Assessing the Forecasting of Comprehensive Loss Incurred by Typhoons: A Combined PCA and BP Neural Network Model

Shuai Yuan[1], Guizhi Wang[1, *], Jibo Chen[1] and Wei Guo[2]

Abstract: This paper develops a joint model utilizing the principal component analysis (PCA) and the back propagation (BP) neural network model optimized by the Levenberg Marquardt (LM) algorithm, and as an application of the joint model to investigate the damages caused by typhoons for a coastal province, Fujian Province, China in 2005-2015 (latest). First, the PCA is applied to analyze comprehensively the relationship between hazard factors, hazard bearing factors and disaster factors. Then five integrated indices, overall disaster level, typhoon intensity, damaged condition of houses, medical rescue and self-rescue capability, are extracted through the PCA; Finally, the BP neural network model, which takes the principal component scores as input and is optimized by the LM algorithm, is implemented to forecast the comprehensive loss of typhoons. It is estimated that an average annual loss of 138.514 billion RMB occurred for 2005-2015, with a maximum loss of 215.582 in 2006 and a decreasing trend since 2010 though the typhoon intensity increases. The model was validated using three typhoon events and it is found that the error is less than 1%. These results provide information for the government to increase medical institutions and medical workers and for the communities to promote residents' self-rescue capability.

Keywords: Typhoon, PCA, BP neural network model, comprehensive loss, LM algorithm.

1 Introduction

Meteorological disasters have caused numerous fatal disasters to life and serious damage to property every year, and according to China marine disaster bulletin, typhoons alone account for approximately 25% of live disasters and 94% of property loss in China. Particularly, the situation usually becomes the most dangerous along the coast. For example, in 2013 typhoons in China's coastal cities hit 1380.34 million people, caused 121 fatalities, and resulted in a direct loss of 152.45 billion RMB according to China marine disaster bulletin. With the acceleration of the urbanization process, while the

---

[1] School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing, 210044, China.

[2] Department of Mathematics & Computer Science, University of North Carolina at Pembroke, Pembroke, NC 28372, USA.

[*] Corresponding Author: Guizhi Wang. Email: wgz@nuist.edu.cn.

number of deaths by typhoons is decreasing, the comprehensive economic loss exhibits an increasing trend.

As such, disaster assessment becomes a critical part of prevention and reduction strategies, and the assessment result is widely regarded as an important basis for the rescuing operations including the distribution of emergency foods and relief funds. Consequently, disaster assessment is of great practical significance for disaster prevention, disaster reduction, safeguarding economic development, and maintaining people's lives and property.

The typhoon problem was drawn an attention as early as in 1920's which [Pielke and Landsea (2008)] employed a method of normalization the damages to study the trend of typhoons in the United States, linking social economic factors with typhoons, and results showed that population and currency factors were important for typhoon damages.

Mille first studied typhoon rainstorms quantitatively [Mille (1958)] through the hurricane model and found that the rainstorm and wind speed are the direct factors affecting the typhoon strength. Emanuel built the Carnot Cycle model [Emanuel (1987)] to estimate the maximum pressure intensity of typhoons, where the results of empirical analysis have showed that the wind speed and the pressure were associated with large intensity changes of typhoons.

Later, the recent works [Lu (1995)] built a disaster damage assessment model based on the number of collapsed houses, direct economic loss, and the area of affected farmland. The results showed that the assessment error for typhoons was less than 5%. Powell established a typhoon disaster assessment model [Powell and Houston (1996)] for estimating the damage by measuring typhoon landing and destroying infrastructure. After that, studying the damage caused by typhoon using the mixed Poisson stochastic model, and concluded [Katz (2002)] that the trend of typhoons in the Poisson process was only a weak evidence and the fluctuation of typhoon quantity was related to the change of the total number of annual damage.

With the improvement of statistical theory such as the multivariate regression and neural network, scholar [Baik and Hwang (2008)] used the multivariate regression method and neural network method to forecast typhoon intensities. To compare the two models, the neural network method had greater predictive potential and the prediction effect was better than the multivariate regression [Lou, Chen and Zheng (2009), Xu, Pan and Miao (2009)]. Zhang et al. [Zhang, Huang and Liu (2011)] provided the soft histogram method to evaluate the risk of land-falling based on the definition of typhoon in the Guangdong Province, China. And they concluded that the highest risk located were the western, central and eastern coast of Guangdong Province.

Matter-element model [Liu, Zhang and He (2010)] has been applied to the evaluation on loss in the 1980s. The matter-element evaluation model was built [Wei, Yang and Liu (2011)] to count for the application by the loss of typhoon Damrey, where the results of empirical analysis have showed that the model had feasible and valid application. Recent works divided [Kumar and Goyal (2013)] the year into four seasons and constructed the principal component analysis (PCA) and BP model to predict the air quality of the four seasons. The results showed that the error of the principal component BP model was 3%, which was

superior to other neural network models, and could predict the air quality effectively.

Recently grey techniques and models have been more widely applied for the prevention and assessment of natural disasters since 1990s [Yang (1997)]. Zhang et al. [Zhang and Zhong (2013)] employed the grayness relational grade to develop the evaluation index and used the evaluation results as input factors to analyze the losses caused by typhoons based on regression models. Results showed that the combined model performed better than the one used separately. In this paper, typhoons are investigated directly by the combined model.

At present, China suffers various impacts from severe typhoons, averaging nine landfalls in a year causing an estimated $3.9 billion in damages and 472 lives [Bakkensen (2013)]. The support vector regression (SVR) and parameter optimization method [Chen, Tang and Sui (2013)] were built which selected severe typhoons occurred in Guangdong Province from 1998 to 2008 were analyzed. The results showed that the SVR based on the genetic algorithm (GA) was the best model, among the three methods mentioned above, to estimate the loss of typhoon. The input-output model [Wang, Li, Chen et al. (2015)] assessed the indirect losses of meteorological disasters due to damaged industrial linkages and obtained that the secondary industry was more vulnerable to the impact of heavy rainstorms. The risk assessment model was presented to optimize the fishing vessel during typhoon emergencies [Zhao, Niu and Bai (2016)]. Results showed that using the hybrid closed loop algorithm was the best model based on Zhejiang Province, China.

