Research Article

Forecasts of Tropical Pacific Sea Surface Temperatures by Neural Networks and Support Vector Regression

Silvestre Aguilar-Martinez1 and William W. Hsieh1,2

1 Department of Physics and Astronomy, University of British Columbia, Vancouver, BC, Canada V6T 1Z1
2 Department of Earth and Ocean Sciences, University of British Columbia, 6339 Stores Road, Vancouver, BC, Canada V6T 1Z4

Correspondence should be addressed to William W. Hsieh, whsieh@eos.ubc.ca

Received 24 May 2009; Accepted 28 September 2009

Recommended by Lakshmi Kantha

Two nonlinear regression methods, Bayesian neural network (BNN) and support vector regression (SVR), and linear regression (LR), were used to forecast the tropical Pacific sea surface temperature (SST) anomalies at lead times ranging from 3 to 15 months, using sea level pressure (SLP) and SST as predictors. Datasets for 1950–2005 and 1980–2005 were studied, with the latter period having the warm water volume (WWV) above the 20°C isotherm integrated across the equatorial Pacific available as an extra predictor. The forecasts indicated that the nonlinear structure is mainly present in the second PCA (principal component analysis) mode of the SST field. Overall, improvements in forecast skills by the nonlinear models over LR were modest. Although SVR has two structural advantages over neural network models, namely (a) no multiple minima in the optimization process and (b) an error norm robust to outliers in the data, it did not give better overall forecasts than BNN. Addition of WWV as an extra predictor generally increased the forecast skills slightly; however, the influence of WWV on SST anomalies in the tropical Pacific appears to be linear.

Copyright © 2009 S. Aguilar-Martinez and W. W. Hsieh. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The El Niño-Southern Oscillation (ENSO) phenomena, the strongest climatic 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), producing warm (El Niño) and cold (La Niña) episodes [1]. As ENSO affects not only the tropical climate but also the extratropical climate [2], the successful prediction of ENSO is of great importance.

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: dynamical models, statistical models, and hybrid (statistical-dynamical) models [3, 4].

Most of the statistical models used for ENSO forecast have been linear models, for example, multivariate linear regression (LR) and canonical correlation analysis (CCA) [5]. However, ENSO has nonlinear features—for instance, the cold SST anomalies during La Niña are centred much further west than the warm anomalies during El Niño. This asymmetry can be explained by nonlinear physics [6].

To better model the nonlinear features of ENSO, neural network (NN) models, originally from the field of computational intelligence, have been used for nonlinear regression [7, 8] and nonlinear CCA [9, 10]. In Wu et al. [8], nonlinear regression by NN was used to predict each of the five leading principal components (PCs) of the tropical Pacific SST anomalies 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 tropical Pacific.

NN methods, generally regarded as forming the first wave of breakthrough in machine learning, became popular in the late 1980s, whereas kernel methods (e.g., support vector regression SVR) arrived in a second wave in the second half of the 1990s [11, 12]. SVR has two advantages over NN models—it avoided the multiple minima problem associated...
with nonlinear optimization used in NN models, and robust error norms are used in SVR instead of the nonrobust mean squared error (MSE) norm.

The warm water volume (WWV) above the 20°C isotherm integrated across the equatorial Pacific basin has been shown to lead eastern equatorial Pacific SST anomalies by about 7 months [13]. This relation is consistent with the recharge-discharge oscillator theory of Jin [14].

In this paper, we test if the forecasting of tropical Pacific SST anomalies by Wu et al. [8] can be improved in two ways: (a) can SVR forecast better than NN, and (b) will adding WWV as an extra predictor improve the forecasts linearly or nonlinearly? Section 2 describes the data sets and Section 3 the data analysis methods. Results are given in Section 4, followed by summary and conclusion in Section 5.

2. Data

The warm water volume (WWV) index is defined as the volume integral of water above 20°C isotherm between 5°N–5°S, 120°E–80°W by Meinen and McPhaden [13]. The WWV index (for the period of January 1980 to December 2005) was obtained from the TAO Project Office of NOAA/PMEL, where ocean analysis data from the Australian Bureau of Meteorology Research Center (BMRC) were used to compute the index.

