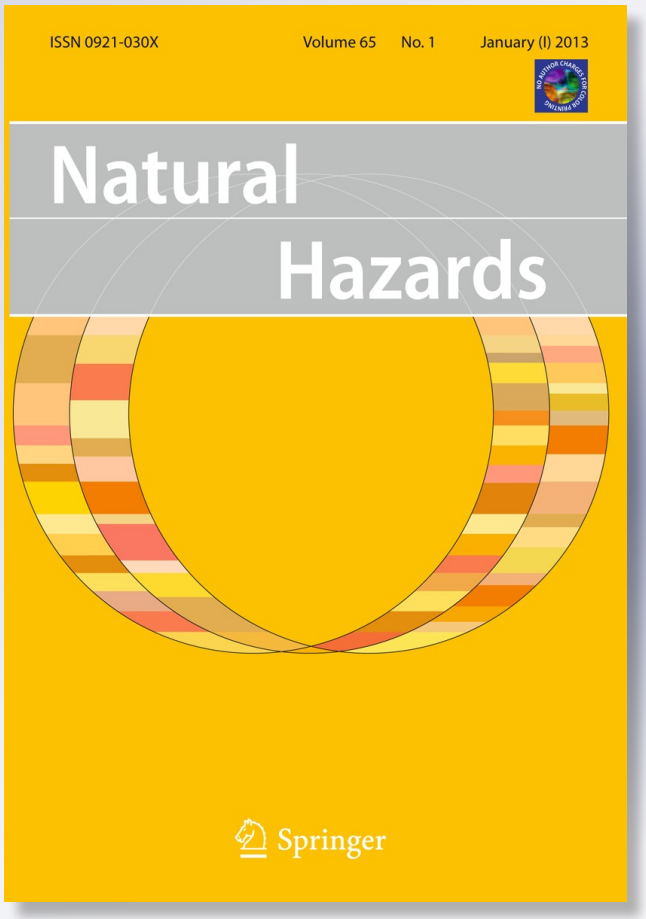
Prediction of tropical cyclone frequency with a wavelet neural network model incorporating natural orthogonal expansion and combined weights

Hexiang Liu, Da-Lin Zhang, Jianwei Chen & Qingjuan Xu

Natural Hazards
Journal of the International Society for the Prevention and Mitigation of Natural Hazards

ISSN 0921-030X
Volume 65
Number 1

Nat Hazards (2013) 65:63-78
DOI 10.1007/s11069-012-0343-x



Springer

Nat Hazards (2013) 65:63–78
DOI 10.1007/s11069-012-0343-x

ORIGINAL PAPER

Prediction of tropical cyclone frequency with a wavelet neural network model incorporating natural orthogonal expansion and combined weights

Hexiang Liu · Da-Lin Zhang · Jianwei Chen · Qingjuan Xu

Received: 21 June 2012 / Accepted: 2 August 2012 / Published online: 21 August 2012

Abstract A nonlinear wavelet neural network (WNN) model with natural orthogonal expansion (NOE) and combined weights is constructed to predict the annual frequency of tropical cyclones (TCF) occurring over the coastal regions of Southern China. Combined weights are obtained by calculating categorical weights, based on the particle swarm projection pursuit, and ranking weights, based on fuzzy mathematics, followed by optimization. The global monthly mean heights at 500 hPa and sea-surface temperature fields are used as two predictors. The linear and nonlinear information of the predictors with reduced dimensions is gathered through the NOE and combined weights, respectively, and treated as the input into the WNN model. This model is first trained with the 55-year (i.e., 1950–2004) TCF data and then used to predict annual TCFs for the subsequent 5 years (i.e., 2005–2009). Results show that the mean absolute and relative errors are 0.6175 and 9.34 %, respectively. The impacts of the combined weights, NOE and WNN as well as the traditional multi-regression approach on the TCF prediction are examined. Results show superior performance of the WNN-based model in the annual TCF prediction.

Keywords Wavelet neural network · Statistical prediction · Tropical cyclone frequency · Combined weights

1 Introduction

Fuzzy mathematics and nonlinear intelligent computation techniques have been widely applied to many scientific and engineering fields during the past two decades (Jin et al. 2000; Iliadis and Spartalis 2005; Song et al. 2006; Bingham and Zhang 2007; Wang and

H. Liu · J. Chen · Q. Xu
School of Mathematical Sciences, Guangxi Teachers Education University, Nanning 530023, Guangxi, People's Republic of China

D.-L. Zhang (✉)
Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742-2425, USA
e-mail: dalin@atmos.umd.edu

Bao 2009). Proper treatment of linear and nonlinear processes and construction of nonlinear mathematical models, especially in optimizing predictors in those models, have attracted considerable attention in order to improve their predictive skills.

Yu et al. (2005) showed the prediction of 24-h rainfall from tropical cyclones (TCs), based on their tracks, intensities, sizes and movements, with a fuzzy multi-targets algorithm. Liu and Zhang (2012) provided reasonable estimations of the hazard risks associated with TCs affecting Southern China using a fuzzy information spread technique and a gray hazard-year prediction model after constructing hazard damage indices with the synthesized linear processing and combined weights method. Lazarovitch et al. (2009) showed the prediction of ground-water distribution under trickle irrigation using an artificial neural network (ANN) model. Zhu et al. (2003) studied the effects of the Madden-Julian Oscillations on the development of TCs over the Indian and West Pacific oceans with the empirical orthogonal functional analysis. Jin et al. (2005) addressed the "overfitting" problem in ANN-forecast models and improved the associated generalization capability by constructing a low-dimensional ANN learning matrix through principal component analysis. Chen et al. (2008) found optimal projection directions in a projection pursuit model by combining a particle swarm optimizing algorithm and penalty function. Yao et al. (2009) examined the monthly and seasonal monsoon intensity index prediction through the construction of a nonlinear genetic neural network ensemble prediction model because of the nonlinear nature involved.

