since 1960, especially during the 1970s.

ENSO is traditionally focussed on the SST variability over the central-eastern equatorial Pacific, where the SST has the largest variance. However, the main large-scale atmospheric convective activities occur in the western equatorial Pacific warm pool region, where the ocean has the warmest climatological surface temperature. Even a small change in the SST over this western Pacific region could lead to dramatic changes in atmospheric convections, which could in turn cause appreciable changes in the extratropical climate through atmospheric teleconnections. It is interesting to see that the NL models in this study give a considerable improvement in forecasting the SST anomalies over the far western Pacific, thus a better forecast for extratropical climate could be expected if the NL predicted SSTs are used in turn to provide bottom boundary conditions for an atmospheric general circulation model.

References

An, S.-I. and F.-F. Jin, (2004). Nonlinearity and asymmetry of ENSO. Journal of Climate, **17**, 2399-2412.

Barnston, A.G., and Ropelewski, C.F. (1992). Prediction of ENSO episodes using canonical correlation analysis. Journal of Climate, **5**, 1316-1345.

Barnston, A.G., van den Dool, H.M., Zebiak, S.E., Barnett, T.P., Ji, M., Rodenhuis, D.R., Cane, M.A., Leetmaa, A., Graham, N.E., Ropelewski, C.R., Kousky, V.E., O'Lenic, E.A., and Livezey, R.E. (1994). Long-lead seasonal forecasts: where do we stand? Bulletin of American Meteorological Society, **75**, 2097-2114.

Bishop, C.M. (1995). **Neural networks for pattern recognition**. Clarendon Pr. Oxford.

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

Kalnay, M.E., et al. (1996). The NCEP/NCAR Reanalysis Project. Bulletin of American Meteorological Society, **77**, 437-471.

Latif, M., Barnett, T.P., Cane, M.A., Flugel, M., Graham, N.E., von Storch, H., Xu, J.S., and Zebiak, S.E. (1994). A review of ENSO prediction studies, Climate Dynamics, **9**, 167-179.

Latif, M., Anderson, D., Barnett, T.P., Cane, M.A., Kleeman, R., Leetmaa, A., O'Brien, J., Rosati, A., Schneider, E. (1998). A review of the predictability and prediction of ENSO. Journal of Geophysical Research, **103(C7)**, 14375-14394.

MacKay, D. (1992). "Bayesian interpolation," Neural Computation, **4(3)**, pp. 415-447.

Monahan, A.H. (2000). Nonlinear Principal Component Analysis: Tropical Indo-Pacific Sea Surface Temperature and Sea Level Pressure. Journal of Climate, **14**, 219-233.

Smith, T.M. and Reynolds, R.W. (2004). Improved Extended Reconstruction of SST (1854-1997) Journal of Climate, **17**, 2466-2477.

Tang, B., Hsieh, W.W., Monahan, A.H., and Tangang, F.T. (2000). Skill Comparisons between Neural Networks and Canonical Correlation Analysis in Predicting the Equatorial Pacific Sea Surface Temperatures. Journal of Climate, **13**, 287-293.

Tangang, F.T., Hsieh, W.W., and Tang, B. (1997). Forecasting the equatorial Pacific sea surface temperatures by neural network models. Climate Dynamics, **13**, 135-147.

Tangang, F. T., Hsieh, W.W., and Tang, B. (1998a). Forecasting the regional sea surface temperatures of the tropical Pacific by neural network models, with wind stress and sea level pressure as predictors. Journal of Geophysical Research, **103**, 7511-7522.

Tangang, F. T., Tang, B., Monahan, A.H., and Hsieh, W.W. (1998b). Forecasting ENSO events– a neural network-extended EOF approach. Journal of Climate, **11**, 29-41.

Trenberth, K.E., Branstator, G.W., Karoly, D., Kumar, A., Lau, N.-C., and Ropelewski, C. (1998). Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. Journal of Geophysical Research, **103(C7)**, 14291-14324.

von Storch, H. and Zwiers, F.W. (1999). **Statistical Analysis in Climate Research**. Cambridge Univ. Pr., 484 pp.

Wallace, J.M., Rasmusson, E.M., Mitchell, T.P., Kousky, V.E., Sarachik, E.S., and von Storch, H. (1998). On the structure and evolution of ENSO-related climate variability in the tropical Pacific: Lessons from TOGA. Journal of Geophysical Research, **103(C7)**, 14241-14260.