Monthly SST data at 2° × 2° resolution were obtained from Version 2 of the NOAA Extended Reconstructed SST (ERSST.v2). Data for the tropical Pacific region (124°E–70°W, 20°N–20°S) during the periods 1950–2005 and 1980–2005 were considered. Monthly sea level pressure (SLP) data available on a 2.5° × 2.5° global grid were downloaded from NCEP/NCAR Reanalysis [15]. The data used here extend from January 1948 to December 2005 (first set) and from January 1980, to December 2005 (second set) over the tropical Pacific Ocean (120°E–70°W, 20°S–20°N).

The reason two data sets of SST and SLP data were used, one covering the period 1950–2005 and the other 1980–2005, is because the WWV data were unavailable prior to 1980, hence the first data set did not contain WWV.

Removal of the climatological seasonal cycle and smoothing by a 3-month running mean were performed on the data to obtain the anomalies. Principal component analysis (PCA) was performed on the SST anomalies, with the five leading principal components (PC) retained. The first five PCs account for 59.9%, 10.8%, 8.0%, 3.5%, and 2.9%, respectively, of the variance for the 1950–2005 SST data set, and 60.3%, 15.0%, 8.0%, 3.1%, and 2.4% for the 1980–2005 data set. The PC time series for 1950–2005 are shown in Figure 1, while the corresponding spacial patterns, that is, the EOFs, are displayed in Figure 2. The first five SST PCs are the predictands for our nonlinear regression models; that is, a separate nonlinear regression model is built to predict each of the five SST PCs into the future by \( \Delta t \) (the lead time).

Our choice on the number of PCs to use for predictors and for predictands was based on Wu et al. [8], where A. Wu (pers. comm.) tested for the optimal number of PCs to use for the LR model (corresponding tests with NN models would be prohibitively expensive). For the predictor PCs, the number of PCs was varied between 1 to 20 for SST, and separately, 1 to 20 for SLP. For the number of predictand SST PCs, using more than 5 decreased the forecast skills of the LR model.

The predictors used are the first 9 and 7 PCs of SST and SLP, respectively, with the SLP PCs also supplied at time lags of 3, 6, and 9 months to provide information on the past evolution of the SLP PCs. Hence there is a total of 9 + 4 × 7 = 37 predictors. For NN and SVR, the 37 predictor time series were further compressed to 12 by a second PCA, also known as extended empirical orthogonal function (EEOF) analysis, or singular spectrum analysis (SSA) [8]. The first 12 SSA PCs were taken to be the final predictors in the NN and SVR models. The forecast lead time \( \Delta t \) is the time interval between the centre of the most recent predictor to the centre of the predictand period.

**Figure 1:** The first five principal components of SST (multiplied by \( 10^{-2} \)) from 1950 to 2005.
3. Nonlinear Regression Models

Following Hsieh [12], we outline the basic ideas behind the nonlinear generalization of linear regression (LR) by neural networks (NN) and support vector regression (SVR). First consider the simple linear regression problem with a single predictand variable $y$:

$$y = \sum_j a_j x_j + a_0$$  \hspace{1cm} (1)

with $x_j$ being the predictor variables, and $a_j$ and $a_0$ being the regression parameters. In the NN approach to nonlinear regression, nonlinear adaptive basis functions $h_j$ (also called hidden neurons) ($j = 1, \ldots, n_h$) are introduced, so the linear regression is between $y$ and $h_j$:

$$y = \sum_j a_j h_j(x; w) + a_0$$  \hspace{1cm} (2)

with typically

$$h_j(x; w) = \tanh \left( \sum_i w_{ij} x_i + w_{0j} \right).$$  \hspace{1cm} (3)

Since $y$ depends on $w$ nonlinearly (due to the nonlinear function tanh), the resulting minimization of the error function (basically the MSE) involves nonlinear optimization, with multiple minima in general, a major disadvantage of the NN approach.

