

ENSO-based tropical cyclone forecasting using CF-ANFIS

Trong Hai Duong¹ · Phi Hung Do² ·
Sy Dzung Nguyen³ · Minh Hien Hoang⁴

Received: 21 December 2015 / Accepted: 20 February 2016 / Published online: 29 March 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract According to literature, there are two aspects of a successful approach for seasonal forecasting of tropical cyclones, including factors relating to the formation and operation of the storms and regression methods. Dealing with the factors, El Niño–Southern Oscillation, and other global factors such as Quasi-Biennial Oscillation, Pacific Decadal Oscillation, etc. and local factors such as sea surface temperature, sea level pressures, etc. were examined for tropical cyclone forecasting. For regression, the most previous works used the linear regression-based model for seasonal tropical forecasting. However, the seasonal tropical forecasting requires high-dimensional data, so the forecasting ability using linear regression will have drawback. In this work, we analyse literatures of forecasting factors and regression methods for tropical cyclone forecasting. A CF-ANFIS algorithm integrating a conjunct space cluster and Cascade-forward neural network are proposed to forecast the number of tropical cyclone making landfall. This algorithm resolves the drawback by considering all forecast factors with high-

dimensional data. The experimental results indicated that the CF-ANFIS for seasonal forecast of tropical cyclones is a significantly effective approach with high accuracy in comparison with traditional ANFIS.

Keywords ANFIS · Machine learning · Forecasting · ENSO · El Niño · Tropical cyclone making landfall

1 Literature review

1.1 Relationship between ENSO and cyclone activities

Many researches have initially focused on studying the relationship between the El Niño phenomenon and occurrences of cyclones [2–4, 7, 8, 14, 15, 22, 26]. Several differences have been found regarding the occurring locations and directions of the storm in different places during El Niño episodes (warm phase of ENSO). Afterwards, there were studies on the impact of La Nina (cold phase of ENSO) and the overall impact of three ENSO phases (warm, cold and neutral phase) on cyclone activities and tropical depression. Since the late 80s, scientists have shown that it is possible to predict seasonal cyclone activities through indicators of ENSO phenomenon [22, 23]. ENSO has a considerable impact on cyclonic activities on a global scale at different levels and on different aspects in terms of the characteristics of the storm.¹ Any changes in ENSO such as climate changes will affect the frequency and orbital motions of tropical storms. Research shows that in El Niño conditions, for example, the frequency of cyclones often decreases on an annual basis in Atlantic, western Pacific and Australian waters, but increases in the Central and Eastern Pacific Ocean. Therefore, it can be noted

✉ Trong Hai Duong
haiduongtrong@gmail.com

Phi Hung Do
dophihung@gmail.com

Sy Dzung Nguyen
nsidung@yahoo.com

Minh Hien Hoang
hnm@netnam.vn

¹ International University, VNU-HCMC, Ho Chi Minh City, Vietnam

² Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam

³ Department of Mechanical Engineering, Ho Chi Minh University of Industry, HUI, Ho Chi Minh City, Vietnam

⁴ Disaster Management Centre, Ministry of Agriculture and Rural Development, Ho Chi Minh City, Vietnam

¹ <http://www.aoml.noaa.gov/hrd/tcfaq/G2.html>.

that the number of tropical storms in some areas reduced can be offset by the increase in cyclones elsewhere due to the global connections of atmospheric circulations in tropical regions.

It is known a tropical storm occurs only in areas where the sea surface temperature is below 27 °C. There exists a relationship between the storm's intensity and genesis location in the different phases of ENSO along with turbulent sea surface temperatures in the Northwest Pacific Ocean [2,3]. Emily A. Fogarty's research² shows a certain relationship between the storm's frequencies each year in China during different phases of ENSO. Chan's research showed a clear effect of the ENSO phenomenon on intense cyclone activities in the North West Pacific Ocean. Several numerical approaches have been employed to analyse the influence of ENSO phenomenon on cyclone activities, but the results have been limited. The statistical analysis method has been used quite extensively in analysing the relationship between ENSO indices and cyclone activities in different regions around the globe.

In the work [19], the authors have conducted seasonal forecasting for the number of storms landing the East coast using regression statistical method based only on factors that predict the ENSO indices (ESOI and NINO3.4). They considered the values of the sea surface temperature disturbance in the Nino3.4 region and the Southern Oscillation Index hemisphere equator (ESOI) of the previous 15 months until the forecast date to analyse and pick out 5 months with the highest correlation of each aforementioned factor for inclusion in the forecast equation. Here the author has paid attention to the forecast during El Nino or La Nina episodes while we also have to predict for the neutral phase; therefore, many studies on forecasting cyclones season have further analysed the relationships between climate factors and other circulations not related to ENSO. For example, one of the important factors in predicting seasonal cyclones in Northwest Pacific region is a factor regarding the Western North Pacific Subtropical High often reflected in some layers of the atmosphere on regional circulation in the region 15–25°N, 115–150°E. Cyclones cannot occur in the subtropical high-pressure area of Northwest Pacific. Subtropical high pressure in the Pacific northwest region, when intensifying and encroaching the west land, will block the cyclone from sweeping north and west, so in these cases, cyclones tend to land more inclined on the west coast.

The study [13] describes an improved statistical scheme for seasonal forecasting of tropical cyclones landing annually ashore along the East coast using data from 1965 to 2005. Based on the factors affecting the impact range of the cyclones in the South China Sea, the factors that affect the cyclones making landfall in the mainland have been identified. This equation is then developed using these factors'

empirical orthogonal functions to predict the number of hurricanes in the first half of the hurricane season (from May to August) in April and in the second half of the hurricane season (from November to December), and the number of storms during the major months of the hurricane season from July to December. This new scheme is approximately 11 % more improved in accuracy when predicting how many storms that make landfall during the entire hurricane season compared to previous studies. Analysis of circulation models show that the conditions within South China Sea appears to be the main factor affecting the number of hurricanes landing here. In the years that the figure is at an above-normal level, the conditions within the East Sea is favourable for the development of storms, and vice versa. The value 500 hPa in the intensity of the subtropical high-pressure level is also seen as a factor in determining whether the cyclone from the Pacific Northwest can approach the South China Sea and whether it can hit the mainland or not.