The above researches on assessing disasters are mainly limited to the assessment of economic damage [Wang, Wu and Chen (2016)]. The comprehensive loss is an important indicator for measuring typhoons, is not included in the evaluation results of these methods. In recent years, scholars have begun to further study [Liu, Zhang and Yang (2017)] the comprehensive loss by using the classification method to assess the blackout in power grid. Based on statistic data of blackout in an actual city, comprehensive loss for blackout was calculated at 99.08 RMB/kWh. The comprehensive loss can be summed up as social indexes and economic indexes, in which social indexes refer to loss of casualties, and the economic indexes refer to direct economic loss and indirect economic loss [Wei and Zhang (1996); Yang (1997) Yang and Zhang (2010)]. Therefore, the comprehensive loss should include

1. The loss of casualties, which is created by medical expenses, take time off work because of disasters, recuperate and lose abilities to work and soon.

2. The direct economic loss, which is the most direct and important indexes reflecting the degree of disasters.

3. The loss of rescue, it refers to prices of labor power and physical resources paid by national economy for disaster rescue.

4. The benefit loss, it refers to the lost value because of the reduction of products caused by disasters.

5. The ecological environment loss. Ecological environment loss is difficult to estimate, we can choose the direct economic loss 1.5 times to replace, because of the damaged filed area is proportional to the ecological environment loss. The above data originates from "China Statistical Yearbook" and "China Meteorological Disaster

Yearbook".

However, the existing researches on typhoon assessment are generally carried out by putting together disaster-inducing factors and disaster environments, and assessment methods for damages include mathematical statistics, fuzzy mathematics, analysis hierarchy processes and multivariate regression method [Imamura and Dang (1997); Ramesh, Nagaraju and Ramanamurthy (2007)]. Since changes in the relation between the hazard factors, hazard bearing factors and disaster factors are not included in the investigation, these models have wide margins of error when applied to actual assessments [Lou, Chen and Qiu (2012)].

PCA and BP neural network model are highly effective in solving the disaster assessment as a single method. But, as far as we know, there are no studies that incorporates both PCA and BP neural network model to assess the comprehensive loss of typhoons. In addition, the BP neural network approach is not optimized in their studies, which would lower the precision [Wang, Song and Yin (2015); Lou, Chen and Zheng (2009); Guo, Dai and Li (2013)]. In this paper, we will apply the Levenberg Marquardt (LM) algorithm to optimize the PCA-BP model.

In this paper, we will investigate the relation among the hazard factors, hazard bearing factors and disaster factors, and subsequently apply the PCA method to preprocess the impact factors. Further, the resulting principal components become the inputs for the BP model, which is optimized by the LM algorithm to measure the comprehensive loss caused by typhoons.

For the remainder of the paper, Section 2 initiates the PCA, BP neural network model and LM algorithm. Section 3 applies the models established in Section 2 to assess the comprehensive loss by typhoons. In Section 4, we discuss the results and elaborate some limitations of the study. Section 5 provides the conclusion.

2 Methods

2.1 Principal component analysis

PCA as a statistical method turns multiple factors into fewer factors; it can not only explore large amount of information of the original data, but also simplify the data and reveal the relationship between variables. Hotelling first proposed the PCA [Hotelling (1933)] in 1923, which now has wide applications in the social economy, enterprise management, geology, biochemistry, medicine and other fields [Li and Sun (2010); Lin, Jiang and Guo (2014); Lu, Lee and Hadley (2014); Tarvainen, Cornforth and Jelinek (2014)]. This paper first proves the reliability of the selected factors by calculating the commonality, then preprocesses the factors according to the PCA, and extracts the principal component to lower the dimensions of data sets.

The commonality [Skinner 1985] refers to the proportion of common factors that explain variable variance. Assuming that each variable is affected by m common factors and a special factor, as follows:

$$X_i = a_{i1}f_1 + a_{i2}f_2 + ... + a_{im}f_m + e_i,$$ (1)

where X_i is the normalized variable with 0 mean and variance of $1, f_1, f_2, ...f_n$ are independent common factors with the mean of 0 and the variance of 1, e_i is the special factor, $a_{i1}, a_{i2}, ...a_{im}$ are factor loadings. The commonality of X_i is defined as:

$$a_{i1}{}^2 + a_{i2}{}^2 + ...a_{im}{}^2. \tag{2}$$

In general, when the commonality is more than 0.5, the variable can be selected in the model. PCA replaces the original indices with a new set of uncorrelated comprehensive indices. Assuming that we have p indices, the (i, j) element of covariance matrix is:

$$x_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j), i, j = 1, 2...p. \tag{3}$$

Then, the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_p$ of the covariance matrix are calculated and the eigenvectors γ_i corresponding to each eigenvalue determined:

$$\gamma_i = (\gamma_{i1}, \gamma_{i2}, ..., \gamma_{ip}), \ i = 1, 2, \cdots, p. \tag{4}$$

And the cumulative contribution rate v of each principal component is defined as $\lambda_i / \sum_{i=1}^{m} \lambda_i$. When the cumulative variance contribution rate is more than 85%, the selected principal components can be considered to fully reflect the information of original variables [Xie, Zhong and Cao (2015)]. Then we have the principal component scores by first calculating the principal component coefficient (PCC) matrix. The correlation between the principal components and the original variables can be obtained by the PCC matrix. The principal component regression equation can be written as:

$$F_i = a_{11}X_1 + a_{21}X_2 + ... + a_{p1}X_p, i = 1, 2...p, \tag{5}$$

where F_i is the i-th principal component. The principal component scores are $l_i = m_i F_i$, where m_i are defined as the normalized value of each variable. This paper uses the principal component scores as the input for the BP neural network model.