For the NN model in (2) and (3), the model parameters $a_j$, $a_0$, $w_{ij}$, and $w_{0j}$ are grouped into a single weight vector $w$. The optimal $w$ is obtained through minimization of the regularized error function:

$$E = \frac{\beta}{2} \sum_{k=1}^{m} \left[ y(x_k; w) - y_{dk} \right]^2 + \frac{\alpha}{2} w^T w,$$  \hspace{1cm} (4)
Figure 3: Forecast correlation skills for the 5 leading PCs of SST. The white, black, and gray bars denote the skill for LR, SVR, and BNN, respectively. For every method, the left (right) bar denotes the score before (after) WWV has been included.
where the first term is the MSE scaled by the factor $\beta/2$, with $y_{\text{det}}$ denoting observed data, and the second term is a weight penalty or regularization term scaled by the factor $\alpha/2$. A relative large $\alpha$ would suppress the use of large weights, thereby making the NN model less nonlinear, hence avoiding overfitting to the noise in the data. In this paper, the hyperparameters $\alpha$ and $\beta$ are chosen based on the Bayesian framework, hence the name Bayesian NN (BNN) model [11]. The MATLAB code “trainbr” from the MATLAB Neural Network toolbox was used for the BNN model.

Finally, the following procedure was used to optimally select the number of hidden neurons ($n_h$). The data set was partitioned into ten segments of which nine were used for building the model, and one was reserved to provide independent data for testing the model forecasts. The nine segments for model building will be referred to as D9, and the single testing segment, D1. D9 was further partitioned into two sets, DT for training and DV for validation. Two-thirds of the data was chosen for training, and the remaining third for validation, with the data chosen at random in blocks of at least 12 months due to autocorrelation in the data.

With the value of $n_h = 1$, and a given random pick of DT and DV, the error function (4) was optimized a total of 50 times from random initial $w$. Since the error function contains many local minimum, there were potentially 50 different models. The model with the lowest MSE over the training set DT was kept, with the error named $\varepsilon_t$. This model was then applied to the corresponding validation set DV and the MSE over DV (named $\varepsilon_v$) was also computed. The set ($\varepsilon_t$, $\varepsilon_v$) characterized this particular choice of DT, DV, and $n_h$ = 1. Next, using the same DT and DV, the whole process was repeated with $n_h = 2$ to obtain ($\varepsilon_t^2$, $\varepsilon_v^2$). This process was carried out for $n_h$ ranging from 1 to 8, such that in the end there were 8 different sets of pairs ($\varepsilon_t^i$, $\varepsilon_v^i$); the $n_h$ corresponding to the lowest $\varepsilon_v$ was stored as $n_{h1}$. A different pair of DT and DV was randomly chosen from the same D9 and the process was repeated until $n_{h2}$ was determined. This was done a total of 20 times which implies the set $s = \{n_{h1}, \ldots, n_{h20}\}$. Finally, the mode of the set $s$ (i.e., the most commonly occurring value in the set $s$) was chosen as the optimal number of hidden neurons that represents the data set D9. With this choice of $\alpha$, a total of 50 models were trained using the entire D9 set and tested on D1. The forecast performance on D1 was evaluated in terms of the ensemble average of the 50 models. This entire process was repeated until all 10 segments had been tested.

Next, instead of adaptive basis functions $h_j(x; w)$ in (3), we use nonadaptive basis functions $\phi_j(x)$; that is, the basis functions do not have adjustable parameters $w$, with

$$y = \sum_j a_j \phi_j(x) + a_0. \quad (5)$$

For instance, Taylor series expansion with 2 predictors $(x_1, x_2)$ would have \{\phi_j\} = $x_1, x_2, x_1^2, x_1 x_2, x_2^2, \ldots$. The advantage of using nonadaptive basis functions is that $y$ does not depend on any parameter nonlinearly, and so only linear optimization is involved without the multiple minima problem. The disadvantage is that one generally needs a very large (or even an infinite) number of nonadaptive basis functions compared to relatively few adaptive basis functions to model a nonlinear relation, hence the dominance of NN models despite the local minima problem.
The high-dimensionality problem has finally been solved with the kernel trick, that is, although $\phi$ is a very high (or even infinite) dimensional vector function, as long as the solution of the problem can be formulated to involve only inner products like $\phi^T(x')\phi(x)$ (with superscript T denoting the transpose), then a kernel function $K$ can be introduced:

$$K(x', x) = \phi^T(x')\phi(x).$$  \hfill (6)