In the study [30] conducted by Haikun and his colleagues, a statistical model was built to simulate the formation of tropical cyclones and another trajectory model used to simulate the tropical cyclones' path, the effects of ENSO phenomenon on the trajectory of tropical cyclones during the major months of the hurricane season (from July to September) in the northwest Pacific region was assessed based on 14 El Niño and La Niña selected from the year 1950 to 2007. The analysis shows that during El Nino episodes, tropical cyclone activities grow significantly stronger in the south latitude 20°N, particularly in the east longitude 130°E. Tropical cyclones with prevailing northwest trajectory and impacting East Asia, including the island of Taiwan, mainland China, the Korean Peninsula and Japan, tend to move more to the west during El Nino years and tend to move more to the north during La Nina years. The numerical simulations also determined that changes related to ENSO in large-scale currents and cyclone forming positions can have a significant impact on the prevailing tropical cyclone trajectory.

In [18], Landsea indicates that one of the biggest impacts of ENSO phenomenon on the Earth's climate system is the alteration of tropical cyclones' characteristics on a global scale. This article by Landsea gives an overview of frequency, intensity and how genesis regions of tropical cyclones have alternated between one another in all storm occurring zones globally following the development phases of ENSO. Together with the impact of the ENSO phenomenon, global factors (such as Quasi-Biennial Oscillation, QBO) and the local factors (such as sea surface temperature, intensity and monsoon rain, sea-level field of pressure and vertical wind shear) can regulate the change of tropical cyclones. The relationship of these factors with the first tropical cyclones, especially a high correlation with the ENSO phenomenon, can be exploited to make seasonal forecasts of the tropical cyclone activities. The author has presented details of

² <http://myweb.fsu.edu/jelsner/Theses/EmilyFogartyMS.pdf>.

the methods developed for specific genesis regions of tropical cyclone activities in the North Atlantic, western North Pacific, South Pacific and Australia.

In the study [28], Wang and colleagues analysed a series of 35-year-long data (1965–1999) that shows the strong impact of the El Niño and La Niña phenomenon with high intensity on tropical cyclone activities in the northwest Pacific, although the total number of tropical cyclones formed over the entire North–West Pacific region did not change significantly from year to year. In the summer and autumn of the El Niño years the frequency of tropical cyclone formation increased considerably in the Southeast corner (0° – 17° N, 140° – 180° E) and decreased in the northwest corner (17° – 30° N, 120° – 140° E). The genesis location of cyclones in the period from July to September moves 6° latitude lower from the average one in many years, while in the period from October to December this moves 18° longitude eastward in dry years compared to the average locations in cold years. After El Niño (La Niña), tropical cyclone formation in the first half of the hurricane season (from January to July) decreases (increases). In dry (cold) years the tropical cyclone average lifespan is about 7 (4) days, and the average total number of tropical cyclone days is 159 (84) days. In autumn during the hot, strong El Niño years, the total number of tropical cyclones hovering north passing the latitude 35° N is 2.5 times more than in the cold, strong La Niña years. This implies that El Niño substantially increases the movement toward the extremes of heat-moisture energy and impacts high latitudes via changing the formation of tropical cyclones and their trajectories. The emergence of tropical cyclones which were increased in the southeast corner of the Northwest Pacific region, is increasing the spin degree in lower levels generated by El Niño, inducing equatorial westerly wind. The reduction in the occurrence of tropical cyclones in the northwest corner is attributed to convergence on the top level caused by the deepening of the east Asia trench and the strengthening of the north-western Pacific subtropical high pressure. The both of these phenomena were caused by the impact of El Niño. Tropical cyclone activities in the month of the hurricane season from July to December are very likely to be forecast based on disturbance of the sea surface temperatures that occurred in the NINO3.4 region earlier from winter to spring, while we can predict tropical cyclones generated in the period from March to July through disturbances of the sea surface temperatures that occurred earlier in the NINO3.4 region from October to December.

In [29], Chang and colleagues pointed out that in September, October, and November, in the neutral year or in the second half of the hurricane season during the El Niño years the number of tropical cyclones making landfall in the Northwest Pacific coastal region, except for Japan and the Korean peninsula, decreased markedly. On the other hand, in the second half of the hurricane season in La Niña years,

the number of tropical cyclones landing in China coast rose significantly in the La Niña years. The reduced number of tropical cyclones making landfall in the second half of the El Niño hurricane season seems related to the shift to the east of the average genesis location where cyclones occur and the disruption of the ridge on 500 mb level near the 130° E longitude. In contrast, the number of tropical cyclones making landfall in the later half of the hurricane season in La Niña years seems to be related to the shift to the west of the average cyclone genesis location and the maintenance of the ridge on the 500-mb level.

In [17], Kim and colleagues suggest that the warming of the Pacific sea is divided into two distinct modes according to the spatial distribution of the sea surface temperature turbulence: warming in the Eastern Pacific (EPW) and warming in the central Pacific (CPW). The 3rd mode is the cooling of the sea surface temperature in the Eastern Pacific region (EPC). These three modes cause different impacts on the operating mode of tropical cyclones over the North Pacific waters regulated by both the various local thermodynamic factors and large-scale circulations models. In years EPW (sea surface warming in the Eastern Pacific) the density and trajectory of tropical cyclone tend to increase in the southeast and decrease in the western Pacific with strong west wind shear. The expansion of the monsoon trough and the weak wind shear on the central Pacific region increased conditions for tropical cyclones in the east compared with the average position of tropical cyclones in many years. In the CPW years (sea water warming in the central Pacific), tropical cyclone activities often shift westward and expand in the area of the Western Pacific Northwest. The shift to the west of CPW entails turbulent heat movement of west wind heat and monsoon trough through the north-western region of the Western Pacific and creates more favourable conditions for the arrival of tropical cyclones. By contrast with the CPW years, there is a decline of tropical cyclone activities in the waters of the East Pacific. Over the EPC years (cooling of sea surface temperature in Eastern Pacific), all investigations on the parameters mostly draw a contrast reflection of the EPW years.

According to literature, forecasting methods can be categorized into qualitative and quantitative approaches. The former usually based on the opinions of people, which refers to a long or medium forecast by asking a group of knowledgeable experts for their opinions with regard to future values of the things being forecasted. The well-known method, called Delphi, involves a group of experts who eventually reach a consensus of a forecast. The later refers to quantitative, mathematical formulations or statistical forecasting, which includes time series models and casual models. The regression, a causal forecasting model that fits curves to the entire data set to minimize the forecasting errors, is often applied for the seasonal forecasting of tropical cyclones (TCs).