Because of the limited skill in the seasonal to inter-annual prediction of TC tracks by meteorological models, many past studies have developed various statistical forecast models in an attempt to predict the frequency of TCs (TCF). For instance, Xiao and Xiao (2010) examined the characteristics of TCFs occurring over Northwest Pacific and their landfalls over China using the TC data during the years of 1951–2008. Kwon et al. (2007) conducted the ensemble prediction of TCFs for the Northwestern Pacific basin. Yin et al. (2010) applied a back-propagation neural network (BPNW) model to the prediction of annual TC occurrences over Northwest Pacific. Lu et al. (2003) used an ANN model, in which mean generating functions and stepwise regression analysis are incorporated, to analyze the annual TCFs affecting Guangxi Province. Ying and Wan (2011) studied the prediction of the annual TCF over China with an optimized regression model, in which a correlative searching method is used to screen predictors and then their multi-collinearity is removed after performing principal component analysis.

While the previous studies have demonstrated considerable success in individually applying the algorithms of fuzzy mathematics, linear principal component analysis, stepwise regression analysis, nonlinear particle pursuit and BPNW to the hazard-risk analysis and TCF prediction, few studies have integrated all the strengths associated with the above individual algorithms for the prediction of TCFs. Thus, it is the intention of this study to fill in the gaps in order to improve our predictive capability of TCFs. Specifically, the objectives of this study are to (1) construct a nonlinear wavelet neural network (WNN) model by coupling natural orthogonal expansion (NOE) and combined weights and (2) apply this WNN-based model to the prediction of annual TCF for the coastal regions of the Southern China. To achieve the above objectives, we define the global monthly mean 500-hPa height (MH500) and sea-surface temperatures (MSSTs) in the 12 months prior to a TC season as two major predictors for the TCF of the TC season. The TCF data and the global MSST and MH500 data during the 56 years of 1949–2004 will be used as a training sample, and the subsequent 5 years of 2005–2009 will be utilized to validate the predicted TCFs with the WNN-based model.

The next section presents the use of linear and nonlinear dimensional reduction algorithms to separate the linear and nonlinear information of the predictors, respectively. Section 3 describes the construction of the WNN model and shows how a multi-dimensional nonlinear dataset can be projected onto a lower-dimensional space. Section 4 shows the application of the WNN-based model to the prediction of the annual TCFs affecting Southern China after model training with the TCFs, MH500 and MSSTs data during the 56 years of 1949–2004. The impact of combined weights, NOE and WNN as well as the traditional multi-regression approach on the TCF prediction will also be examined.

2 Linear and nonlinear dimensional reduction

In constructing our TCF prediction model, an important first step is to perform the linear and nonlinear dimensional reduction of the preprocessed data associated with the two predictors through NOE and combined weights, respectively.

2.1 Linear dimensional reduction

While multi-indicator variables could provide rich information describing an event, they often increase the workload of data collection and subsequent data processing. In addition, because of certain inter-relationship among these variables, some multi-collinearity issue may arise. In this regard, NOE could identify a predictor matrix with the fewest possible indicators containing concise linear information to replace a multi-indicator predictor matrix, thereby reducing the matrix dimensions. In particular, since the principal components of NOE are orthogonal to each other, they do not have any multi-collinearity influence. Moreover, NOE can extract true signals from the original matrix that are often contaminated by random noises, because the principal components are relatively less sensitive to random noises.

Suppose that a predictor matrix $\mathbf{X}$ (i.e., MH500 or MSST in the present case) can be expressed in terms of a time function $\mathbf{Z}$ and a space vector $\mathbf{V}$, that is, $\mathbf{X} = \mathbf{VZ}$, the decomposition of $\mathbf{X}$ can be achieved by the principal component decomposition method (Jin et al. 2005). Specifically, for a given predictor matrix,

$$\mathbf{X}' = \left(x'_{ij}\right)_{n\times m}, \tag{1}$$

a normalization procedure may be performed using

$$x_{ij} = \frac{x'_{ij} - \bar{x}_i}{s_i}, \quad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m),$$

where $\bar{x}_i, s_i$ are the i-th row sample mean and standard deviation, respectively (Xie and Liu 2006), and the normalized predictor matrix is written as

$$\mathbf{X} = (x_{ij})_{n\times m}, \tag{2}$$

Its covariance matrix is

$$\mathbf{S} = \frac{1}{n}\mathbf{X}\mathbf{X}^{\mathrm{T}} \tag{3}$$

where $\mathbf{X}^{\mathrm{T}}$ is the transposed matrix of $\mathbf{X}$. By obtaining eigenvalues of the real symmetric matrix $\mathbf{S}$, that is, $\lambda_1, \lambda_2, \ldots, \lambda_m$ ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$), and their corresponding eigenvector, $\mathbf{V} = (v_1, v_2, \ldots, v_m)$, we can express each principal component as the linear combination of the original multi-indicator variables, that is,

$$z_i = v_{i1}x_1 + v_{i2}x_2 + \cdots + v_{in}x_n, \tag{4}$$

or a more general matrix expression

$$\mathbf{Z} = \mathbf{V}^{\mathrm{T}}\mathbf{X} = (z_1, z_2, \ldots, z_n)^{\mathrm{T}} \tag{5}$$

With the above procedures, we can significantly reduce the number of indicators in the original data and treat these principal component predictors as part of the model input.

2.2 Nonlinear dimensional reduction

The combined weights method, coupling a nonlinear data processing technique and fuzzy mathematical theory, is capable of reducing a multi-dimensional vector space to a lower-dimensional space such that data analysis can be markedly simplified. To obtain the combined weights of a predictor, say matrix (1), its relative membership values r_{ij} need to be first calculated through the membership weighting function,

$$r_{ij} = \frac{x_{ij} - \min\limits_j x_{ij}}{\max\limits_j x_{ij} - \min\limits_j x_{ij}} \quad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m) \tag{6}$$

where $\max\limits_j x_{ij}$ and $\min\limits_j x_{ij}$ are the maximum and minimum values of the j-th indicator (Xie and Liu 2006), respectively. This will yield a fuzzy weighted evaluating TCF matrix $\mathbf{R} = (r_{ij})_{n\times m}$ after standardizing the predictor sample x_{ij}. Then, categorical and ranking weights, based on the projection pursuit-particle swarm optimization algorithm and fuzzy mathematical theory, respectively, will be calculated, as detailed below, and optimized to obtain the combined weights (Jin et al. 2003; Liu and Zhang 2012).