Wu, A. and Hsieh, W.W. (2002). Nonlinear canonical correlation analysis of the tropical Pacific wind stress and sea surface temperature. Climate Dynamics, **19**, 713-722.

Table 1: Number of hidden neurons used in the NN models as the predictand varies from the SST PC_1 to PC_5, and lead time from 3 to 15 months.

lead time	PC_1	PC_2	PC_3	PC_4	PC_5
3	1	2	2	1	4
6	1	2	3	6	3
9	5	2	4	2	5
12	5	2	1	3	1
15	2	7	1	3	6

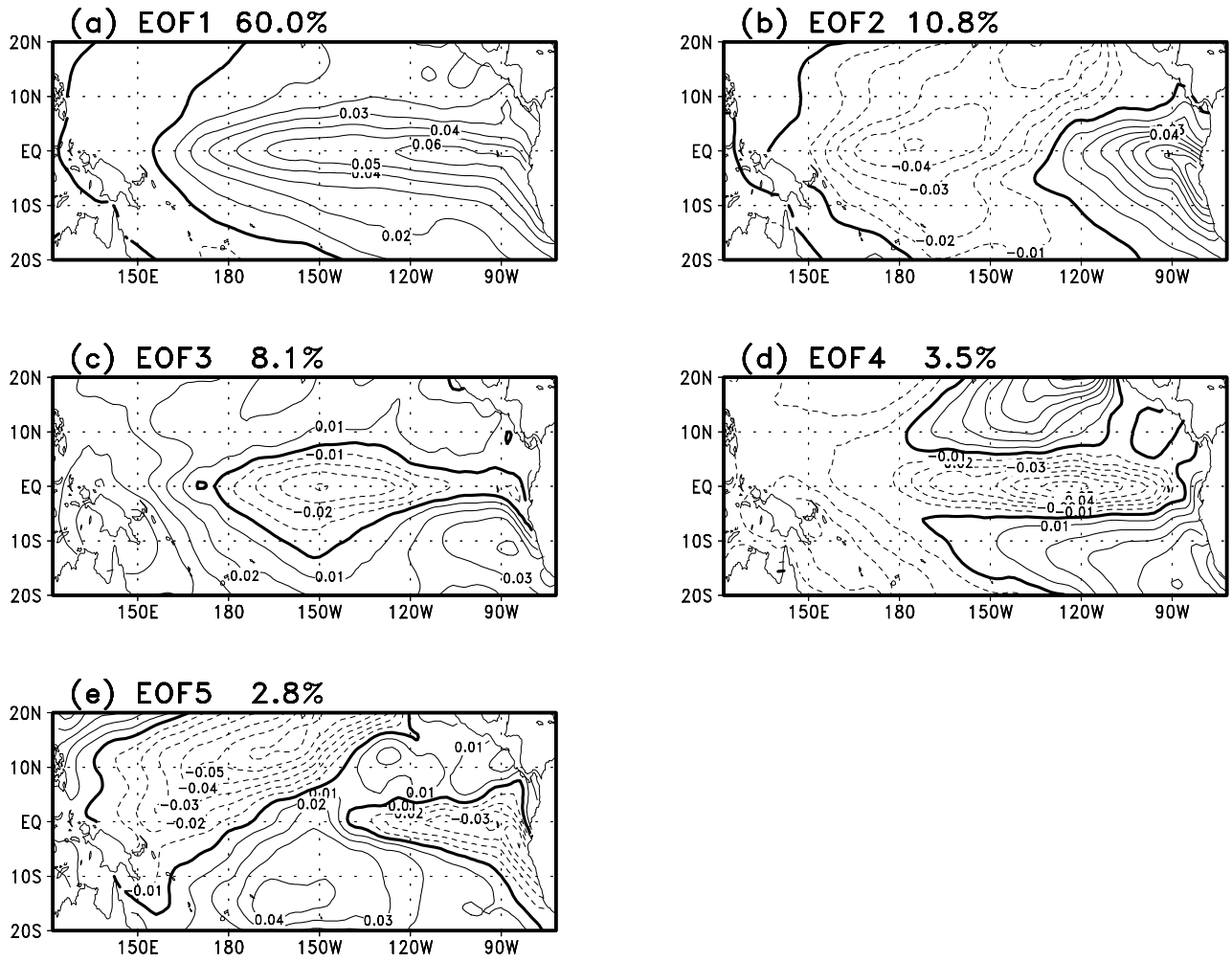

Figure 1: The 5 leading EOF spatial modes (i.e. the eigenvectors of the principal component analysis) of the tropical Pacific SST anomalies. The EOFs have been normalized to unit norm, and the contour interval is 0.01. The percentage variance explained by each mode is given in the panel title.