2.2 BP neural network model

2.2.1 Basic principle

A BP neural network model is a multilayer feed forward neural network composed of several network layers. Fig. 1 illustrates the structure diagram of three-layer BP neural network models. Rumelhart et al. [Rumelhart, Hinton and Williams (1986)] first proposed the BP neural network model, which now has numerous applications for pattern recognition, image processing, communication, drought and urban accident [Straub and Schroder (1996); Grossberg (1988); Nakamura, Yoshida and Engelmann (2000); Jia, Pan and Yuan (2015); Behbahani, Amiri and Imaninasab (2018)].
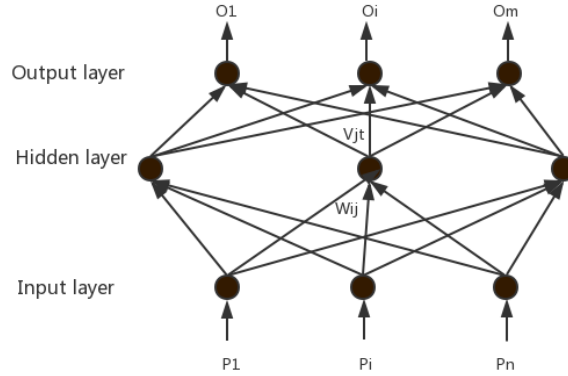
Figure 1: BP neural network model

A three-layer BP neural network can be utilized to approximate arbitrary nonlinear functions. Concurrently, the BP neural network model has the ability of self-learning: It can extract regularly information from data and memorize the weights, and hence it possesses a generalization ability. According to the activation function, the relationship between input and output of the input layer is:

$$\left[s_1, s_2, .., s_p\right] = \left[x_1, x_2, ...x_m\right] \times \begin{bmatrix} w_{11} & w_{12} & ... & w_{1p} \\ w_{21} & w_{22} & ... & w_{2p} \\ ... & ... & ... & ... \\ w_{m1} & w_{m2} & ... & w_{mp} \end{bmatrix} + \left[\theta_1, \theta_2, ...\theta_p\right], \tag{6}$$

where $S_k = \left(s_1, s_2, ...s_p\right)$ is the hidden layer's weighted input, $X_k = \left(x_1, x_2, ...x_m\right)$ is the sample input, w_{ij} is the weight from input layer to hidden layer, and θ_j is the hidden layer threshold for each node .

The relation between hidden layer weighted input and output is：

$$b_j = g\left(s_j\right), \tag{7}$$

where $B_p = \left(b_1, b_2, ...b_p\right)$ is the hidden layer output.

The relation between the output layer's weighted input and hidden layer output is:

$$\left[l_1, l_2, .., l_p\right] = \left[b_1, b_2, ...b_p\right] \times \begin{bmatrix} v_{11} & v_{12} & ... & v_{1n} \\ v_{21} & v_{22} & ... & v_{2n} \\ ... & ... & ... & ... \\ v_{p1} & v_{p2} & ... & v_{pn} \end{bmatrix} + \left[r_1, r_2, ...r_n\right], \tag{8}$$

where $L_p = \left(l_1, l_2, ...l_p\right)$ is the output layer's weighted input, v_{jt} is the weight from input layer to the output layer, r_l is the output layer threshold for each node.

The relation between the output layer's weighted input and target output is：

$$\hat{y}_t = g(l_t), \tag{9}$$

where $\hat{Y}_k = (\hat{y}_1, \hat{y}_2, \ldots \hat{y}_n)$ is the actual output.

The error function between target output and the actual output of the BP neural network model is:

$$E_k = \frac{1}{2}\|Y_k - \hat{Y}_k\|_2^2 = \frac{1}{2}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2, \tag{10}$$

where $Y_k = (y_1, y_2, \ldots y_n)$ is the target output.

2.2.2 Activation function settings

The activation function is a transformation function between input and output. Common activation functions are divided into threshold type, linear type and S type (Sigmoid). When the data is entered in the network, it is first transmitted from the input layer to the hidden layer. After activating the function, it is then transmitted to the next hidden layer and finally to the output layer. During this process, the data of each layer must be transformed by the corresponding activation function.

Because the Sigmoid function has the function of nonlinear amplification coefficient, the output value can be changed to (-1,1). For larger input samples, the amplification coefficient is small, while for smaller input samples, the amplification coefficient is larger, so the Sigmoid function is used to approximate the nonlinear input and output relationship. The Sigmoid function is microscopic and can prevent network saturation, thus providing good support for model accuracy. If the Sigmoid function is used in the output layer, the output is limited to (-1,1). So a linear activation function is used so that the output can be any value.

In addition, the BP network differs from other forward neural networks in that all hidden layer nodes in the network take Sigmoid functions, and the activation function requirements of the input layer must be microscopic. Therefore, we take the Sigmoid function for the activation function of the input layer and hidden layer. Tab. 1 shows two common Sigmoid functions, the tansig function and logsig function. As for the activation function of the output layer, we take linear function purelin.

Table 1: Activation functions

Function name	Function performance	Calculation formula	Range
logsig	Logarithmic activation function	$g(n) = \dfrac{1}{1+e^{-n}}$	(0,1)
tansig	Tangent activation function	$g(n) = \dfrac{2}{1+e^{-2n}} - 1$	(-1,1)
purelin	Linear activation function	$g(n) = n$	(-1,1）

2.3 LM algorithm optimization

The slow convergence rate of the standard BP algorithm will lead to increase in error. Hence, we need to optimize the BP algorithm in practical applications. At present, popular optimization algorithms include: Gauss-Newton algorithm, Conjugate Gradient algorithm and so on. This paper uses the LM algorithm to adjust the weights and optimizes the BP neural model, which holds advantages such as good robustness, fast convergence, non-uniform iteration direction, and hence greatly improves the network convergence speed and generalization ability. The LM algorithm modifies network weights according to the following formula:

$$\omega(n+1) = \omega(n) - [J^T(n)J(n) + \mu I]^{-1}J^T(n)E_n ,\tag{11}$$

where ω is the weight, I is an identity matrix, μ is the user-defined learning rate, J is the Jacobian matrix containing the first derivative of the error performance function.