William Gray and his team pioneered the seasonal hurricane prediction enterprise using regression-based linear statistical models [15]. In a study, Chu [7], Fan and Wang [10] presented a multivariate linear regression model applied to predict the seasonal tropical cyclone count in the vicinity of Taiwan. The model is based on the least absolute deviation so that regression estimates are more resistant than those derived from the ordinary least square method. Kim et al. [16] used least absolute deviation (LAD) regression and the Poisson regression method. Poisson model is being slightly more skilful than the LAD model. Goh and Chan [11–13] presented an improved prediction scheme for the number of TCs making landfall on the coast of south China. The schemes for the early, late, and JD seasons all provide reasonable results. Chu and Zhao [5,6] applied a hierarchical Bayesian change point analysis to detect abrupt shifts in the TC time series over the central North Pacific (CNP). Chu [4] extended the probabilistic Bayesian framework suggested in the prior works from the CNP [6], with a particular focus toward the vicinity of the Taiwan area. Different from prior studies, he adopts a feature classification approach based on the fuzzy clustering analysis of TC tracks.

2 Forecasting methods

According to the literatures review, the most previous works applied the linear regression-based models for seasonal tropical forecasting. However, the higher order polynomial models are used, the overall degree of error will be reduced. The seasonal tropical forecasting requires high-dimensional data, so the forecasting ability also is reduced for higher order polynomial models. According to Chu [4–7], Bayesian-based model for seasonal typhoon activity forecasting is more effective than linear regression-based models. In another aspect, Azizi [1] shows that ANFIS model provides better forecasting accuracy in comparison with Bayesian model. Hence, ANFIS model is recommended to be used for production estimation under random uncertainties. According our study, the ANFIS has been not yet applied for seasonal tropical forecasting up to our previous research [9]. Another reason to choose ANFIS model is that it can be used to combine all predictor factors for forecasting, while other approaches only use several factors by transforming high-dimensional data to low-dimensional data.

This work is an improvement of our previous research [9]. The aim of the previous work [9] was to offer usefully realistic supports for seasonal forecast of tropical cyclone activities along the Vietnam coast. The forecasting factors include ENSO, atmospheric and oceanographic data related to formation conditions and tropical cyclone activity in the study area in the 62-year period (1951–2012). The conjunct space cluster-based adaptive neuro-fuzzy inference system

was applied for seasonal forecasting of tropical cyclones. This model integrated a conjunct space-based cluster method and a perceptron (called P-ANFIS). The perceptron is a linear regression model, so P-ANFIS still gets drawback with high-dimensional data. Here, an improved version of P-ANFIS is proposed by using cascade-forward neural network [1] instead of a perceptron, which is called CF-ANFIS.³ The experimental results indicated that the CF-ANFIS for seasonal forecasting of tropical cyclones using ENSO is a significant effective approach with high accuracy in comparison with P-ANFIS.

3 Tropical cyclone forecasting using CF-ANFIS

An adaptive neuro-fuzzy inference system (ANFIS) is a machine learning technique capturing the advantage of both neural networks and fuzzy logic principles in a single framework. ANFIS has a learning capability to generate a set of fuzzy IF-THEN rules to be approximate nonlinear functions, which is used to make inference. Given data set T_{Σ} as follows

$$(\bar{x}_i, y_i), \quad \bar{x}_i = [x_{i1} \ x_{i2} \ \dots \ x_{in}], \quad i = 1 \dots P, \quad (1)$$

where \bar{x}_i is the i th input vector of the data set and y_i is the output; P is number of samples. A black-box type model expressing a mathematical relationship between input and output spaces of a system based on this given data set can be expressed by mapping as follows:

$$f: \mathfrak{R}^n \rightarrow \mathfrak{R}^1 \\ \bar{x}_i \mapsto y_i | y_i = f(\bar{x}_i).$$

System modelling is defined by extrapolating the function f based on these known data. It is well known that there are some existing algorithms to solve this problem based on the fuzzy inference system, called T-S model, which was proposed by Takagi and Sugeno [27]. The k th. rule of this fuzzy model is as following:

$$R^{(k)}: \text{If } x_{i1} \text{ is } B_1^k \text{ and } \dots \text{ and } x_{in} \text{ is } B_n^k;$$

then

$$y_{ki} = \sum_{j=1}^n a_j^{(k)} x_{ij} + a_0^{(k)},$$

where $\bar{x} = [x_{i1}, x_{i2}, \dots, x_{in}]^T$ is the i th input data vector of T_{Σ} ; $B^{(k)}$ is input fuzzy set; $[a_0^{(k)}, a_1^{(k)}, \dots, a_n^{(k)}]^T$ is the weight vector; y_{ki} is output data; $j = 1 \dots n$, n is dimension

³ <http://www.mathworks.com/help/nnet/ref/cascadeforwardnet.html>.

of input data set; $k = 1 \dots m$, m is the number of IF-THEN rules.

In this paper, the above mathematical model is expressed by the ANFIS. ANFIS is one of the most popular types of fuzzy neural network [20, 25–27]. The clustering techniques are commonly used to create fuzzy rules of ANFIS.

According to [27], each data cluster can be considered as a crisp frame on which different types of MFs (and firing strengths) can be adapted. Some serious drawbacks often affect the clustering algorithms adopted in this context, according to the particular data spaces where they are applied. To overcome such problems, Panella et al. [24] analysed various clustering methods adopting for ANFIS, including clustering in input space, clustering in output space and clustering in input–output space (conjunct space) [26]. The clustering based on the data set only in the input space, which assumed that points potentially belonging to the same cluster in the input space are mapped into points potentially belonging to the same cluster in the output space. Its disadvantage is that the output clusters could not reflect the real structure of the mapping in the output space. On the other the hand, the clustering method considers only output space, which can be ensure that the possibility to discover the real structure of the mapping in the output space. Unfortunately, there can be contradictory rules having similar input MFs but different output coefficients, which is unacceptable in ANFIS networks. To overcome these problems, M. Panella [24–26] and Dzung [20] presented a clustering method in conjunct space for ANFIS network construction. The clustering in input–output space mentioned in the previous works combines a linear cluster (e.g. hyperplane cluster) and Simpson’s min–max models for classification (min–max classification).