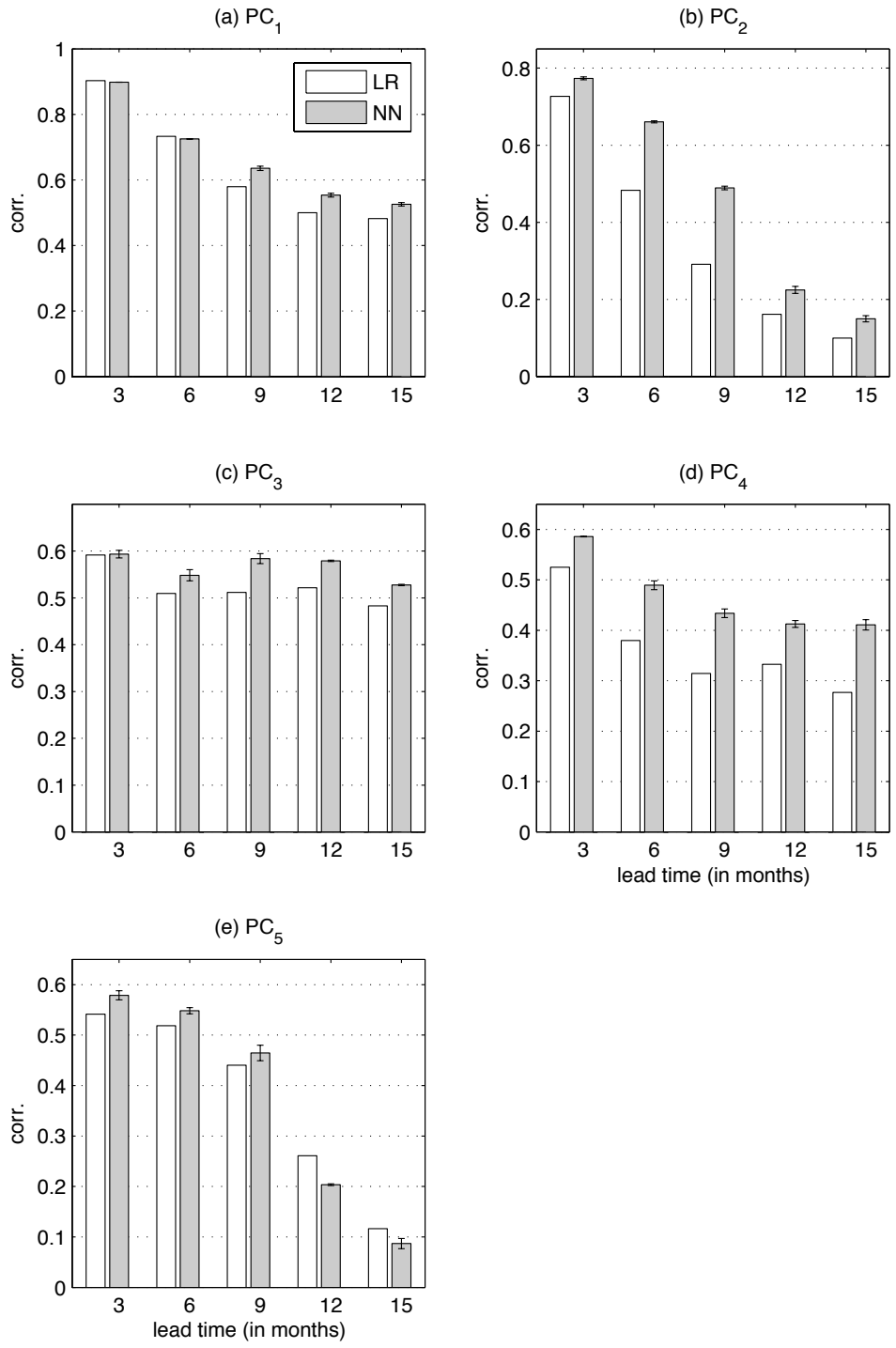

Figure 2: Cross-validated correlation skills of the 5 leading PCs of the tropical Pacific SST anomalies predicted by the LR models (white vertical bars), and NN models (grey bars). The error bars indicate ± 1 standard error of the forecast skills from the 30-member NL ensemble.