In summary, the calculation steps of the LM algorithm are described as follows:

(1) Giving the allowable values of target error ε and initializing the weights and threshold vectors.

(2) Calculating the network output, the error function and Jacobian matrix.

(3) Calculating modified value of weights $\Delta\omega = \omega(n+1) - \omega(n)$.

(4) If $E < \varepsilon$, then the training stops. If $E \geq \varepsilon$, then the sum of the squared errors between the target output and the actual output is repeatedly calculated. When the sum of the squared errors reaches the target error, the training stops and the network converges.

In this paper, PCA and BP neural network are incorporated so that the output of the PCA becomes BP's input. In contrast, if the direct input is provided to the BP neural network model, it will not only enlarge the complexity of the network, but also affects the convergence speed and self-correction ability of the model.

Because many factors are disorderly and complicated, this paper can extract the important information of the factors through PCA, and integrate them into specific indicators, which greatly reduces the number of variables and more clearly reveals the relationship between variables, thus improving the calculation accuracy. The purpose of the combined model lies to greatly clarify the input structure of the neural network, enhance the fitting ability of neural network model and improve the prediction accuracy of the model.

3 Analysis of empirical results

Disaster data of typhoons from 2005-2015 (most recent data available) are derived from the "China Meteorological Disaster Yearbook (Fujian Volume)". The economic data are referred to the "China Statistical Yearbook". This paper chooses Fujian Province as the research area because it is a coastal province and suffers most frequently from typhoon disasters among all the provinces in China (33 typhoons making landfall in Fujian province in 2005-2015).

3.1 Selection of impact factors

The disaster situation is composed of hazard factors, hazard bearing factors and disaster factors. The hazard factors of typhoon refer to the wind speed and rainstorm carried by typhoon, and the hazard bearing factors are the main body of direct loss caused by typhoon. Disaster factors refer to the integration of people and objects in a disaster. We select 15 factors from three aspects, hazard factors, hazard bearing factors and disaster factors, to study the comprehensive losses caused by typhoons, in which hazard factors are $x_1 - x_2$, disaster factors are $x_3 - x_8$ and hazard bearing factors are $x_9 - x_{15}$ (Tab. 2). In this setting, hazard factors and disaster factors are regarded as the comprehensive influencing factors of typhoons, and hazard bearing factors are regarded as the strain of the disaster. $x_{10} - x_{11}$ can reflect the exposure of hazard bearing factors, $x_{14} - x_{15}$ can reflect the vulnerability of hazard bearing factors, which not only measure the positive impact on the resilience of typhoons, but also consider their negative impact.

Table 2: Factor description

Variable	Hazard factors	Disaster factors	Hazard bearing factors
x_1	Maximum wind speed		
x_2	Maximum rainstorm		
x_3		Number of collapsed houses	
x_4		Number of damaged houses	
x_5		Damaged of fields area	
x_6		Destructed fields area	
x_7		People affected by disaster	
x_8		Direct economic loss	
x_9			Emergency transferred population
x_{10}			Proportion of male population
x_{11}			Gross regional product (GRP)
x_{12}			Number of medical institutions
x_{13}			Number of medical workers
x_{14}			Gross domestic product per capita
x_{15}			Health expenditure per capita

3.2 The calculate of principal component

The commonality of each factor is firstly calculated (Tab. 3). The commonality of each factor is more than 0.5, which proves (Refer to "Section 2.1") the reliability of the selected factors.

Table 3: The commonality of factors

Variable	Initial	Commonality	Variable	Initial	Commonality
x_1	1.000	0.902	x_9	1.000	0.874
x_2	1.000	0.710	x_{10}	1.000	0.722
x_3	1.000	0.924	x_{11}	1.000	0.964
x_4	1.000	0.962	x_{12}	1.000	0.767
x_5	1.000	0.956	x_{13}	1.000	0.924
x_6	1.000	0.822	x_{14}	1.000	0.969
x_7	1.000	0.921	x_{15}	1.000	0.958
x_8	1.000	0.828			
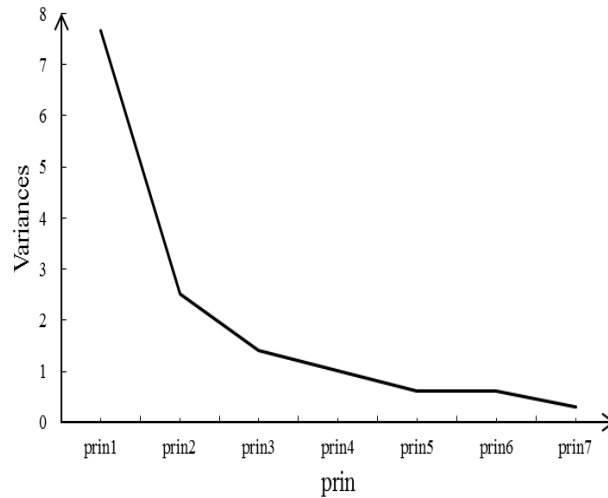
Figure 2: Scree plot

In order to eliminate the impact of dimension, the data is further standardized. Then the appropriate principal components are selected by PCA. Fig. 2 gives the scree plot of PCA, and it can be shown that when five principal components are taken, the trend begins to stabilize. In addition, for the top five principal components, their total cumulative contribution rate has reached 88.02% (Tab. 4), and hence we select the top five principal components for analysis. The principal component was calculated and the results are obtained in Tab. 5.