Instead of the unmanageable $\phi$, one now works only with the very manageable kernel function. From Bishop [11] or Hsieh [12], the nonlinear regression solution can be written in the following form:

$$y = \sum_{k=1}^{n} a_k K(x_k, x),$$  \hfill (7)

where there are $k = 1, \ldots, n$ data points $x_k$ in the training data set. If a Gaussian kernel function is used, that is,

$$K(x', x) = \exp\left(-\frac{||x - x'||^2}{2\sigma^2}\right),$$  \hfill (8)

Figure 5: Correlation scores before and after WWV was added to the predictor set for the Niño 4, Niño 3.4, Niño 3, and Niño 1+2 regions.
where $\sigma$ controls the width of the Gaussian function, then $y$ is simply a linear combination of Gaussian functions. These kernel methods have become popular for nonlinear regression, classification, and so forth. For nonlinear regression, the most common kernel method used is probably the support vector regression (SVR) method.

As mentioned earlier, SVR improves on NN methods by avoiding multiple local minima. It also uses a more robust error norm, the $\epsilon$-insensitive error:

$$E_\epsilon (y - y') = \begin{cases} 0, & \text{if } |y - y'| < \epsilon, \\ |y - y'| - \epsilon, & \text{otherwise.} \end{cases}$$  \hspace{1cm} (9)

Note that for outliers, that is, large $y - y'$, this error function behaves like the robust mean absolute error (MAE) instead of the nonrobust MSE used in NN models.

The SVR problem is solved by minimizing the following function:

$$F = C \sum_{k=1}^{n} E_\epsilon (y(x_k) - y_d k) + \frac{1}{2} ||a||^2,$$  \hspace{1cm} (10)

where $a$ contains the regression parameters in (5), and $C$ is an (inverse) regularization parameter. Details for solving the SVR problem are given in [11, 12]. The SVR code used was that developed by Chang and Lin [16].
The SVR model has three hyperparameters $C$, $\epsilon$, and $\sigma$ (controlling the width of the Gaussian kernel in (8)), which need to be determined by a grid search. SVR forecasts were tested by a tenfold cross validation as in the BNN and linear regression models; that is, the data set D1 was reserved for model forecast testing while D9 was used for model development. A sevenfold cross validation was used on D9, that is, 1/7 of D9, was reserved for model validation while 6/7 of D9 was used for training the model for given $C$, $\epsilon$, and $\sigma$ values. By rotating the training and validation data, the MSE was computed over validation data for all of D9, for each set of $C$, $\epsilon$, and $\sigma$ values, and the lowest validated MSE gave the optimal values for the three hyperparameters. The search for the hyperparameters was exhaustive; that is, a coarse grid search was first used to identify the region of minimum validated MSE, and then a finer grid search was used to pinpoint the optimal $C$, $\epsilon$, and $\sigma$ values.

4. Forecast Results

Individual forecasts of the first five PCs of SST were conducted using LR, BNN, and SVR. Forecasts were made at lead times of 3, 6, 9, 12, and 15 months, and the performance
of these methods was compared. The analysis was applied to two records, 1950–2005 and 1980–2005, with latter data set containing the warm water volume (WWV) index. Due to space limitation, we will concentrate on the latter data set. Details on the first data set are given in Aguilar-Martinez [17].

For the 1980–2005 record, from the training and validation procedure for BNN, out of a total of 250 cases (as there were 5 predictand PCs, 5 lead times and 10 validation segments, giving $5 \times 5 \times 10$ cases), 1 hidden neuron was selected in 230 cases, 2 hidden neurons were selected in 19 cases, and for only 1 case, more than 2 hidden neurons were selected (see Table 4.2 in [17]). In summary, two hidden neurons were predominantly selected only for lead times of 3 and 6 months for PC2 (but one hidden neuron was predominant at other lead times), while all other PCs predominantly selected only one hidden neuron. While one hidden neuron allows the BNN to model a nonlinear relation, more complex nonlinearity needs more than one hidden neuron.