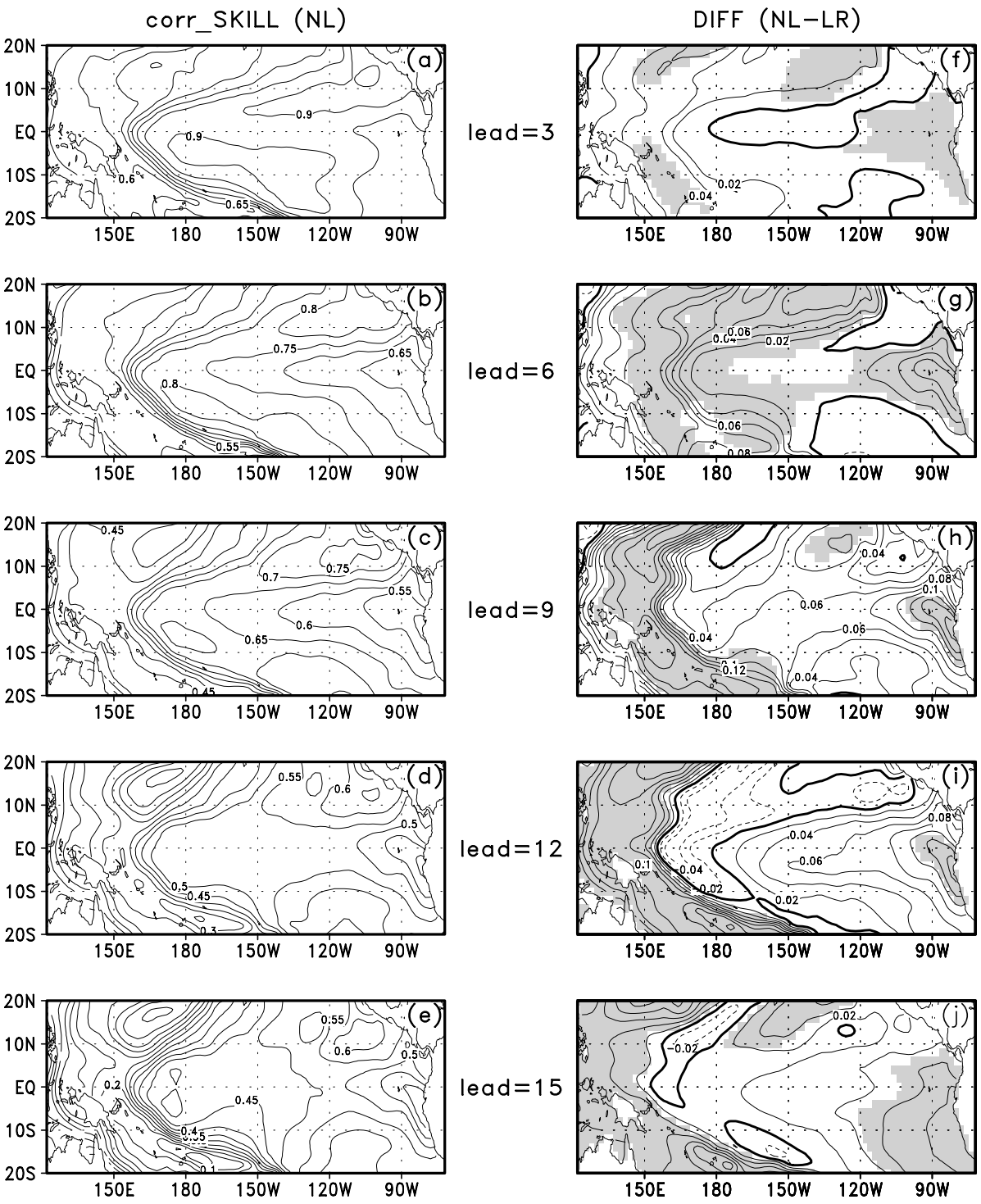

Figure 3: Correlation skills of the NL models at lead time from 3 to 15 months (panels a–e in the left column) and the difference in correlation skills between the NL models and LR models (NL minus LR, panels f–j, right column). The contour interval is 0.05 in panels a–e, and 0.02 in panels f–j, with negative contours dashed, and the gray areas in panels f–j indicate that the difference is statistically significant at the 5% level based on a right-tailed t-test.