Table 4: Variance contribution rate

Prin	Eigenvalues	Variance contribution rate	Cumulative contribution rate
Prin1	7.661	51.076	51.076
Prin2	2.514	16.759	67.835
Prin3	1.404	9.362	77.197
Prin4	1.014	6.763	83.960
Prin5	0.609	4.060	88.021

Table 5: PCC matrix

Variable	prin1	prin2	prin3	prin4	prin5
x_1	0.140	0.431	-0.158	0.011	-0.639
x_2	0.225	-0.120	0.116	0.444	-0.329
x_3	0.243	0.232	-0.432	-0.168	0.272
x_4	0.232	0.248	-0.473	-0.246	0.188
x_5	0.310	0.217	0.234	0.127	0.126
x_6	0.290	0.058	0.271	0.151	0.258
x_7	0.300	0.165	0.285	0.196	0.135
x_8	0.220	0.411	0.111	0.070	0.144
x_9	0.312	0.098	0.208	-0.176	-0.131
x_{10}	-0.202	0.353	0.006	-0.059	0.388
x_{11}	-0.310	0.278	0.068	0.120	0.138
x_{12}	-0.274	0.245	0.089	0.122	0.149
x_{13}	-0.021	0.044	0.506	-0.740	-0.060
x_{14}	-0.312	0.276	0.085	0.106	0.135
x_{15}	-0.302	0.297	0.104	0.100	0.149

According to the coefficient matrix in Tab. 5, five principal components expressions are obtained:

$$prin1 = 0.140x_1 + 0.225x_2 + 0.243x_3 + 0.232x_4 + 0.310x_5 + 0.290x_6 + 0.300x_7 + 0.220x_8$$
$$+0.312x_9 - 0.202x_{10} - 0.310x_{11} - 0.274x_{12} - 0.021x_{13} - 0.312x_{14} - 0.302x_{15}$$
(12)

$$prin2 = 0.431x_1 - 0.120x_2 + 0.232x_3 + 0.248x_4 + 0.217x_5 + 0.058x_6 + 0.165x_7 + 0.411x_8$$
$$+0.098x_9 + 0.353x_{10} + 0.278x_{11} + 0.245x_{12} + 0.044x_{13} + 0.276x_{14} + 0.297x_{15}$$
(13)

$$prin3 = -0.158x_1 + 0.116x_2 - 0.432x_3 - 0.473x_4 + 0.234x_5 + 0.271x_6 + 0.285x_7 + 0.111x_8$$
$$+0.208x_9 + 0.006x_{10} + 0.068x_{11} + 0.089x_{12} + 0.506x_{13} + 0.085x_{14} + 0.104x_{15}$$
(14)

$$prin4 = 0.111x_1 + 0.444x_2 - 0.168x_3 - 0.246x_4 + 0.127x_5 + 0.151x_6 + 0.196x_7 + 0.070x_8$$
$$+0.312x_9 - 0.202x_{10} - 0.31x_{11} - 0.274x_{12} - 0.021x_{13} - 0.312x_{14} - 0.302x_{15} \tag{15}$$

$$prin5 = -0.639x_1 - 0.329x_2 + 0.272x_3 + 0.188x_4 + 0.126x_5 + 0.258x_6 + 0.135x_7 + 0.144x_8$$
$$-0.131x_9 + 0.388x_{10} + 0.138x_{11} + 0.149x_{12} - 0.06x_{13} + 0.135x_{14} + 0.149x_{15} \tag{16}$$

The coefficients of the hazard factors and disaster factors in *prin*1 are all positive, and hence the comprehensive variable *prin*1 describes the overall disaster level. Higher value of *prin*1 would indicate a stronger destructive disaster.

The maximum coefficient in *prin*2 is 0.431, which corresponds to the maximum wind speed describing the typhoon intensity.

The coefficients of the number of collapsed houses and the number of damaged houses is -0.432 and -0.473 in *prin*3, and hence *prin*3 describes the damaged condition of houses. These are negative loadings. The greater the *prin5*, the less the number of collapsed houses and damaged houses meaning that the damaged condition of houses is slight.

In *prin*4, the maximum coefficient -0.740, corresponds to the number of medical works, and hence *prin*4 describes the medical rescue. The greater the *prin*4, the less the medical workers, meaning that the medical rescue operation is more difficult and more personnel would be needed.

Table 6: The principal component scores

Number	Prin1	Prin2	Prin3	Prin4	Prin5	Number	Prin1	Prin2	Prin3	Prin4	Prin5
0505	3.40	-0.84	3.56	-4.28	-0.39	1013	-0.93	0.26	-0.3	-0.12	-0.62
0519	4.67	0.60	1.18	1.03	0.17	1103	-1.88	-2.26	-0.24	0.17	1.86
0601	4.84	0.03	1.97	1.32	0.70	1111	-0.33	-0.91	0.26	0.69	0.67
0604	5.99	1.30	-0.64	1.05	0.99	1205	-2.02	-0.57	-0.08	-0.04	0.70
0605	2.95	-0.66	-1.10	-0.67	0.52	1209	-1.46	-0.14	0.30	-0.12	1.06
0608	5.72	3.48	-4.05	-1.80	0.11	1307	-1.86	1.33	0.40	-0.03	-0.46
0709	3.56	-0.74	0.67	0.77	-0.58	1308	-2.61	0.44	0.07	-0.20	0.73
0713	1.31	-1.76	-0.80	-0.60	0.15	1312	-2.00	1.42	0.17	-0.05	-0.62
0716	0.85	-2.31	-0.84	0.06	-1.25	1315	-3.32	-0.05	-0.23	-0.42	0.52
0801	-0.49	-2.67	-1.17	-0.50	0.47	1319	-1.93	1.25	0.21	0.31	-0.47
0807	-0.09	-2.49	-0.85	-0.22	-0.09	1323	-1.96	1.67	0.00	0.09	-0.87
0808	2.18	-1.41	0.15	1.51	-1.63	1407	-3.41	0.45	0.05	-0.21	0.57
0903	-0.03	-1.81	0.10	0.71	0.57	1410	-2.64	1.24	0.29	0.00	-0.19
0907	1.66	-0.09	0.75	0.88	-0.47	1510	-3.75	0.85	-0.25	-0.35	0.33
1006	-1.30	-1.01	-0.35	0.40	-0.86	1513	-0.14	4.10	1.52	0.62	0.69
1010	-1.38	-0.21	-0.80	-0.21	-0.91	1521	-2.62	1.50	-0.33	-0.12	-0.61
1011	-0.92	0.02	-0.35	-0.00	-1.06						

In addition to the maximum coefficient of the typhoon intensity in *prin*5, 0.639, the next loading corresponds to the proportion of the male population, 0.388, indicating that the more people are able to-participate to strengthen the labor force. This is the self-rescue capability.