2.2.1 Determining the categorical weights $A_1 = \{a_{i1},\ i = 1, 2, \ldots, n\}$

Projection pursuit is a statistical method that is typically used to process the nonlinear, non-normal distributed data in a multi-dimensional space. It will give the projection direction that characterizes the multi-dimensional data structures or properties, and reduce a multi-dimensional space problem to a lower-dimensional one, thereby simplifying our analysis and understanding. In mathematical terms, this is equivalent to projecting points in a multi-dimensional space to lower-dimensional one while keeping the categorical information of predictors in their original multi-dimensional space. In the present case, the projection pursuit method begins from synthesizing the original multi-indicator matrix to a one-dimensional projection series along the projected direction $A_1 = \{a_{i1};\ i = 1, 2, \ldots, n\}$ with the membership weighting matrix r_{ij} ($i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, m$),

$$r_{ij} = \frac{x_{ij} - \min_j x_{ij}}{\max\limits_j x_{ij} - \min\limits_j x_{ij}} \quad (i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m) \tag{7}$$

To determine the projected direction, it is necessary to ensure the one-dimensional distribution of the projection points locally as dense as possible but globally with the clustered points as widely spread as possible. To this end, we define the projection objective function:

$$Q_a = S_z \cdot D_z, \tag{8}$$

where S_z is the standard deviation of z_j,

$$S_z = \sqrt{\frac{\sum_{j=1}^{m} (z_j - \bar{z})^2}{m-1}}, \tag{9}$$

and D_z is the local density of the projected points,

$$D_z = \sum_{i=1}^{n} \sum_{j=1}^{n} (K - d_{ij}) u(t) (K - d_{ij}). \tag{10}$$

In Eq. (10), K is the window radius of the local density; $d_{ij} = |z_i - z_j|$ is the distance between points; $t = K - d_{ij}$; and the unit step function $u(t)$ is 1 for $t \geq 0$, and 0 for $t < 0$. When selecting K, it is necessary to ensure that the average number of the projected points within the window will not be too small to avoid the generation of large moving-average errors, and meanwhile, it will not increase too fast with increasing n. Thus, we take an empirical value of $K = 0.1S_z$, following Jin et al. (2003).

Given a predictor sample, the projection objective function Q_a changes with projection direction, and different projection directions will reflect different data characteristics. An optimized projection direction should capture the most important characteristics of the multi-dimensional data, so the projection direction can be estimated by maximizing the projection objective function:

$$\begin{cases} \max \ Q_a = S_z \cdot D_z \\ s.t. \sum_{i=1}^{n} a_i^2 = 1 \quad \text{and} \quad a_i \geq 0 \end{cases} \tag{11}$$

where *s.t.* is an acronym of "subject to". To solve the complicated nonlinear optimization problem using $A_1 = \{a_{i1}; i = 1, 2, \ldots, n\}$ as the optimization series, we define the fitness function as

$$f(a_{i1}) = Q_a. \tag{12}$$

Next, a particle (or point) swarm optimization algorithm is employed to optimize the projection direction (Bonabeau et al. 2000). In this algorithm, the fitness function (12) regulates the search direction, according to each particle's position and velocity. Each particle can pursuit its current optimal particle by memory through its interaction with others, and it can also keep finding an optimal region in a complicated solution space. If a better solution is found, it will be used as a base to search the next solution. Every interactive process is not entirely random, but the finding and updating solution processes are all optimized. This optimization algorithm will allow one to find k units with high fitness from the particles that have evolved to the last generation, and obtain k better-projected directions as the categorical weights of the predictors $A_1 = \{a_{i1}, i = 1, 2, \ldots, n\}$ after dimensional reduction.

2.2.2 Determining ranking weights $A_2 = \{a_{i2}, i = 1, 2, \ldots, n\}$

The analytic hierarchy process in fuzzy mathematics treats the variable under study (e.g., TCFs herein) as a system, and the essence of system evaluation is a process of optimizing and ranking. In general, if an element of r_{ij} in $\mathbf{R}$ is large, so is its ranking influence on the

predictor. That is, given the i-th indicator of the predictor matrix $\mathbf{R}$, if the algebraic sum s_i along the i-th line is large, so is its optimized ranking influence on the predictor. In this study, we derive the ranking information of the predictors' sample sets, with some techniques from Jin et al. (2003), and then determine each predictor's ranking weights.

Specifically, we start from the fuzzy evaluation matrix $\mathbf{R} = (r_{ij})_{n\times m}$, and let $s_i = \sum_{j=1}^{m} r_{ij}$. From

$$c_{ij} = \begin{cases} \frac{s_i - s_j}{\max\{s_i\} - \min\{s_i\}}(c_m - 1) + 1, & s_i \geq s_j \\ \frac{1}{\frac{s_i - s_j}{\max\{s_i\} - \min\{s_i\}}(c_m - 1) + 1} & s_i < s_j \end{cases} \tag{13}$$

where $\max\{s_i\}$ and $\min\{s_i\}$ are the maximum and minimum values of s_i, respectively, and $c_m = \min\{9, \text{int}[\frac{\max\{s_i\}}{\min\{s_i\}} + 0.5]\}$ is an important reference constant, we construct a decision matrix $\mathbf{C} = (c_{ij})_{n\times m}$ and obtain a level 1–9 decision matrix $\mathbf{C}$ to be used to calculate the ranking weights of each predictor. The eigenvalues of $\mathbf{C}$ are used to calculate consistency index (**CIC**). If **CIC** < 0.10, it means that the decision matrix meets the consistency requirements, and the ranking weights $A_2 = \{a_{i2},\ i = 1,\ 2,\ \ldots,\ n\}$ can be obtained. If it does not, an optimization algorithm in MATLAB 7.0 may be used to obtain a modified decision matrix until it meets the consistency requirements.