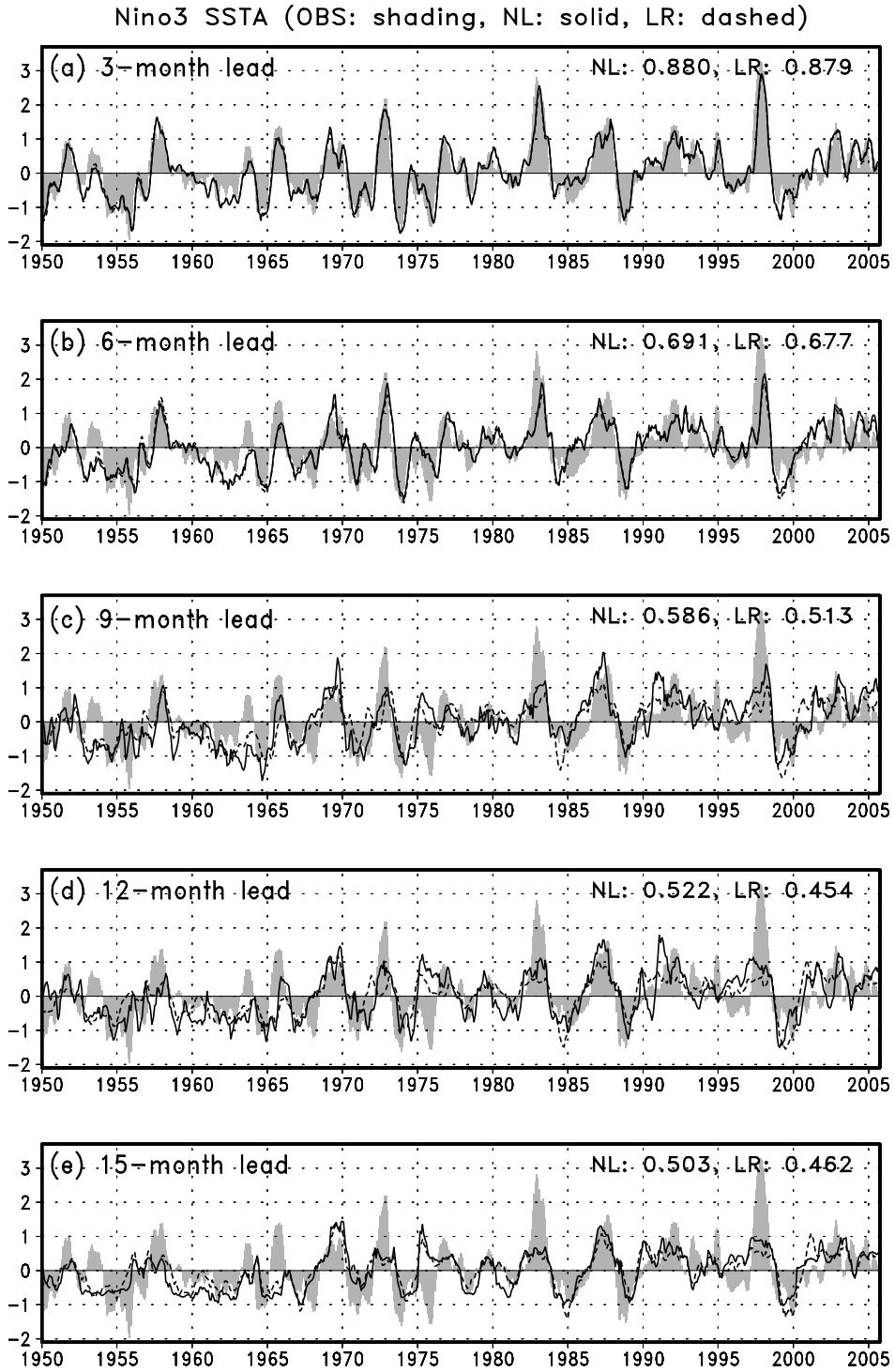

Figure 4: Time series of the Niño3 SST anomalies predicted by the NL models (solid curves) and LR models (dashed) at (a) 3-month, (b) 6-month, (c) 9-month, (d) 12-month and (e) 15-month lead time. The observations are indicated by the grey shading. In the upper-right corner of each panel, the forecast correlation skills for the NL models and LR models are given.