Relating to building the clustering data space, the algorithm for parting data system, named PDS [1], is used to build pure hyper-boxes in the input data space and hyperplanes in the output data space. The value corresponding to the k th data hyperplane is calculated as follows:

$$y_k^{(i)} = \sum_{j=1}^n a_j^{(k)} x_{ij} + a_0^{(k)}, \quad i = 1 \dots P, \tag{2}$$

where $[a_0^{(q)}, \dots, a_n^{(q)}]^T$, is a vector of parameters. Relating to the NN, it can be sketched out as follows. It is a Cascade-forward neural network consisting of one input layer, one hidden layer and one output layer. In the input layer, number of input signals depends on optimal structure of the clustering data space built by the data separating process to establish data clusters. Number of neurons at the hidden layer is adaptively established in training process of the NN.

The hyperplane clustering algorithm is expressed as follows:

Initialization. The C -means algorithm is used to initialize hyperplanes by clustering the input space into M clusters $\Gamma^{(k)}$, $k = 1 \dots M$. The correspondence between such clusters and initialized hyperplanes is based on following criterion: If an input pattern \bar{x}_i , $i = 1..P$ belongs to the cluster $\Gamma^{(q)}$, $1 \leq q \leq M$, then the corresponding input–output pair (\bar{x}_i, y_i) is assigned to the hyperplane A_q .

Step 1. The coefficients of each k th hyperplane, $y_t = \sum_{j=1}^n a_j^{(k)} x_{tj} + a_0^{(k)}$, $k = 1 \dots M$, is updated using the pairs assigned to either in the initialization or in the successive step 2, where index t spans all the pairs assigned to the k th hyperplane using suited least-squares techniques.

Step 2. Each pair (\bar{x}_i, y_i) is assigned to a hyperplane A_q , which has the minimum orthogonal distance d_i from it. The stop condition is determined using a convergence quantity $\sigma = (|D - D^{(old)}|/D^{(old)})$ where D is current approximation error defined by $D = \frac{1}{P} \sum_{i=1}^P d_i$ and $D^{(old)}$ is the previous approximation error. The algorithm will be stopped if it satisfies $\sigma \leq$ a predefined threshold ε ; otherwise, it goes back to step 1.

The previous algorithm is a linear clustering that only yields the linear consequent of Sugeno rules. According to [25], several clusters of the input space could be associated with the same hyperplane. To solve this problem, the well-known Simpson’s min–max models for classification (min–max classification) were applied by Panella [24, 25] and Dung [20]. The combination of the hyperplane clustering and the max-min classification on the input space supports to effectively determine the ANFIS network for a given number of rules. The min–max classification technique uses hyperboxes (HBs) which have boundary hyperplanes parallel to the coordinate axes of the patterns of the training set.

We consider a set of the patterns T_t covered by the t th min–max hyperbox HB_t . The HB_t is determined using two vertexes, the max vertex $\bar{\omega}_t = [\omega_{t1} \omega_{t2} \dots \omega_{tn}]$ and the min vertex $\bar{v}_t = [v_{t1} v_{t2} \dots v_{tn}]$, where $\omega_{tj} = \max(x_{ij} | \bar{x}_i \in T_t)$ and $v_{tj} = \min(x_{ij} | \bar{x}_i \in T_t)$. If T_t consists of the patterns associated with the cluster labelling m only, then the HB_t will be considered as a pure hyperbox labelling m and denoted $pHB_t^{(m)}$. An HB can be considered as a crisp frame on which different types of membership functions (MFs) can be adapted. Here, the original Simpson’s MF is adopted, in which the slope outside the HB is established by the value of the fuzziness parameter γ ,

$$\mu_{pHB_t^{(m)}}(\bar{x}_i) = \frac{1}{n} \sum_{j=1}^n [1 - f(x_{ij} - \omega_{tj}, \gamma) - f(v_{tj} - x_{ij}, \gamma)]$$

$$f(x, y) = \begin{cases} 1 & \text{if } xy > 1 \\ xy & \text{if } 0 \leq xy \leq 1 \\ 0 & \text{if } xy < 0, \end{cases} \tag{3}$$

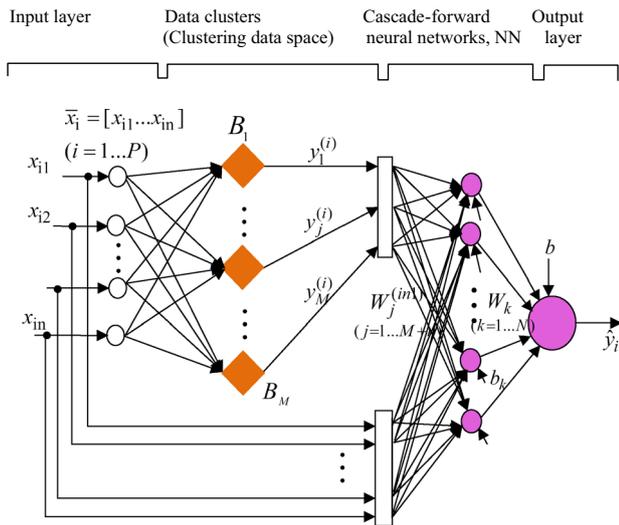


Fig. 1 Structure of FNS

where $t = R_m$, and R_m is the number of pure hyperboxes labelling m . Several pHB can be associated with the same cluster labelling m ; thus the overall input MF, $\mu_{B_i^{(m)}}(x_i)$, is calculated as follows:

$$\mu_{B_i^{(m)}}(\bar{x}_i) = \max\{\mu_{\text{pHB}_1^{(m)}}(\bar{x}_i), \dots, \mu_{\text{pHB}_{R_m}^{(m)}}(\bar{x}_i)\}. \tag{4}$$

There are several approaches for min–max classification such as Simpson’s min–max models, ARC in [25] and CSHL in [20]. Here, the CSHL algorithm is applied for this work.

The combination of the hyperplane clustering followed by the min–max classification on the input space can be determined of the ANFIS network for a given number of rules. In this paper, the above mathematical model is expressed by a fuzzy-neuron structure (FNS). The proposed structure of the FNS depicted in Fig. 1 is a combination of a clustering data space and a Cascade-forward neural network, NN.