The correlation skills evaluated over testing data for the SST PCs for the 1980–2005 record are shown in Figure 3. Two correlation scores are provided for each method, with the bar on the left being the score prior to the inclusion
of the WWV index as a predictor. For PC1, before WWV was included, BNN had slightly higher correlation scores than SVR at lead times of 9, 12, and 15 months, while SVR had marginally higher scores than BNN and LR at short lead times (3 and 6 months), (with hyperparameters chosen for the SVR given in Table 4.3 of [17]). Inclusion of the WWV index in the forecast of PC1 generally increased the correlation scores for all three models, while at the same time reducing or eliminating the differences between them. This increase in the correlation scores is expected given that the WWV along with SST is the main variables in most theoretical discharge/recharge oscillator models [14, 18, 19].

The nonlinearity of ENSO variability is manifested through PC2. The correlation skills for PC2 at lead times of 3–9 months are generally superior for the nonlinear methods, with SVR attaining the best performance (Figure 3). In general, adding the WWV index did not improve the scores for PC2. The scores for PCs 3–5 also showed that the inclusion of the WWV index did not significantly affect the forecast skills of these PCs.

To find out if the SST response to WWV could be governed by nonlinear dynamics, the performance of the models was tested using the WWV index as the sole predictor. The correlation scores for PC1 are displayed adjacent to the values
obtained using the full set of predictors in Figure 4(a), with the score bar on the right obtained from the single predictor (WWV) models. For these models with only the WWV predictor the correlation scores provided by linear regression were higher than nonlinear regression at all lead times except for 15 months, indicating a rather linear relation between SST and WWV, in accordance with Meinen and McPhaden [13] and McPhaden et al. [20]. As PC1 predictions of these models show relatively low scores at very short lead times (3 months), WWV is therefore a useful predictor mainly at intermediate lead times. The forecast of PC2 using WWV as the single predictor yielded low correlation scores for all models, as indicated in Figure 4(b), indicating that WWV was not useful for predicting PC2, the mode showing the largest degree of nonlinearity in Figure 3.

SST anomaly fields were reconstructed by summing the five predicted PCs multiplied by their corresponding EOF spatial patterns. SST anomalies averaged over the Niño 4 (160°E–150°W, 5°S–5°N), Niño 3.4 (170°W–120°W, 5°S–5°N), Niño 3 (150°W–90°W, 5°S–5°N), and Niño 1+2 (90°W–80°W, 10°S–0°) regions were then computed. The correlation skills and root mean squared error (RMSE), before and after WWV was introduced into the predictor set, are given in Figures 5 and 6, respectively, for the various Niño

Figure 10: SVR forecast performance and its difference from LR for the period 1980–2005 with WWV included as a predictor. The contour interval is 0.1 in the right column and 0.05 in the left.
regions. Before the WWV index was added, BNN provided a better correlation skill than SVR at lead times of 9–15 months for the Niño 3.4 and 4 regions. Incidentally, correlation skills for the Niño 4 region show the largest observed discrepancies among the different models, with BNN winning at longer lead times (12–15 months) and SVR at shorter lead times (3 and 6 months). Incorporating the WWV tended to reduce the differences in correlation among the 3 methods due to the mainly linear coupling between WWV and SST [18]. The advantage of the nonlinear methods is the most apparent in the region of largest variability, Niño 1 + 2, over short lead times of 3–9 months. However, the correlation scores for LR show improvements at longer lead times relative to the nonlinear methods. This is consistent with the forecast results of PC1 (Figure 4(a)), where LR compares well with the nonlinear methods at 12–15 months lead times.

Contour maps of the correlation scores between the predicted and observed anomalies for the nonlinear models are provided in Figures 7 and 8. The right column in these figures shows the difference in the correlation scores between the nonlinear model and LR. In Figures 7(d), 7(f), and 7(h), the region where BNN has slightly higher correlation skill than LR shifts westward along the equator as the lead time increases from 6 to 9 to 12 months. At lead times of 6 and 9 months, SVR attains correlation scores slightly higher than those for LR off the equator near the dateline (Figures 8(d) and 8(f)).