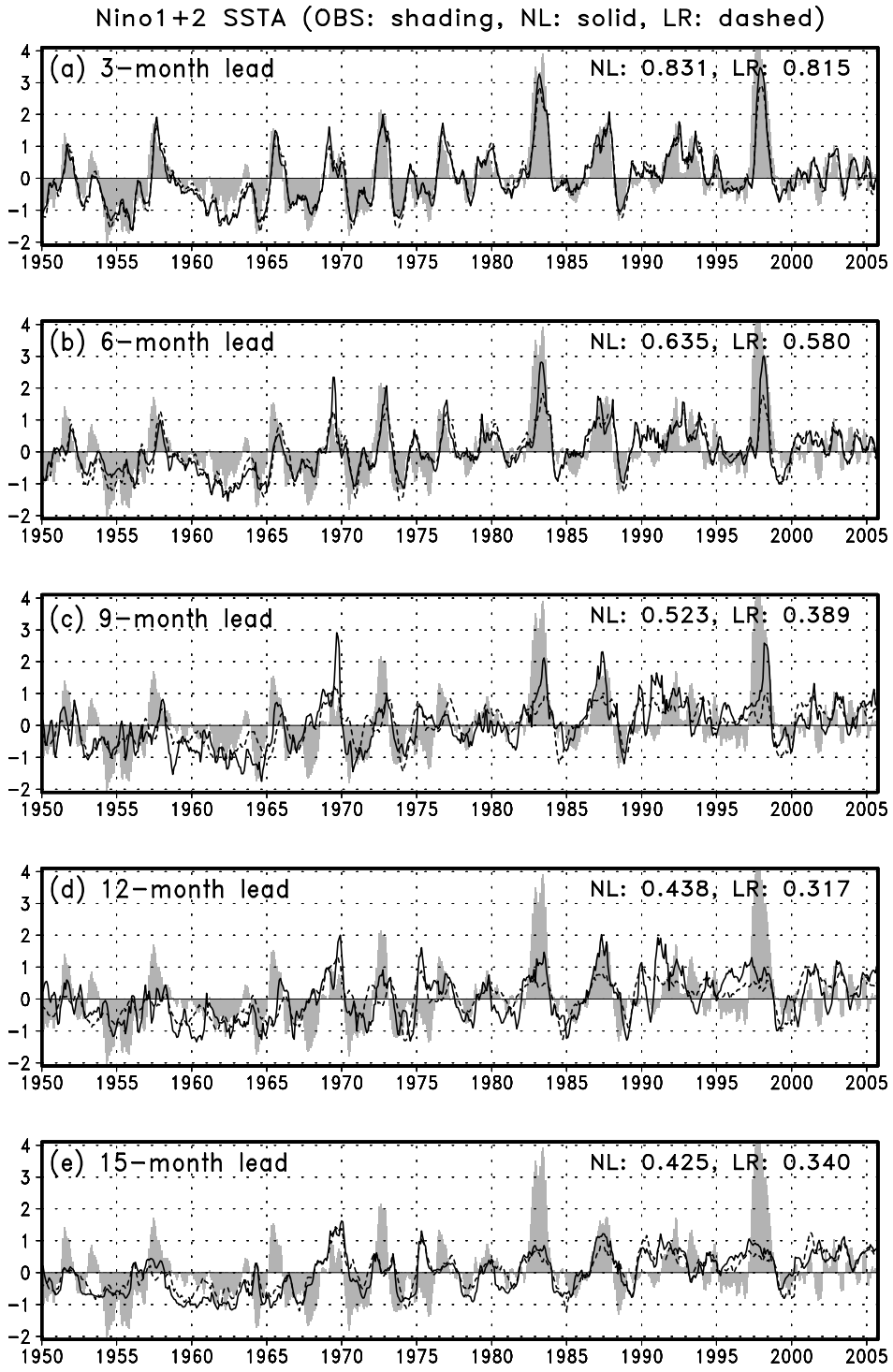

Figure 5: Similar to Fig. 4, but for the Niño1+2 SST anomalies.