2.2.3 Determining combined weights A = {a_i, i = 1, 2, …, n}

The combined weights $A = \{a_i,\ i = 1,\ 2,\ \ldots,\ n\}$ can be obtained by optimizing the objective function,

$$\begin{aligned} \min F &= \sum_{j=1}^{m}\sum_{i=1}^{n}(\mu|a_{i1} - a_i|r_{ij} + (1-\mu)|a_{i2} - a_i|r_{ij}) \\ s.t. &\sum_{i=1}^{n} a_i = 1 \quad \text{and} \quad a_i \geq 0, \quad i = 1, 2, \ldots, n \end{aligned} \tag{14}$$

where we may take the categorical coefficient as $\mu = 0.5$, assuming that each categorical weight has the same reference value. So, the original multi-dimensional nonlinear predictor matrix $\mathbf{X}$ can now be simplified to a one-dimensional linear predictor series $\mathbf{A}$, that is,

$$\mathbf{Z}^* = \mathbf{R}\,\mathbf{A} = (z_{11}^*, z_{21}^*, \ldots, z_{n1}^*)^T \tag{15}$$

This one-dimensional predictor series should be highly correlated with the prognostic variable under study (e.g., TCF), and it will be used as part of the input into the WNN-based prediction model to be derived in the next section.

3 Construction of a WNN-based prediction model

Numerous studies showed that there are often complicated linear and nonlinear relationships between prognostic variables and predictors. Thus, it is desirable to capture the key linear information in the original multi-indicator array of each predictor, and meanwhile gather as much nonlinear information as possible. As mentioned in the preceding section, the linear information in the original data array is obtained by performing NOE twice, while its nonlinear information is gathered by performing NOE once, followed by the

combined weights derived through the particle pursuit and fuzzy mathematics. These algorithms are relatively simple and easy to execute, but their resulting functions are powerful with few parameters to be adjusted.

The generalization of a nonlinear model is another key factor in evaluating how well the model performs. A WNN model typically couples the wavelet transformation with BPNW, and it replaces the S-type activation function in BPNW with a transfer function that uses the wavelet basis function as hidden nodes. Thus, such a WNN model contains superior properties of the functional approximation and error tolerance from wavelet transformation, and the self-learning ability from neural network. Unlike a normal neural network prediction model, such a WNN model could not only determine the network structures objectively, but also has a fine generalization performance.

In this study, we intend to develop a nonlinear prediction model by coupling NOE and combined weights with a WNN model, which takes advantage of the strengths in NOE and combined weights and all the superior properties of a WNN model. Specific steps in constructing this WNN-based model are given as follows:

i. Generate randomly the stretch coefficient e_j and the transfer coefficient p_j of the wavelet function in the range [0, 1], design a weighting function w_{ij} associated with the input and hidden layers and a weighting function w_{jk} associated with the hidden and output layers, and specify the overall convergence error of the model.
ii. Determine the number of nodes for the respective input, output and hidden layers, based on the sample characteristics.
iii. Determine the training and prediction samples through correlation analysis: (a) form a new data series after applying NOE and correlation analysis to the original data array; (b) gather numerous indicators after performing NOE on the new data series; (c) obtain combined weights after applying the particle pursuit and fuzzy mathematics algorithms to the new data series, followed by optimization; and (d) gather further key indicators after performing dimensional reduction for the new data series with the combined weights. The final predictors associated with MH500 and MSST found in (b) and (c) will be used as the input into the WNN-based model.
iv. Given the input data series x_i ($i = 1,2, \ldots, k$), the output of the hidden layer should be

$$g_j(x) = g_j\left(\frac{\sum_{i=1}^{k} w_{ij}x_i - p_j}{e_j}\right) \quad (j = 1, 2, \ldots, l), \tag{16}$$

where $g(j)$ is the j-th node output value of the hidden layer; e_j and p_j are the stretch and transfer coefficients of the wavelet basis function g_j, respectively. We use the Morlet mother wavelet basis function for $g_j(x)$, that is, $g_j(x) = \cos(1.75x)\mathrm{Exp}(-x^2/2)$, which gives better results, based on numerous model experiments.

v. A nonlinear WNN-based prediction model is finally constructed after obtaining outputs from the WNN model $y_k = \sum_{i=1}^{l} w_{ik}g(i), \quad (k = 1, 2, \ldots, m)$ and performing the WNN-integrated training, where l and m are the number of nodes in the hidden and output layers, respectively.

Figure 1 shows the flow chart of the above-mentioned steps starting from the data processing to dimensional reduction, model training and the prediction of the annual TCFs using the WNN-based model.

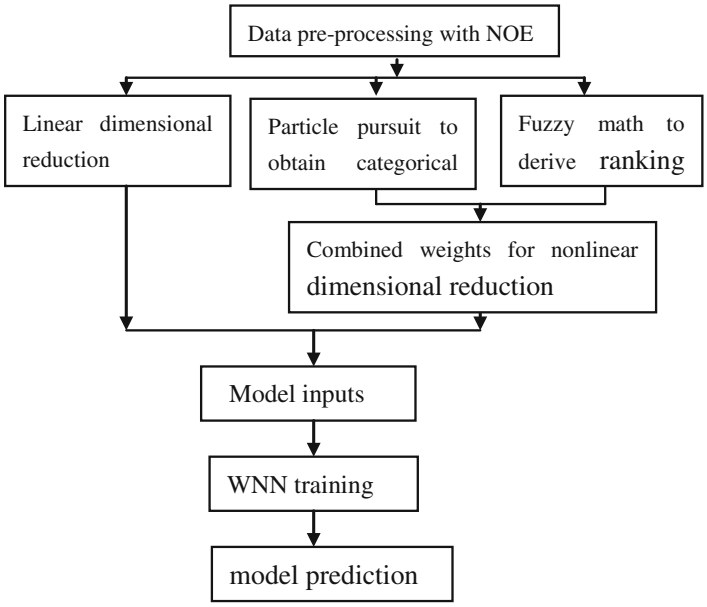


Fig. 1 The flow chart of a nonlinear WNN prediction model by coupling NOE and combined weights

4 Model application and analysis

In this section, we apply the WNN-based prediction model constructed in the preceding section to the TCF prediction for the coastal areas of Southern China consisting of the Provinces of Guangdong, Guangxi and Hainan. These regions typically suffer the most frequent passage, the most severe impact and the longest duration of TCs in China. However, the annual TCFs over these areas vary substantially from year to year, for example, from the highest 9 to the lowest 1 during the past 61 years (i.e., 1949–2009; see Tables 1 and 2). In addition, the relationships between the TCFs and predictors are complicated, involving both linear and nonlinear processes. Thus, the above-mentioned variable annual TCFs and complicated relationships provide a severe test to examine the reliability of the WNN-based model in predicting the TCFs over Southern China.