Although adding the WWV index as a predictor slightly enhanced the forecast skills, it reduced the difference in prediction skill between the nonlinear and linear models in some cases. For instance, the slight prediction advantage at lead times of 9–12 months of BNN relative to LR was reduced when WWV was added, as can be seen by comparing Figures 7(f) and 9(f) and Figures 7(h) and 9(h). The correlation score of SVR with WWV added is given in Figure 10. Comparing LR against itself after WWV was added reveals a slight increase in correlation skill emerging at long lead times of 9–15 months concentrated on the central-eastern equatorial region of the tropical Pacific (see Figure 4.15 in [17]).

The analysis was repeated for the longer record of 1950–2005. Compared with the results for the 1980–2005 period, the results from 1950–2005 showed an even smaller difference between the nonlinear and linear methods. This is not surprising since the ENSO episodes during 1950–1979 were weaker and less nonlinear than those from 1980 onward [21]. Hence there is less advantage in using nonlinear methods in the pre-1980 period.

5. Summary and Conclusion
The forecast results of the 5 leading PCs of SST indicate some evidence of nonlinear structure in the ENSO cycle. The nonlinear structure is mainly present in the second mode of the SST field (Figure 3). Overall, the improvement in forecast skills by the nonlinear models over LR was modest, as found by Tang et al. [22] and Wu et al. [8], with the nonlinear methods being most advantageous in the eastern region of Niño 1+2 at 3–9-month lead (Figure 5). As forecast lead time increased from 6- to 12-months, the region where forecast correlation skill for BNN appeared higher than that for LR shifted westward, from the eastern equatorial Pacific to the central and western equatorial Pacific. While SVR did not show significant overall forecast skill advantage relative to LR over BNN, the regions where SVR showed improved skill over LR were more off-equatorial and further west (near the dateline) than for BNN, at 6 and 9 month leads.

Although SVR has two structural advantages over NN models, namely, (a) no multiple minima as there is no nonlinear optimization involved, and (b) an error norm robust to outliers in the data, it did not give overall better forecasts than BNN. Presumably (a) the use of an ensemble average in BNN to deal with multiple solutions from multiple minima was quite adequate, and (b) the data set did not contain drastic outliers to utilize the advantage of the robust SVR model.

Addition of WWV as an extra predictor generally increased the forecast skills slightly; however, the influence of WWV on SST anomalies along the central-eastern tropical Pacific region appears to be linear as the difference between nonlinear and linear models often diminished when WWV was included. As the WWV embodies the large-scale low-frequency dynamics [20], the response of the SST field to the WWV index comes with a particular time delay, at about 6 to 12 months. This observation is in agreement with our results, where the maximum correlation between the SST field and WWV was recorded at around 6–12-month lead time.

Acknowledgments
The authors were supported by the Natural Sciences and Engineering Research Council of Canada and the Canadian Foundation for Climate and Atmospheric Sciences. The authors are grateful to Dr. Aiming Wu for his helpful comments. The WWV index of Meinen and McPhaden [13] was kindly provided by the TAO Project Office of NOAA/PMEL.

References
[1] S. G. Philander, *El Niño, La Niña, and the Southern Oscillation*, Academic Press, San Diego, Calif, USA, 1990.
[2] J. M. Wallace, E. M. Rasmusson, T. P. Mitchell, V. E. Kousky, E. S. Sarachik, and H. von Storch, “On the structure and evolution of ENSO-related climate variability in the tropical Pacific: lessons from TOGA,” *Journal of Geophysical Research C*, vol. 103, no. C7, pp. 14241–14259, 1998.
[3] A. G. Barnston, H. M. van den Dool, S. E. Zebiak, et al., “Long-lead seasonal forecasts—where do we stand?” *Bulletin of the American Meteorological Society*, vol. 75, no. 11, pp. 2097–2114, 1994.
[4] L. Goddard, S. J. Mason, S. E. Zebiak, C. F. Ropelewski, R. Basheer, and M. A. Cane, “Current approaches to seasonal-to-interannual climate predictions,” *International Journal of Climatology*, vol. 21, no. 9, pp. 1111–1132, 2001.
[5] A. G. Barnston and C. F. Ropelewski, “Prediction of ENSO episodes using canonical correlation analysis,” *Journal of Climate*, vol. 5, no. 11, pp. 1316–1345, 1992.
[6] S.-I. An and F.-F. Jin, “Nonlinearity and asymmetry of ENSO,” *Journal of Climate*, vol. 17, no. 12, pp. 2399–2412, 2004.