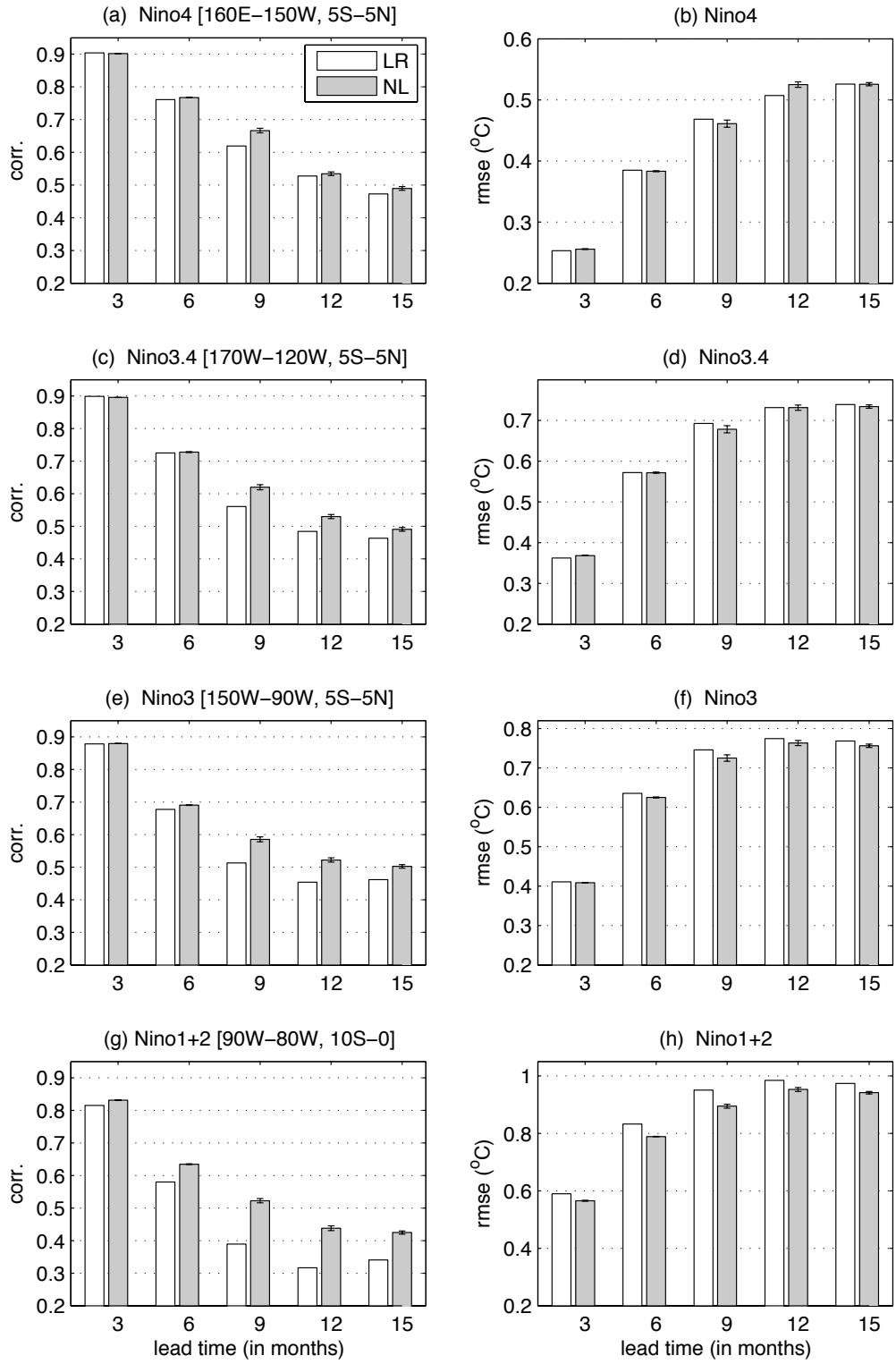

Figure 6: Correlation (left column) and RMSE (right column) skills of SST anomaly forecasts over the Niño4, Niño3.4, Niño3 and Niño1+2 regions. The LR skills are denoted by white vertical bars, and NL skills, by grey bars. The error bars indicate ±1 standard error of the forecast skills from the 30-member NL ensemble.