Structure of the NN is created as follows:

Input vector of the NN is $M + n$ dimensions. The number of neuron N in the hidden layer is $N = N_0$ at the beginning time of the training process and adaptively adjusted in the net training process. The number of neurons in the output layer is 1. The ‘sum’ function is used for input of all of neurons. The ‘purelin’ function, $f(s) = s$, is used for output of the neuron at output layer. Transfer function is used for all of neurons at the hidden layer as follows:

$$f(s) = \frac{2}{1 + \exp(-2s)} - 1.$$

The i th input–output sample of the data set used for training the NN is established as follows:

$$[\mathfrak{N}_i \ \wp_i] = [(\bar{x}_i \ y^{(i)}), y_i]. \tag{5}$$

This is combined by the i th input–output sample of the data set (1) and values of corresponding hyperplanes $y_k^{(i)}$ calculated by (2), $y^{(i)} = [y_1^{(i)} \ y_2^{(i)} \ \dots \ y_M^{(i)}]$.

Training process the NN to adjust net parameters is performed as follows:

Let W be the weight vector of the NN, $W = [w_1 \ w_2 \ \dots \ w_H]^T$. The error equation of the NN can be written as follows:

$$\begin{aligned} E_r(W) &= \frac{1}{P} \sum_{i=1}^P (\wp_i - \hat{\wp}_i(W))^2 \\ &= \sum_{i=1}^P e_i^2(W) \\ &= V^T(W)V(W), \end{aligned}$$

where

$$\begin{aligned} V(W) &= [v_1(W) \ v_2(W) \ \dots \ v_P(W)]^T \\ &= [e_1(W) \ e_2(W) \ \dots \ e_P(W)]^T. \end{aligned}$$

Using the algorithm Levenberg–Marquardt, the weight vector W of the NN at the $(k + 1)$ th loop, signed W_{k+1} , is calculated as follows:

$$W_{k+1} = W_k - [J^T(W_k)J(W_k) + \mu I]^{-1} J^T(W_k)V(W_k),$$

where I is unit-square matrix, size H ; μ is an adaptive index; and J is the matrix Jacobian as follows:

$$J(W_k) = \begin{bmatrix} \frac{\partial v_1}{\partial w_1} & \frac{\partial v_1}{\partial w_2} & \dots & \frac{\partial v_1}{\partial w_H} \\ \frac{\partial v_2}{\partial w_1} & \frac{\partial v_2}{\partial w_2} & \dots & \frac{\partial v_2}{\partial w_H} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial v_P}{\partial w_1} & \frac{\partial v_P}{\partial w_2} & \dots & \frac{\partial v_P}{\partial w_H} \end{bmatrix}_{(W_k)}.$$

Via the aspects aforementioned, the process of building the FNS can be briefly presented as follows.

Initializing:

- The initial number of data clusters: $M = M_0$.
- The number of neurons in the hidden layer: N_0 .

Step 1. Cluster the data space: call the algorithm PDS of [20,21].

The obtained result is a cluster data space having M_{op} pure data clusters in which M_{op} is the optimal number of data clusters.

Step 2. Building structure of the FNN as follows: input: (x_i, y_j^i) , output: $y_j^i, i = 1 \dots P; j = 1 \dots M$, with N neurons in the hidden to be N .

Step 3. Training the FNS.

During the training process, the FNS is evaluated and updated based on E_N as follows:

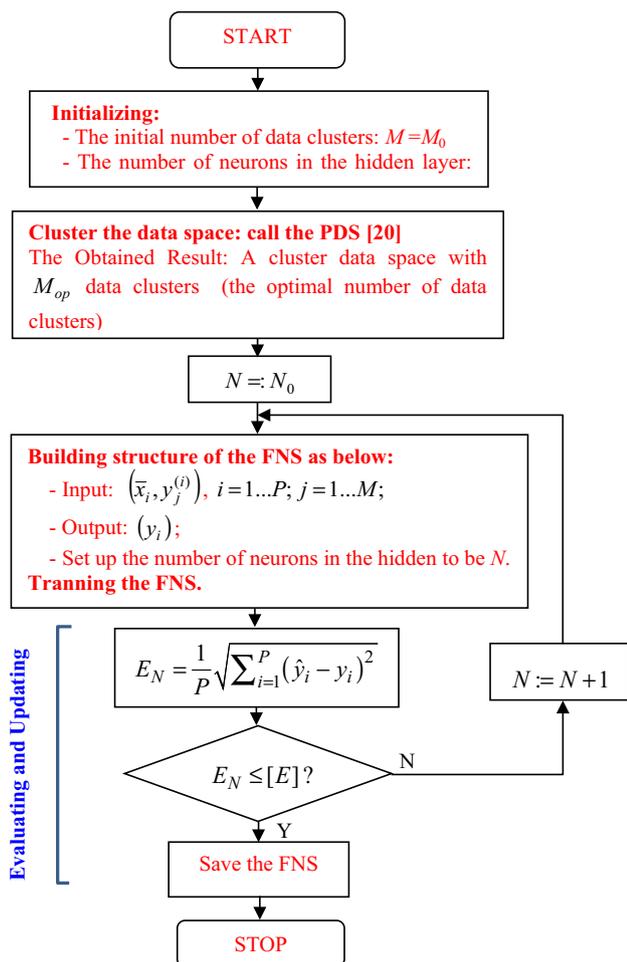


Fig. 2 Framework of the algorithm for building FNS

$$E_N = \frac{1}{P} \sqrt{\sum_{i=1}^P (\hat{y}_i - y_i)^2} \tag{6}$$

If $E_N \leq [E]$, the current structure of the FNS is the required one. The training process is stopped. Conversely, increasing the number of neurons in the hidden layer is performed and the training is resumed:

$$N =: N + 1. \tag{7}$$

The flowchart of this work is depicted in Fig. 2.

4 Experimental results

4.1 Dataset

According to literature, the ENSO events affect the different characteristics in typhoon’s activity in Western North Pacific, South China Sea and in Vietnam. It causes the changes in the

Table 1 The number of tropical depressions in the Vietnamese coast (1951–2011)