By utilizing Eqs. (12)-(16), the principal component scores for each of the 33 typhoons in Fujian Province during 2005-2015 are calculated (Tab. 6).

In Tab. 6, the scores of *prin*1 are generally larger than other principal components, and this implies more original information is extracted with *prin*1. Hence, *prin*1 corresponds to the comprehensive loss.

In addition, the principal component scores have positive and negative results because data are normalized throughout the process, which indicate the positional relationship with the average value. It can be seen that the scores of *prin*1 have decreased from the Typhoon 1006 to about -2.0, indicating that the comprehensive losses are lower than the average value, and reflecting that the government and the public are paying more attention to the impact of typhoons. Although typhoons occurred in 2013 the most often, but fortunately the damage is not the worst. The degree of awareness and response to the disasters have also been strengthened. For example, we can take measures such as increasing the medical workers, medical institutions, and medical expenses to promote people's self-rescue capability and awareness of prevention and to reduce the disasters of loss.

Further, the principal component scores of the Typhoons 0601, 0604 and 0608 are higher at about 5.0. In fact, these three typhoons are indeed causing huge comprehensive losses of 196.962, 215.582 and 200.596 billion RMB respectively. And it is estimated that an average annual loss of 138.514 billion RMB occurred for 2005-2015. Therefore, the principal component scores accurately reflect the actual damages, cover a large amount of information of the original variables, and can replace the original input for the BP's input.

3.3 BP neural network model setting

Based upon the principal component scores in Section 3.2, the BP neural network model is built to estimate the comprehensive losses caused by typhoons. To ensure the reliability and validity of the model and according to the chronological order of typhoon, 30 typhoon samples are used in training samples for simulation and fitting of the BP neural network, and the rest 3 samples (Typhoons 1410, 1513, and 1521) are taken to examine the model. When the relative error between the fitted and actual values of the test sample is less than 1%, the establishment of the BP model to achieve the target accuracy can provide relevant departments with corresponding suggestions to reduce the comprehensive losses caused by the typhoons.

The principal component scores are used as the BP neural network input and the comprehensive losses are the network output. Therefore, the node of the input layer is set to 5 and the node of the output layer is set to 1. The model parameters are set to: training steps=5000, target error=0.01, learning rate=0.01, momentum coefficient=0.9. After multiple trainings, it is found that when the node of the hidden layer is 10, the mean-squared error reaches the minimum value, and hence the structure of the 5-10 BP neural network is determined.

Since the activation function used by the BP neural network has a limited range of input

and output, the data needs to be normalized before the model is built. Based on the normalized data, the error and regression effects of the training samples are obtained, and the target accuracy and correlation coefficient of the model are investigated. Fig. 3 and Fig. 4 show the error curve and regression plot.
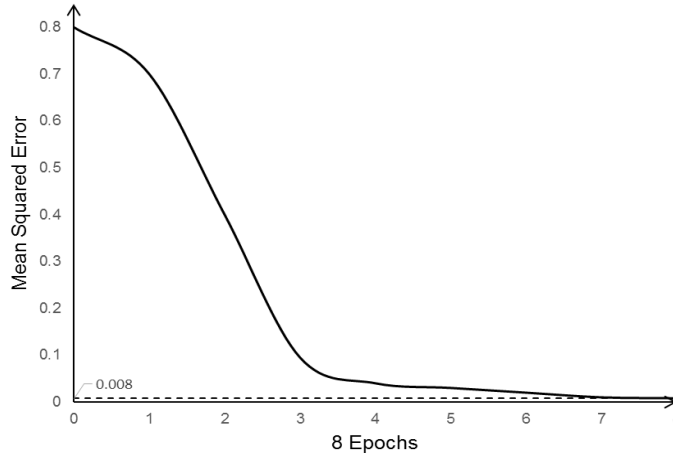

Figure 3: The error curve

According to the Fig. 3, the error begins to reach the target accuracy of 0.01 and continues to decline after 8 trainings, the network stops training and the final error is 0.008. The linear combination of the training samples obtained from Fig. 4 is:

$$\hat{Y}_k \approx 0.98 \times Y_k \pm 0.0021 , \tag{17}$$

where the correlation coefficient reaches 0.9856, implying that the fitting effect of the equation is well and the overall training effect is good. The 5-10-1 BP neural network model can be constructed.
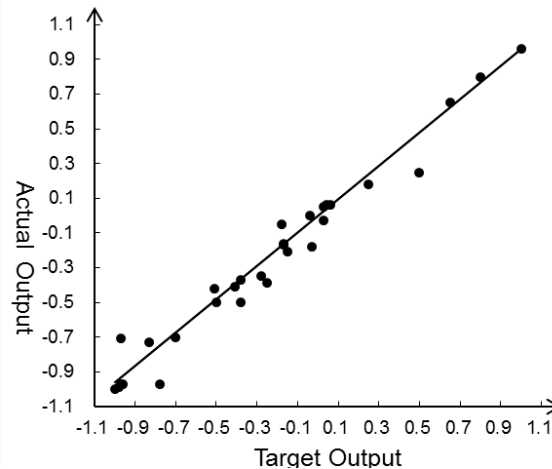
Figure 4: Regression plot