4.1 Data source and processing

For this study, we treat the landfalling TCF over Southern China as the prognostic variable. Several TCF datasets collected during the years of 1949–2009 are used, that is, including those from Typhoon Annual Book for years 1949–1988 and Tropical Cyclone Annual Book for years 1989–2009 (see Tables 1, 2). The global MH500 and MSST fields, used as the two major predictors, are from the National Centers of Environmental Prediction (NCEP) 1°-resolution reanalysis. Statistically, we may expect the MH500 field to roughly indicate the relative strength of the subtropical high in determining the tracks of TCs through the Southern China (Chen and Ding 1979; Deng et al. 1999), and the MSST field to reveal TCFs and the genesis/intensity of TCs over the region under study (Aberson 1998; Kaplan and DeMaria 1999; Wu et al. 2010). We treat the first 56-year (i.e., 1949–2004) data (but for 55-year TCFs, i.e., 1950–2004) as the model-training sample, and the remaining 5-year (i.e., 2005–2009) data as the independent prediction sample.

Table 1 The observed TCFs and the trained TCFs with the WNN-based model during the 55-year period (i.e., 1950–2004)

Years	Observed	Trained	Years	Observed	Trained	Years	Observed	Trained	Years	Observed	Trained
			1963	6	6.1902	1977	3	2.8954	1991	6	5.3859
1950	1	2.0528	1964	5	5.1439	1978	4	4.5821	1992	8	7.2607
1951	2	1.2909	1965	8	7.8101	1979	4	4.5139	1993	6	6.6609
1952	7	6.0542	1966	4	4.6601	1980	8	7.9323	1994	9	7.6595
1953	5	6.5785	1967	7	7.4038	1981	7	7.3529	1995	8	7.8002
1954	2	2.5254	1968	5	3.9459	1982	4	3.8078	1996	7	6.3232
1955	4	4.63	1969	4	4.7888	1983	5	4.3307	1997	3	3.0007
1956	5	4.6008	1970	4	5.8127	1984	6	6.5517	1998	3	3.1643
1957	5	5.6878	1971	9	8.3335	1985	5	5.145	1999	5	4.6195
1958	6	5.6701	1972	2	3.012	1986	6	7.1083	2000	4	4.2935
1959	4	4.567	1973	8	6.2411	1987	5	4.671	2001	7	7.648
1960	6	4.7647	1974	8	7.6208	1988	5	5.0177	2002	6	4.7627
1961	7	6.3504	1975	5	5.8247	1989	7	7.1172	2003	4	5.1775
1962	7	6.717	1976	5	5.1378	1990	6	7.6749	2004	5	5.0741

Table 2 Comparison of the WNN-predicted annual TCF after averaging 30 forecast tests to the observed one and their associated absolute and relative errors during the years of 2005–2009

Years	Observed	Predicted	Absolute error	Relative error (%)
2005	6.0000	7.4531	1.4531	24.22
2006	4.0000	3.9294	−0.0706	1.76
2007	5.0000	4.6248	−0.3752	7.50
2008	9.0000	7.9062	−1.0938	12.15
2009	9.0000	9.0950	0.0950	1.06
Average	6.6000	6.6017	0.6175	**9.34**
Standard deviation	2.3022	1.8663	0.5253	**9.47**

The 5-year averaged errors and standard deviations are also given

Before applying the WNN-based prediction model, we calculate first the correlations between the above-mentioned 55 TCFs in a series and the 55 MH500 values in a series from each month of a 12-month period prior to the June month of individual TC seasons (e.g., from June to December in 1949 and January to May in 1950 for the TCF of 1950) for the 55 years (i.e., 1949–2003) at each (1° × 1°) grid box. Repeat the above procedures for MSST. This will initially give 12 global maps of the correlations associated with each predictor. Figure 2a, b shows examples of the correlations between the TCFs and MH500, and the TCFs and MSST for the month of Septembers, respectively. In these maps, we are only concerned with the grid boxes at which (a) the absolute correlation coefficients are greater than 0.32 (i.e., reaching the 98 % significance level) and (b) threshold (a) is met for a continuous coverage of at least 30 grid boxes. For the examples given in Fig. 2a, b, 64 and 157 grid boxes meet the specified thresholds, respectively.

With the high-correlated grid boxes obtained above, their NOE analysis of the associated correlation arrays will give new correlations between the 55-year TCFs and numerous (dimensionless) principal components. We pick 27 principal components with the absolute correlation coefficient of greater than 0.20 as preliminary indicators, 22 of which are associated with MH500 in the present case. Although only five preliminary

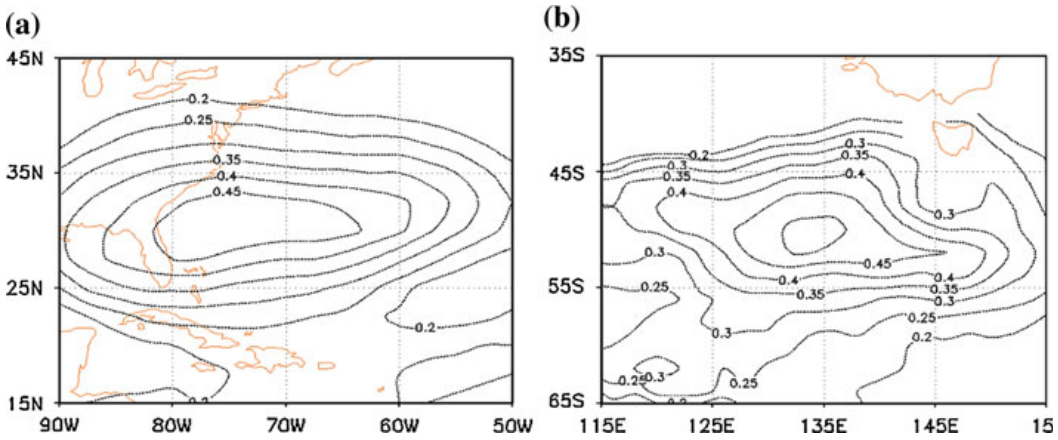


Fig. 2 Horizontal distribution of the correlation coefficients (**a**) between the 55-year TCFs and the 55-year MH500 in the month of Septembers prior to individual TC seasons over the southeastern US-Atlantic region; and (**b**) as in (**a**) but for MSST over the Southern Indian Ocean. Only the areas of significance are shown