Year	Number of landfall TC	Year	Number of landfall TC	Year	Number of landfall TC
1951	2	1972	6	1993	5
1952	8	1973	10	1994	6
1953	1	1974	7	1995	7
1954	3	1975	2	1996	8
1955	2	1976	2	1997	4
1956	3	1977	3	1998	6
1957	1	1978	6	1999	2
1958	0	1979	3	2000	2
1959	1	1980	4	2001	2
1960	3	1981	3	2002	0
1961	2	1982	4	2003	2
1962	4	1983	6	2004	2
1963	2	1984	4	2005	5
1964	10	1985	4	2006	4
1965	3	1986	4	2007	5
1966	2	1987	4	2008	5
1967	1	1988	1	2009	5
1968	5	1989	8	2010	4
1969	2	1990	8	2011	4
1970	3	1991	3		
1971	7	1992	5		

origin of typhoon formation, frequency, intensity, track and in other characteristics of acted typhoons in these regions. We collect factors relating to the formation and activity of the storms in the study area. In particular, the indices of El Niño–Southern Oscillation (ENSO) including warming phase (El Niño), cooling phase (La Niña) and neutral phase relate to the activity of the tropical storms and tropical depressions in the Vietnamese coast. In addition to ENSO, other global climate factors (such as the stratospheric Quasi-Biennial Oscillation, Pacific Decadal Oscillation (PDO), North Atlantic Oscillation, Arctic Oscillation, Antarctic Oscillation, the northern hemisphere oscillation, long-wave radiation equatorial Pacific, etc.) are selected as forecasting factorsaê. The local factors (such as sea surface temperature, monsoon intensity and rainfall, sea level pressures, tropospheric vertical shear, perceptible water, low-level relative vortices, and vertical wind shear) can also help modulate tropical cyclone variability.

The annual number of tropical depressions in the Vietnamese coast from 1952 to 2011 is shown as Table 1.

The data set with 30 of the original principal (initial factor) affects the number of storms in years (Table 2) are used to establish training set.

Table 2 30 factors for TC forecasting

No.	Factors
1	MEI
2	SOI: Anomaly Data
3	SOI: Standardized Data
4	Reqsoi—Equatorial SOI data (1951–2012)
5	SOIx—Extratropical Southern Oscillation Index: SOI*
6	Northern Oscillation Index: NOI
7	NINO 1 + 2
8	NINO 3
9	NINO 4
10	NINO 3.4
11	Northern Atlantic Oscillation (NAO)
12	Tropical/ Northern Hemisphere (TNH) Pattern
13	Arctic Oscillation index since January 1950
14	PDO index 1900–2012
15	Pacific Transition Pattern (PTI)
16	QBI data 1948–2012
17	Repac—Equatorial Eastern Pacific SLP (Standardized Anomalies) 1949–2012
18	Rindo—Indonesia SLP (Standardized Anomalies)
19	WP 1948–2012
20	Standardized Northern Hemisphere Teleconnection Indices (1981–2010 Clim) (NAO)
21	Standardized Northern Hemisphere Teleconnection Indices (1981–2010 Clim) (EA)
22	Standardized Northern Hemisphere Teleconnection Indices (1981–2010 Clim) (WP)
23	Standardized Northern Hemisphere Teleconnection Indices (1981–2010 Clim) (EP/NP)
24	Standardized Northern Hemisphere Teleconnection Indices (1981–2010 Clim) (PNA)
25	Standardized Northern Hemisphere Teleconnection Indices (1981–2010 Clim) (EA/WR)
26	Standardized Northern Hemisphere Teleconnection Indices (1981–2010 Clim) (SCA)
27	Standardized Northern Hemisphere Teleconnection Indices (1981–2010 Clim) (TNH)
28	Standardized Northern Hemisphere Teleconnection Indices (1981–2010 Clim) (POL)
29	Standardized Northern Hemisphere Teleconnection Indices (1981–2010 Clim) (PT)
30	Standardized Northern Hemisphere Teleconnection Indices (1981–2010 Clim) (Expl. Var.)

4.2 Result

We used the samples from 1951 to 2000 to train the CF-ANFIS network. The test samples are from 2001 to 2011.

The prediction error for the i th factor is calculated as follows (Fig. 3):

$$\text{error}_i = \frac{\text{factor}_i - \text{factor}_i^{\text{NF}}}{\text{factor}_i} * 100(\%)$$

The dataset with 30 of the factors affecting the number of storms in years (see Table 2) are used to establish training set in two cases. Case 1: Survey 2 stages in each year, first 6 months and last 6 months. Case 2: each factor is used to determine initial 4-factor up to the value of the average quarterly (3 months in a row, 1–3, 4–6, 7–9, 9–12. Accordingly, set the training set with case 1

$P = 61$ input–output data samples corresponding to the 61-year survey period, from 1951 to 2011; each sample $n = 60$ data input (the factor) and 1 output (the number of storms in the corresponding year.) for case 2, we have $P = 61$ but each sample $n = 120$ data input, the factor number, 1 output is the number of storms in the respective years. In both of cases, the sample data input–output i is denoted as $\vec{X}_i = [X_{i1} \ X_{i2} \ \dots \ X_{in}]$, and $y_i, i = 1 \dots P$ (see Fig. 3).

The mean square error of case 1 is $\text{LMS} = 8.89 \times 10^{-4}$; case 2's $\text{LMS} = 2.9366 \times 10^{-6}$. This result shows that the factors is divided by quarter help the increasing number of data dimensions from 60 to 120, this leads to increase the accuracy of ANFIS in finding out the relationship between the elements affecting climate and storms occurring during the year. Qualitatively, the reason for the increased accuracy, because of the tropical climate, four sea-

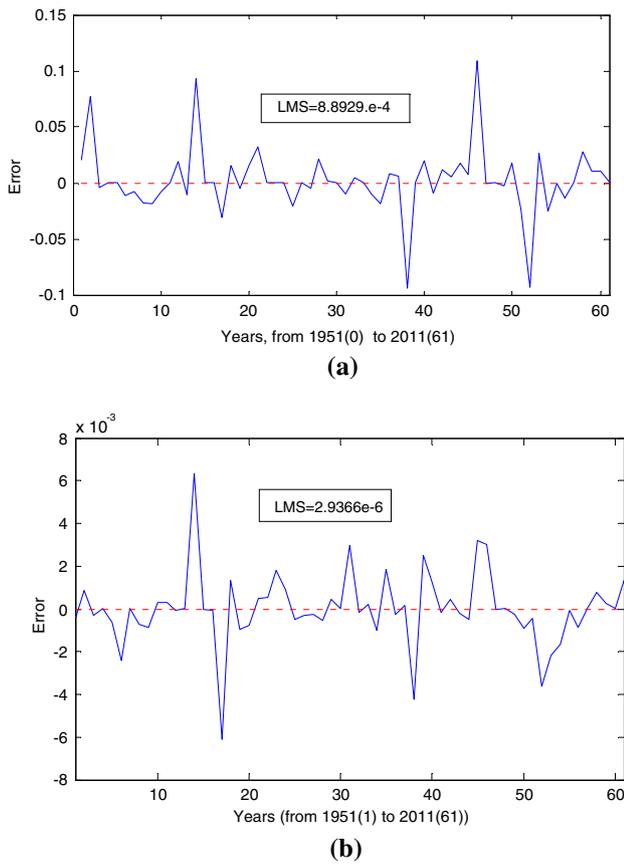


Fig. 3 Difference between the output of training data in two cases: case 1 (a) and case 2 (b)