Further we use the trained 5-1-10 BP neural network model to fit the comprehensive losses of the test sample, and compare the actual value of the training sample with the BP fitting value. Fig. 5 shows the concrete results, indicating that the actual values are basically consistent with the BP fitting values. This shows that the BP neural network is fully learned and the trained network can embody the comprehensive losses caused by the typhoons.
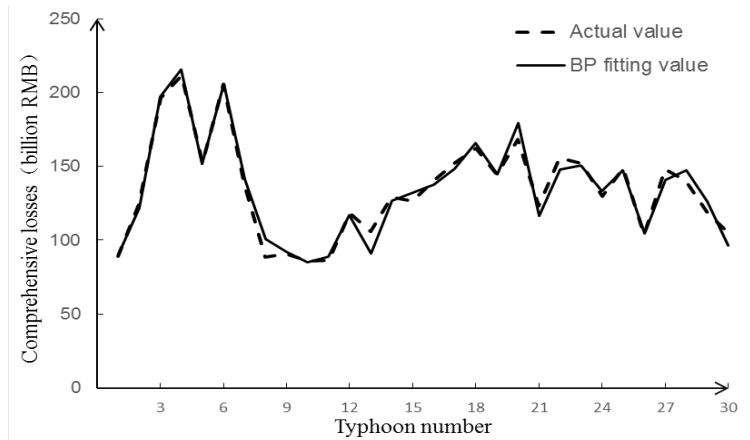
Figure 5: Actual value and BP fitting value

According to Tab. 7, the BP neural network model has a good prediction effect on the comprehensive losses of typhoons. The relative errors of the Typhoons 1410, 1513 and 1521 are respectively 0.9%, 0.58% and -0.5%, which are far less than 10%.

Next, from the trend in Fig. 5, the comprehensive losses from the Typhoon1006 to the Typhoon 1407 have been reduced. This is related with nationwide preventive measures in the past 10 years to carry out activities and popularize knowledge so as to improve people's self-rescue capability. At the same time, the relative error of the comprehensive loss of the training sample is within 10%, which fully shows that the BP neural network model has high accuracy and good effect. Therefore, the study of the typhoons by establishing the BP neural network model can play a role in reducing loss and disaster prevention.

Table 7: The actual value, fitting value and error of comprehensive loss of typhoons

Number	The actual value (billion RMB)	The fitting value (billion RMB)	Estimated error	Relative error (%)
1410	145.574	144.2178	1.3562	0.9
1513	247.835	246.3915	1.4435	0.58
1521	132.036	132.7879	-0.75	-0.5

4 Discussions

This study analyzes the impact factors and comprehensive loss of typhoons in Fujian province from 2005-2015. According to the results of Tab. 6, the comprehensive loss

caused by typhoons in 2006 reached a maximum of 215.582 billion RMB, and the comprehensive loss decreased in 2010. Moreover, Fujian province had begun to attach importance to the impact of typhoons, and carried out popularization of safety precautions since 2010, a provision of further support for our findings.

Assessment the loss of typhoons indicate that quantitative assessment is crucial for disasters assessment, which is in line with the work by Lou et al. [Lou, Chen and Zheng (2009)]. But the assessment model precision is 7% in error because Lou did not consider the impact of hazard bearing factors and disaster factors on the loss. However, Guo et al. [Guo, Dai and Li (2013)] uses the principal component BP model to predict the deviation of typhoons, and found that the maximum wind speed and the central pressure are the main factors that determined the typhoons deviation. The results indicate that the precision of this model is 6% in error and the model can predict 12 hours warning before typhoons.

Moreover, the principal component BP model can be applied to the air pollution field. Kumar et al. [Kumar and Goyal (2013)] constructed a principal component BP model to predict the air quality from 2000-2006. The results show that the accuracy of the principal component BP model is superior to other neural network models and the error of model is about 3%.

In this paper, PCA is used to reduce the dimension of hazard factors, hazard bearing factors and disaster factors, which eliminates the correlation between factors. Then we use LM algorithm to optimize the BP model. By forecasting the comprehensive loss of typhoons in Fujian province, we achieve the accuracy of 99%, or 1% in error. Comparing with the above research results, the accuracy of the model is further improved.

Due to the remarkably uncertain nonlinear relationship between factors and typhoons, PCA can eliminate the uncertain nonlinear relationship between the factors. The PCA is applied to evaluate the comprehensive loss of typhoons, which can reduce the uncertainly of factors selected in the previous researches. At the same time, extracting the principal component scores as BP's input, we can obtain the results which are consistent with disasters.

Although this paper is based on Fujian province as the research object and the final mathematical model cannot be directly applied to other provinces, assessment factors selected in this model are relatively easy to obtain. As long as there are data from other provinces, the corresponding model can also be constructed to predict the results using the method of this paper. Therefore, the PCA-BP model can be used in the assessment of typhoon damages in other areas, and this is a promotion value.

In addition, because typhoon data is difficult to collect and meteorological department no longer statistics the number of damaged houses in recent years. We use methods for dealing with missing data to solve, which reduce the accuracy of the model in a certain extent. However, any model can only approximate the reality of reasonable simplification.

5 Conclusions

In this paper, a joint model which utilize the PCA and the BP neural network model optimized by the LM algorithm is employed to assess the damages caused by typhoons for a coastal province of China. Firstly, we comprehensively study the relationship between the

hazard factors, hazard bearing factors and disaster factors by calculating the commonalities, and the problem of vulnerability and exposure of hazard bearing body are investigated. Then, we apply the PCA to reduce the dimensions and apply appropriate weights to obtain the principal component scores, which become the simplified input for the BP neural network model. Finally, the BP neural network model is optimized by the LM algorithm to assess the comprehensive loss. Moreover, an empirical analysis is conducted for the comprehensive loss by typhoons for Fujian province in 2005-2015. Based on the related literatures and empirical analysis, the following conclusions can be drawn.

1. Five principal components, *prin*1-*prin*5 are extracted from 15 impact factors by using PCA, which correspond to five principal indices, overall disaster level, typhoon intensity, medical rescue, damaged condition of houses, and self-rescue capability. Due to climate changes, the stronger wind speed will increase the typhoon intensity over time; however, the loss decreases since 2010, which is largely due to widely implemented prevention and rescue strategies.