[7] F. T. Tangang, W. W. Hsieh, and B. Tang, “Forecasting the equatorial Pacific sea surface temperatures by neural network models,” *Climate Dynamics*, vol. 13, no. 2, pp. 135–147, 1997.

[8] A. Wu, W. W. Hsieh, and B. Tang, “Neural network forecasts of the tropical Pacific sea surface temperatures,” *Neural Networks*, vol. 19, no. 2, pp. 145–154, 2006.

[9] W. W. Hsieh, “Nonlinear canonical correlation analysis of the tropical Pacific climate variability using a neural network approach,” *Journal of Climate*, vol. 14, no. 12, pp. 2528–2539, 2001.

[10] A. J. Cannon and W. W. Hsieh, “Robust nonlinear canonical correlation analysis: application to seasonal climate forecasting,” *Nonlinear Processes in Geophysics*, vol. 15, no. 1, pp. 221–232, 2008.

[11] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, USA, 2006.

[12] W. W. Hsieh, *Machine Learning Methods in the Environmental Sciences*, Cambridge University Press, Cambridge, UK, 2009.

[13] C. S. Meinen and M. J. McPhaden, “Observations of warm water volume changes in the equatorial Pacific and their relationship to El Niño and La Niña,” *Journal of Climate*, vol. 13, no. 20, pp. 3551–3559, 2000.

[14] F.-F. Jin, “An equatorial ocean recharge paradigm for ENSO—part I: conceptual model,” *Journal of the Atmospheric Sciences*, vol. 54, no. 7, pp. 811–829, 1997.

[15] E. Kalnay, M. Kanamitsu, R. Kistler, et al., “The NCEP/NCAR 40-year reanalysis project,” *Bulletin of the American Meteorological Society*, vol. 77, no. 3, pp. 437–471, 1996.

[16] C.-C. Chang and C.-J. Lin, “LIBSVM: a library for support vector machines,” http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[17] S. Aguilar-Martinez, *Forecasts of tropical Pacific sea surface temperatures by neural networks and support vector regression*, M.S. thesis, University of British Columbia, Vancouver, Canada, 2008.

[18] A. J. Clarke, S. van Gorder, and G. Colantuono, “Wind stress curl and ENSO discharge/recharge in the equatorial Pacific,” *Journal of Physical Oceanography*, vol. 37, no. 4, pp. 1077–1091, 2007.

[19] F.-F. Jin, “An equatorial ocean recharge paradigm for ENSO—part II: a stripped-down coupled model,” *Journal of the Atmospheric Sciences*, vol. 54, no. 7, pp. 830–847, 1997.

[20] M. J. McPhaden, X. Zhang, H. H. Hendon, and M. C. Wheeler, “Large scale dynamics and MJO forcing of ENSO variability,” *Geophysical Research Letters*, vol. 33, no. 16, Article ID L16702, 2006.

[21] S.-I. An, Z. Ye, and W. W. Hsieh, “Changes in the leading ENSO modes associated with the late 1970s climate shift: role of surface zonal current,” *Geophysical Research Letters*, vol. 33, no. 14, Article ID L14609, 2006.

[22] B. Tang, W. W. Hsieh, A. H. Monahan, and F. T. Tangang, “Skill comparisons between neural networks and canonical correlation analysis in predicting the equatorial Pacific sea surface temperatures,” *Journal of Climate*, vol. 13, no. 1, pp. 287–293, 2000.
Submit your manuscripts at http://www.hindawi.com