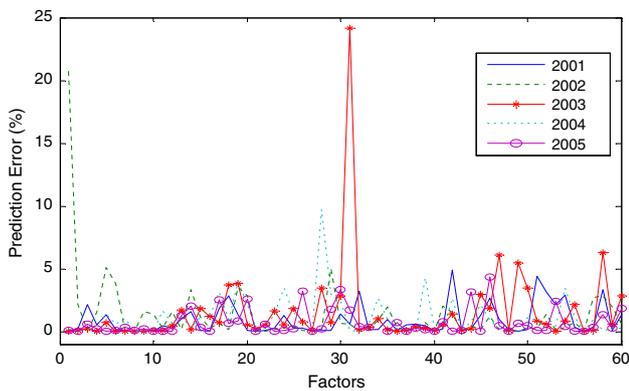


Fig. 4 Prediction error of 60 factors affecting to TC in the years, from 2001 to 2005

sons a year, quite clear, and therefore, the factor divided by quarter reflects better characteristics than divided by two seasons.

Figures 4 and 5 present the prediction error of 60 factors affecting to TC in the years, from 2001 to 2010.

Figure 6 shows the predicted number of TC versus reality in the years, from 2001 to 2011. The discrepancy between

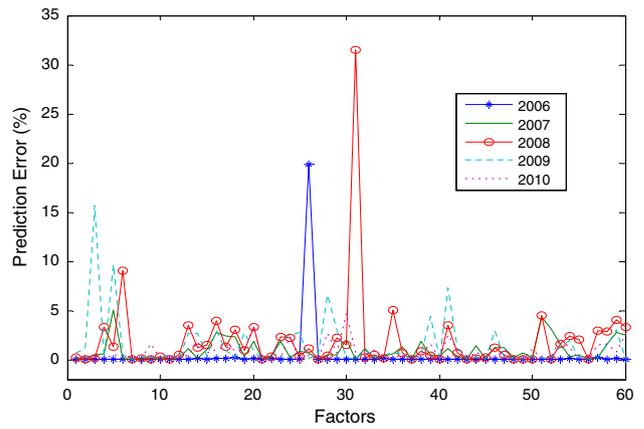


Fig. 5 Prediction error of 60 factors affecting to TC in the years, from 2006 to 2010

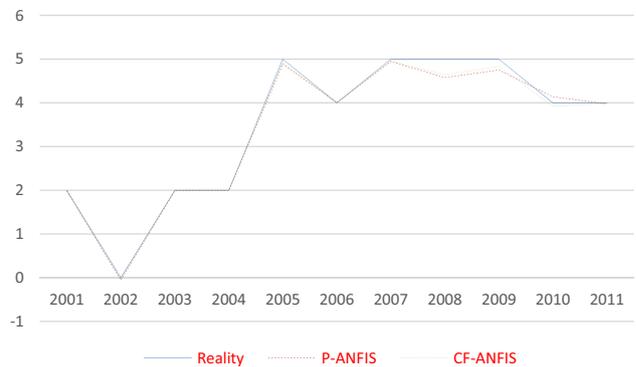


Fig. 6 The number of predicted TC and reality in the years from 2001 to 2011

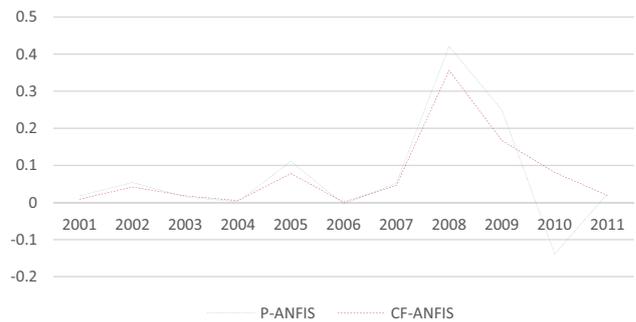


Fig. 7 Discrepancy between the number of predicted TC and reality in the years, from 2001 to 2011

the number of predicted TC and reality in the years, from 2001 to 2011 is presented as Table 2 and Fig. 7.

The difference between the actual number of TCs and the number of TCs that was predicted in a given period (year) is measured as follows:

Table 3 Discrepancy between the number of predicted TC and reality in the years, from 2001 to 2011

Year	Number of TC in reality, y_i^{data}	Error prediction for period i , $\text{err}_i = y_i^{\text{data}} - y_i^{\text{NF}}$ (P-ANFIS)	Error prediction for period i , $\text{err}_i = y_i^{\text{data}} - y_i^{\text{NF}}$ (CF-ANFIS)
2001	2	0.018	0.009
2002	0	0.054	0.042
2003	2	0.015	0.018
2004	2	0.002	0.005
2005	5	0.111	0.078
2006	4	-0.004	0.001
2007	5	0.052	0.046
2008	5	0.421	0.356
2009	5	0.249	0.167
2010	4	-0.139	0.081
2011	4	0.024	0.019

$$\text{err}_i = y_i^{\text{data}} - y_i^{\text{NF}},$$

where, y_i^{data} and y_i^{NF} are the actual and predicted number of TCs in year i , respectively.

The mean absolute deviation: 0.075

$$\text{err}_{\text{mean}} = \frac{\sum_{i=2001}^{2011} |\text{err}_i|}{11(\text{year})} = 0.099.$$

The discrepancy between the actual number and the predicted number of TCs indicated that the forecast (in what?) is slightly lower than reality (see Fig. 7; Table 3). The mean error in the set of forecasts is only 0.099 and 0.075 for P-ANFIS and CF-ANFIS, respectively. From Fig. 6, we can say that the accuracy of prediction is significantly high for both P-ANFIS and CF-ANFIS.

5 Conclusions

This paper has presented a conjunct space cluster-based ANFIS for the seasonal forecasting of tropical cyclones making landfall along the Vietnam coast. The experimental result has indicated that the conjunct space clustering-based ANFIS is an effective approach with high accuracy for the seasonal forecasting of tropical cyclones. However, the perceptron-based ANFIS still has drawbacks in comparison with Cascade forward neural network-based ANFIS.