indicators are associated with MSST, their contributions to the 55-year TCFs are not small. That is, their correlations with the TCFs are as high as −0.54, whereas 14 preliminary indicators associated with MH500 are less than 0.3. These 27 indicators with the highly correlated principal components will be treated as the model-training sample. Clearly, this will ensure the high correlation of the two predictors with the TCFs during the 55 years.

4.2 Prediction of annual TCF for Southern China

The above 27 preliminary indicators will be reduced to nine indicators after applying stepwise regression, and then performing the NOE of the nine indicators will yield three principal components that have both the high variances (i.e., 19.3075, 15.2321 and 9.9372) and the high correlation between the TCFs and the predictors (0.6185, −0.3560 and −0.3356) for the 56 years under study. Meanwhile, to effectively capture the nonlinear information of each predictor, the number of the selected nine indicators will be reduced to one through the combined weights, in which the correlation coefficient between this indicator and the TCFs is 0.4433. A total of four principal indicators, that is, the three principal components after the NOE plus one principal component after applying the combined weights, will be used as the model input into the WNN model. The 56-year data sample during the years of 1949–2004 will be network-trained via the WNN prediction model.

According to Kolmogorov's mapping network theorem (Krylov 2002), a three-layer network can accomplish the projection from any n to m dimensions. Thus, we decide to construct such a three-layer network with one hidden layer. In general, too many nodes in a hidden layer will affect network performance, leading to too long a learning process, whereas too fewer nodes could cause a bad network error tolerance. Our initial experimentation indicates that six nodes are appropriate for the hidden layer, giving rise to a 4-6-1 network structure. During the network training, we take the learning probability of the neural network weight and the wavelet parameter as 0.01 and 0.001, respectively (Shi et al. 2010; Jin 2004), and set the sum of errors, that is, $\sum|\text{actual value} - \text{default value}|$, as the target function. If the sum of errors is set too small, the training process may enter a local valley with poor results. Our network training experience indicates that when the sum of errors equals about 8, the mean relative error of the sample predictors can be kept at about 14 %, giving the best training result. In this case, we have 5 % probability to obtain predictions with the mean relative error of about 10 %. Figure 3 compares the averaged

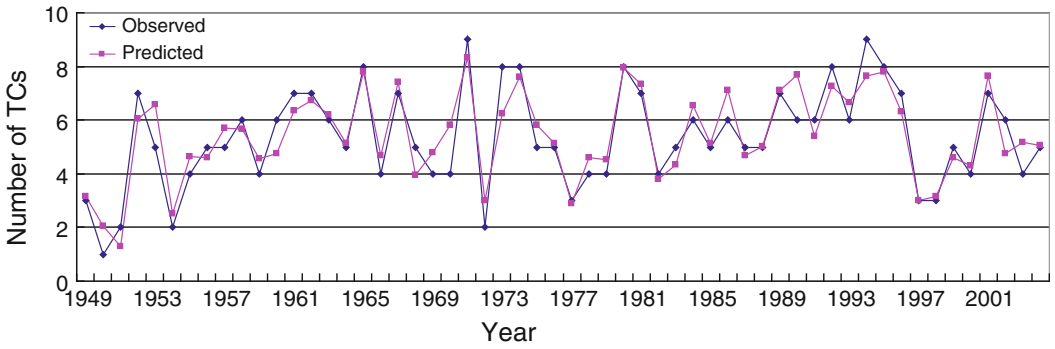


Fig. 3 Comparison of the time series of the trained annual TCF (*pink*) from the nonlinear WNN model to the observed one (*blue*) during the years of 1949–2004 (see text for more details)

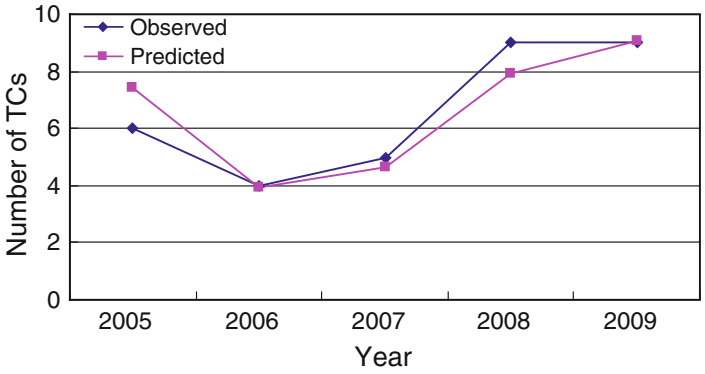


Fig. 4 Comparison of the WNN-predicted (*pink*) annual TCF after averaging 30 forecast tests to the observed one (*blue*) during the years of 2005–2009

annual TCF from the selected 30 trained TCFs with smaller relative errors after training the 56-year data sample for 600 times, at which the above-mentioned conditions are met. Obviously, the trained TCFs compare very favorably to the observed, revealing that the WNN nonlinear prediction model, based on the NOE and combined weights, does have a great data-fitting capability.