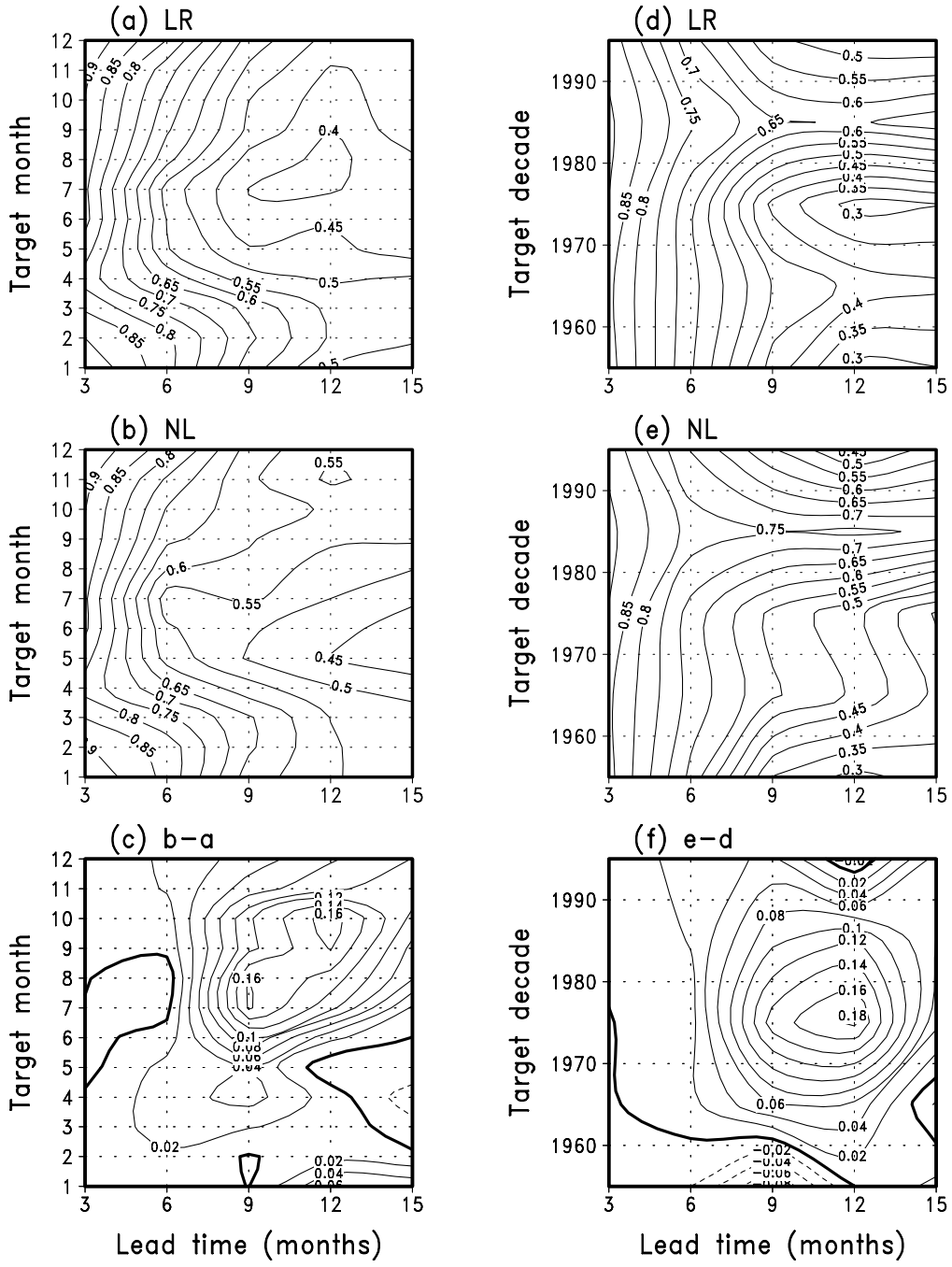

Figure 7: Correlation skills for the Niño3 SST index as a function of lead time and target month from (a) the LR models and (b) NL models, with the difference (NL minus LR) shown in (c). Similarly, the interdecadal changes of the Niño3 correlation skills are shown in the right column (panels d–f). The contour interval is 0.05 in the upper four panels, and 0.02, in the two bottom panels, with negative contours dashed and zero contours thickened.