Acknowledgments This research is funded by International University, VNU-HCM under Grant No. T2015-05-IT/HĐ-ĐHQ-TQLKH.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Azizi, A., Ali, Y.A., Ping, L.W.: Model development and comparative study of Bayesian and ANFIS inferences for uncertain variables of production line in tile industry. *WSEAS Trans. Syst.* **11**(1), 22–37 (2012)
2. Chan, J.C.L.: Tropical cyclone activity in the Northwest Pacific in relation to the El Niño/Southern Oscillation phenomenon. *Mon. Weather Rev.* **113**, 599–606 (1985)
3. Chan, J.C.L.: Interannual variability of tropical cyclones making landfall over China. In: ESCAP/WMO Typhoon Committee Annual Review Rev., pp. 85–94 (1993)
4. Chu, P.-S.: Bayesian forecasting of seasonal typhoon activity: a track-pattern-oriented categorization approach. *J. Clim.* **23**, 6654–6668 (2010)
5. Chu, P.-S., Zhao, X.: Bayesian change-point analysis of tropical cyclone activity: the central North Pacific case. *J. Clim.* **17**, 4893–4901 (2004)
6. Chu, P.-S., Zhao, X.: A Bayesian regression approach for predicting seasonal tropical cyclone activity over the central North Pacific. *J. Clim.* **15**, 4002–4013 (2007)
7. Chu, P.S., et al.: Climate prediction of tropical cyclone activity in the vicinity of Taiwan using the multivariate least absolute deviation regression method. *Terr. Atmos. Ocean. Sci.* **18**(4), 805–825 (2007)
8. Dong, K.: El Niño and tropical frequency in the Australian region and the northwest Pacific. *Aust. Meteorol. Mag.* **36**, 219–225 (1988)
9. Duong, T.H., Nguyen, D.C., Nguyen, S.D., Hoang, M.H.: An adaptive neuro-fuzzy inference system for seasonal forecasting of tropical cyclones making landfall along the Vietnam coast. In: *Advanced Computational Methods for Knowledge Engineering. Studies in Computational Intelligence*, vol. 479, pp. 225–236. (2013)
10. Fan, K., Wang, H.: A new approach to forecasting typhoon frequency over the western North Pacific. *weather Forecast.* **24**, 974–986 (2009). doi:10.1175/2009WAF222194.1
11. Goh, A.Z.C., Chan, J.C.L.: Variations and prediction of the annual number of tropical cyclones affecting Korea and Japan. *Int. J. Climatol.* **32**(2), 178–189 (2012). doi:10.1002/joc.2258
12. Goh, A.Z.C., Chan, J.C.L.: Interannual and interdecadal variations of tropical cyclone activity in the South China Sea. *Int. J. Climatol.* **30**, 827–843 (2010b)
13. Goh, A.Z.C., Chan, J.C.L.: An improved statistical scheme for the prediction of tropical cyclones making landfall in South China (2010c)

14. Gray, W.M., et al.: Atlantic seasonal hurricane frequency: part I: El Niño and 30 mb quasi-biennial oscillation influences. *Mon. Weather Rev.* **112**, 1649–1668 (1984)
15. Gray, W.M., et al.: Predicting Atlantic seasonal hurricane activity 6–11 months in advance. *Weather Forecast.* **7**, 440–455 (1992)
16. Kim, H.S., et al.: Seasonal prediction of summertime tropical cyclone activity over the East China Sea using the least absolute deviation regression and the Poisson regression. *Int. J. Climatol.* **30**, 210–219 (2010)
17. Kim, H.M., Webster, P.J., Curry, J.A.: Modulation of North Pacific tropical cyclone activity by three phases of ENSO. *J. Clim.* **24**, 1839–1849 (2011)
18. Landsea, C.W.: El Niño–Southern Oscillation and the seasonal predictability of tropical cyclones. In: Diaz, H.F., Markgraf V. (eds.) *El Niño: Impacts of Multiscale Variability on Natural Ecosystems and Society* (2000). (In press)
19. Liu, K.S., Chan, Johnny C.L.: Climatological characteristics and seasonal forecasting of tropical cyclones making landfall along the South China Coast. *Mon. Weather Rev.* **131**, 1650–1662 (2003)
20. Nguyen, S.D., Ngo, K.N.: An adaptive input data space partitioning solution to the synthesis of neuro-fuzzy models. *Int. J. Control Autom. Syst.* **6**, 928–938 (2008)
21. Nguyen, S.D., Choi, S.B.: A new neuro-fuzzy training algorithm for identifying dynamic characteristics of smart dampers. *Smart Mat. Struct.* **21**(8), 1–14 (2012). (Article No. 085021)
22. Nicholls, N.: A possible method for predicting seasonal tropical cyclone activity in the Australian region. *Mon. Weather Rev.* **107**, 1221–1224 (1979)
23. Nicholls, N.: Recent performance of a method for forecasting Australian seasonal tropical cyclone activity. *Aust. Meteorol. Mag.* **40**, 105–110 (1992)
24. Panella, M., Rizzi, A., Frattale Mascioli, F.M., Martinelli, G.: ANFIS synthesis by hyperplane clustering. In: *Proc. of IFSA/NAFIPS*, vol. 1, pp. 340–345, Vancouver (2001)
25. Panella, M., Gallo, A. S.: An input–output clustering approach to the synthesis of ANFIS networks. *IEEE Trans. Fuzzy Syst.* **13**(1), 69–81 (2012)
26. Shapiro, L.J.: The relationship of the quasi-biennial oscillation to Atlantic tropical storm activity. *Mon. Weather Rev.* **117**, 1545–1552 (1989)
27. Takagi, T., Sugeno, M.: Fuzzy identification of systems and applications to modeling and control. *IEEE Trans. Syst. Man Cybern.* **15**(1), 16–132 (1985)
28. Wang, B., Chan, J.C.L.: How strong ENSO events affect tropical storm activity over the western North Pacific*. *J. Clim.* **15**, 1643–1658 (2002)
29. Wu, M.C., Chang, W.L., Leung, W.M.: Impacts of El Niño–Southern Oscillation events on tropical cyclone landfalling activity in the western North Pacific. *J. Clim.* **17**, 1419–1428 (2004)
30. Zhao, H., Wu, L., Zhou, W.: Assessing the influence of the ENSO on tropical cyclone prevailing tracks in the western North Pacific. *Adv. Atmos. Sci.* **27**(6), 1361–1371 (2010)