Now, we may use this trained WNN model to "predict" the subsequent 5-year TCFs, that is, 2005–2009, using the MH500 and MSST data from each month of a 12-month period prior to the June month of the TC season of concern (e.g., from June to December in 2004 and January to May in 2005 for the TCF prediction of 2005). Clearly, the TCFs and the MH500 and MSST data are independent from those used in the above training process. Figure 4 compares the predicted annual TCF, after averaging 30 forecast tests, to the observed one for each year from 2005 to 2009. The standard deviations of the predicted TCFs vary from 0.5 to 2.3. Table 2 shows that the mean absolute and relative errors are 0.6175 and 9.34 %, respectively. We have estimated the root-mean-squared error (RMSE) as 0.832. These results are indeed encouraging, indicating the effectiveness of the WNN-based model in predicting the annual TCFs for the coastal region of the Southern China.

4.3 Impact of combined weights, NOE and WNN

As mentioned in Sect. 4.1, the above encouraging results are obtained with four predictors, that is, the three principal component predictors plus one (combined weights) predictor, after performing NOE, combined weights and WNN. In this subsection, the importance of each of the three processes is examined using the results presented in Sect. 4.2 as a control model.

Specifically, in Test A, only the three principal component predictors are used as the input into the WNN model, but excluding the one predictor from combined weights. In Test B, the combined weight algorithm is applied multiple times, rather than going through NOE, until the 27 preliminary predictors is dimensionally reduced to one predictor, which is then used as the input into the WNN model. Table 3 shows that the mean absolute and relative errors for Tests A and B are 0.9823, 15.15 % and 0.7147, 11.64 %, respectively, with the RMSEs of 1.264 and 6.914. These errors are all obviously greater than those in the control prediction model, that is, 0.6175 and 9.34 % (see Table 2). Thus, the use of NOE and combined weights to extract the respective linear and nonlinear information of the

Table 3 As in Table 2 except for Test A (left: excluding the combined weights predictor) and Test B (right: without going through NOE)

Years	Observed	Test A			Test B		
		Predicted	Absolute error	Relative error (%)	Predicted	Absolute error	Relative error (%)
2005	6.0000	5.8731	−0.1269	2.12	7.1190	1.1190	18.65
2006	4.0000	4.3301	−0.6699	13.40	5.0352	0.0352	0.70
2007	5.0000	5.0448	1.0448	26.12	4.8627	0.8627	21.57
2008	9.0000	6.2100	−2.7900	31.00	7.6089	−1.3911	15.46
2009	9.0000	8.7202	−0.2798	3.11	8.8345	−0.1655	1.84
Average	6.6000	6.03564	0.9823	**15.15**	6.69206	0.7147	**11.64**
Standard deviation	2.3022	1.6670	1.3988	**13.13**	1.7106	0.9897	**9.72**

Table 4 As in Table 2 except for Test C (use of a multi-regression model)

Years	Observed	Predicted	Absolute error	Relative error (%)
2005	6.0000	7.3147	1.3147	21.91
2006	4.0000	4.4524	0.4524	11.31
2007	5.0000	4.2459	−0.7541	15.08
2008	9.0000	6.8053	−2.1947	24.39
2009	9.0000	9.9390	0.9390	10.43
Average	6.6000	6.55146	1.1309	**16.62**
Standard deviation	2.3022	2.3369	1.4308	**6.27**

predictors followed by the WNN process is effective in obtaining highly reliable prediction results.

In order to objectively evaluate the dimensional reduction of the predictors with multi-dimensional information and the TCF prediction performance of the WNN model, three tests are designed to see how well the traditional regression models would perform. Note that multi-regression models have been used at today's operational weather prediction centers for predicting the tracks and intensities of TCs (Aberson 1998; Kaplan and DeMaria 1999). Thus, in Test C, all the four predictors are used as the input into a linear multi-variable regression prediction model, instead of the WNN model. Table 4 shows that the mean absolute and relative errors for this prediction model are 1.1309 and 16.62 %, respectively, with an RMSE of 1.281. These errors are clearly greater than those given in Table 2.

In Tests D and E, the model input is the same as Tests A and B, respectively, except that a linear regression prediction model is used instead of the WNN model. Table 5 shows the associated errors of two tests. A comparison of Table 3 and Table 5 indicates that with the same predictors and the same independent testing sample, the WNN-based model produces higher prediction accuracies, suggesting further the effectiveness of the model in capturing the essence of the correlations between the TCFs and predictors.

Table 5 As in Table 3 except for Test D (left) and Test E (right), in which the WNN algorithm is replaced by a linear regression prediction model

Years	Observed	Predicted	Absolute error	Relative error (%)	Predicted	Absolute error	Relative error (%)
2005	6.0000	7.1385	1.1385	18.97	6.5523	0.5523	9.21
2006	4.0000	5.6145	1.6145	40.36	4.3999	0.3999	9.99
2007	5.0000	4.9184	−0.0816	1.63	4.8118	−0.1882	3.76
2008	9.0000	6.5883	−2.4117	26.8	5.1660	−3.8340	42.60
2009	9.0000	9.0324	0.0324	0.36	8.7270	−0.2730	3.03
Average	6.6000	6.6584	1.0557	**17.62**	5.9314	1.0494	**13.72**
Standard deviation	2.0322	1.5802	1.5580	**17.01**	1.7597	1.8055	**18.77**

5 Concluding remarks

In this work, we constructed a nonlinear WNN-based model by coupling NOE and combined weights in order to predict the annual TCF for the coastal regions of the Southern China. The global MH500 and MSST fields are used as the two major predictors. Because of the typical linear and nonlinear relationships between the TCFs and predictors, NOE is used to perform linear dimensional reduction in the predictors, while the combined weights method, which combines categorical weights, based on the particle pursuit, and ranking weights, based on the fuzzy mathematics, is used to conduct the nonlinear dimensional reduction. The projection of the multi-dimensional nonlinear data to a lower-dimensional space provides more useful linear and nonlinear information for the training of the WNN-based model. In constructing the WNN-based model, the S-type activation function in ANN is replaced by the Morlet mother wavelet basis function, and the global convergence error in WNN is set as the objective function.

The WNN-based model is tested with a 61-year dataset, using the first 56-year (i.e., 1949–2004) data as a training sample and the subsequent 5-year (i.e., 2005–2009) data to validate the prediction of TCFs. Results show that the mean absolute and relative errors are 0.6175 and 9.34 %, respectively. The impacts of using the combined weights, NOE and WNN as well as the traditional multi-regression approach for the TCF prediction are also examined. Results show superior performance of the WNN-based model in the TCF prediction.

Of course, the above results do not imply that the WNN model so constructed is perfect, especially in terms of physical processes involved in TCFs. In the future, it may be desirable to include vertical wind shear as one additional predictor to explore its impact on annual TCFs through the genesis of TCs (Kaplan and DeMaria 1999; Frank and Ritchie 2001; Zhu et al. 2003; Zhang and Kieu 2005). It will also be of interest to gain insight into why the annual TCFs over certain areas are highly correlated with MSST or MH500. Nevertheless, our results appear to have important implications to the construction of such a WNN-based prediction model for other fields (e.g., natural disasters, economics and finance).

Acknowledgments We would like to thank Prof. Zhaohua Wu of the Florida State University for his initial editorial assistance. This work was funded by China's NSFC Grants 41065002 and 11061008, the Key Scientific Research Grant 0993002-4 of Guangxi Province and the research Grant 200911MS151 of the

Department of Education of Guangxi Province. The second author was supported by the NSF Grant ATM0758609.

References

Aberson SD (1998) Five-day tropical cyclone track forecasts in the North Atlantic basin. Weather Forecast 13:1005–1015

Bingham HB, Zhang H (2007) On the accuracy of finite-difference solutions for nonlinear water waves. J Eng Math 58:211–228

Bonabeau E, Dorigo M, Theraulaz G (2000) Inspiration for optimization from social insect behavior. Nature 406(6):39–42

Chen L, Ding Y (1979) An introduction to Typhoons of West Pacific (in Chinese). China Science Press, Beijing

Chen G-Z, Wang J-Q, Xie X-M (2008) Application of particle swarm optimization for solving optimization problem of projection pursuit modeling (in Chinese). Comput Simul 25(8):159–165

Deng Z, Tu Q, Feng J, Xiong F (1999) The relation of the landfalling typhoon frequency to the Pacific SST field (in Chinese). Quart J Appl Meteor 10:54–60

Frank WM, Ritchie EA (2001) Effects of vertical wind shear on the intensity and structure of numerically simulated hurricanes. Mon Weather Rev 129:2249–2269

Iliadis LS, Spartalis SI (2005) Fundamental fuzzy relation concepts of a D.S.S. for the estimation of natural disasters' risk (the case of a trapezoidal membership function). Math Comput Model 42:747–758

Jin L (2004) Weather forecast of the neural network modeling of theoretical methods and applications. Meteorological Press, Beijing pp 46–47

Jin L, Ju W-M, Miao G-L (2000) Study on ANN-based multi-step prediction model of short-term climate variation. Acta Meteor Sinica 17:157–164

Jin JL, Wei YM, Ding J (2003) System evaluation model based on combined weights (in Chinese). Math Pract Theory 33:51–58

Jin L, Kuang X-Y, Huang H–H et al (2005) Study on the overfitting of the artificial neural network forecasting model. Acta Meteor Sinica 19:90–99

Kaplan J, DeMaria M (1999) An updated statistical hurricane intensity prediction scheme (SHIPS) for the Atlantic and eastern north Pacific basins. Weather Forecast 14:326–337

Krylov NV (2002) Introduction to the theory of random processes. Amer. Math. Soc, Providence

Kwon HJ, Lee WJ, Won SH et al (2007) Statistical ensemble prediction of the tropical cyclone activity over the Western North Pacific. Geophys Res Lett 34. doi:10.1029/2007GL032308

Lazarovitch N, Poulton M et al (2009) Water distribution under trickle irrigation predicted using artificial neural networks. J Eng Math 64:207–218

Liu H-X, Zhang D-L (2012) Analysis and prediction of hazard risks caused by tropical cyclones in southern China with fuzzy mathematical and grey models. Appl Math Model 36:626–637

Lu H, Jin L, Miao QL, Wang YH (2003) Prediction model of the annual frequency of tropical cyclones affecting Guangxi (in Chinese). J Nanjing Inst Meteor 26:56–62

Shi F, Wang XC, Yu L, Li Y (2010) MATLAB neural network 30 case analysis (in Chinese). Beijing Aerospace Industry University Press, Beijing, pp 210–212

Song ZY, Ling L, Nielsen P, Lockington D (2006) Quantification of tidal watertable overheight in a coastal unconfined aquifer. J Eng Math 56:437–444

Wang J, Bao Q (2009) Energy demand forecasting model in China based on wavelet-neural network (in Chinese). J Syst Sci Math Sci 29:1542–1551

Wu L, Tao L, Ding Q (2010) Influence of sea surface warming on environmental factors affecting long-term changes of Atlantic tropical cyclone formation. J Clim 23:5978–5989

Xiao F, Xiao Z (2010) Characteristics of tropical cyclones in China and their impacts analysis. Nat Hazards 54:827–837

Xie J-J, Liu C-P (2006) Fuzzy mathematics method and its applications (in Chinese), 3rd edn. Huazhong University of Science and Technology Press, Huazhong, pp 61–62

Yao C, Jin L, Zhao H-S (2009) Ensemble prediction of monsoon index with a genetic neural network model. Acta Meteor Sinica 23(6):701–712

Yin YZ, Luo Y, Marco G et al (2010) Forecast method study on yearly tropical cyclone frequency in Northwest Pacific based on the BP neural network technique (in Chinese). J Trop Meteor 26(5):614–619

Ying M, Wan RJ (2011) The annual frequency prediction of tropical cyclones affecting China (in Chinese). J Appl Meteor Sci 22(1):66–76

Yu P-S, Chen S-T, Yang T-C (2005) The potential of fuzzy multi-objective model for rainfall forecasting from typhoons. Nat Hazards 34:131–150

Zhang DL, Kieu CQ (2005) Shear-forced vertical circulations in tropical cyclones. Geophys Res Lett 32. doi:10.1029/2005GL023146

Zhu CW, Nakazawa T, Li JP (2003) Modulation of twin tropical cyclogenesis by the MJO westerly wind burst during the onset period of 1997/98 ENSO. Adv Atmos Sci 20(6):882–898