2. Typhoons occurred in 2013 the most often, but fortunately the damage is not the worst. The comprehensive loss caused by typhoons in 2006 reached the maximum damage, with a loss of 215.582 billion RMB. For the typhoons 1410, 1513 and 1521 happened in 2014-2015 (Tab. 6) that were used to examine the joint model established from 30 typhoons utilizing the PCA and BP neural network model, the comprehensive losses were estimated as 145.574, 247.835, 132.036 billion RMB respectively, which contain the relative errors of 0.9%, 0.58% and -0.5% and justify the accuracy of the model.

In summary，huge damages that already happened and future potential damages by typhoons cannot be overlooked. The study suggests to take effective actions: for the government to increase medical institutions and medical workers and for communities to promote residents' self-rescue capability.

References

Baik, J.; Hwang, H. S. (2008): Tropical cyclone intensity prediction using regression method and neural network. *Journal of the Meteorological Society of Japan*, vol. 76, no. 5, pp. 711-717.

Bakkensen, L. A. (2013): The impact of climate and socioeconomic change on typhoon losses in China. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2637373.

Behbahani, H.; Amiri, A. M.; Imaninasab, R. (2018): Forecasting accident frequency of an urban road network: a comparison of four artificial neural network techniques, vol 37, no 7, pp. 767-780. https://onlinelibrary.wiley.com/doi/full/10.1002/for.2542.

Chen, S. H.; Tang, D. L.; Sui, G. J. (2013): Estimating of typhoon disaster loss for Guangdong Province based on SVR model. *Marine Environmental Science*, vol. 32, no. 6, pp. 939-943.

Emanuel, K. (1987): The dependence of hurricane intensity on climate. *Nature*, vol. 326, no. 6112, pp. 483-485.

Grossberg, S. (1988): Nonlinear neural networks: principles, mechanisms, and architectures. *Neural Networks*, vol. 1, no. 1, pp. 47-61.

Guo, Z. Y.; Dai, X. Y.; Li, X. X. (2013): Predict typhoon-induced storm surge deviation in a principal component back-propagation neural network model. *Chinese Journal of Oceanology and Limnology*, vol. 31, no. 1, pp. 219-226.

Hotelling, H. (1933): Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417-441.

Imamura, F.; Dang, V. T. (1997): Flood and typhoon disasters in Viet Nam in the half century since 1950. *Natural Hazards*, vol. 15, no. 1, pp. 71-87.

Jia, H.; Pan, D.; Yuan, Y. (2015): Using a BP neural network for rapid assessment of populations with difficulties accessing drinking water because of drought. *Human & Ecological Risk Assessment: An International Journal*, vol. 21, no. 1, pp. 100-116.

Katz, R. W. (2002): Stochastic modeling of hurricane damage. *Journal of Applied Meteorology*, vol. 41, no. 7, pp. 754-762.

Kumar, A.; Goyal, P. (2013): Forecasting of air quality index in Delhi using neural network based on principal component analysis. *Pure & Applied Geophysics*, vol. 170, no. 4, pp. 711-722.

Li, H.; Sun, J. (2010): Forecasting business failure in China using case-based reasoning with hybrid case representation. *Journal of Forecasting*, vol. 29, no. 5, pp. 486-501.

Lin, N.; Jiang, J.; Guo, S. (2014): Functional principal component analysis and randomized sparse clustering algorithm for medical image analysis. *PloS One*, vol. 10, no. 7, pp. 616-625.

Liu, S.; Zhang, J.; He, Z. (2010): Study on assessment model of typhoon disaster losses based on GIS. *Journal of Catastrophology*, vol. 25, no. 2, pp. 64-67.

Liu, Z. F.; Zhang, Z. B.; Yang, B. (2017): Evaluation of great blackout social comprehensive loss of power grid. *Power System Technology*, vol. 41, no. 9, pp. 2928-2938.

Lou, W. P.; Chen, H. Y.; Zheng, F. (2009): Economic loss assessment of typhoon based on principal component analysis and neural network. *Geographical Research*, vol. 28, no. 5, pp. 1243-1254.

Lou, W. P.; Chen, H. Y.; Qiu, X. F. (2012): Assessment of economic losses from tropical cyclone disasters based on PCA-BP. *Natural Hazards*, vol. 60, no. 3, pp. 819-829.

Lu, M.; Lee, H. S.; Hadley, D. (2014): Logistic principal component analysis for rare variants in gene-environment interaction analysis. *Transactions on Computational Biology & Bioinformatics*, vol. 11, no. 6, pp. 1020-1028.

Lu, W. F. (1995): Assessment and prediction of disastrous losses due to tropical cyclones in Shanghai. *Journal of the Natural Disasters*, vol. 4, no. 3, pp. 40-45.

Miller, B. I. (1958): O the maximum intensity of hurricanes. *Journal of the Atmospheric Sciences*, vol. 15, no. 15, pp. 184-195.

Nakamura, K.; Yoshida, H.; Engelmann, R. (2000): Computerized analysis of the likelihood of malignancy in solitary pulmonary nodules with use of artificial neural networks. *Radiology*, vol. 214, no. 3, pp. 823-830.

Pielke, R. A.; Landsea, C. W. (2008): Normalized hurricane damages in the United States: 1925 95. *Weather & Forecasting*, vol. 13, no. 3, pp. 621-631.

Powell, M. D.; Houston, S. H. (1996): Hurricane Andrew's landfall in South Florida. Part II: Surface wind fields and potential real-time applications. *Weather & Forecasting*, vol. 11, no. 3, pp. 329-349.

Ramesh, K. J.; Nagaraju, A. R.; Ramanamurthy, M. V. (2007): Framework development of hydrometeorological observational network and hazard mitigation modelling systems in respect of floods and cyclones. *Natural Hazards*, vol. 41, no. 3, pp. 531-548.

Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. (1986): Learning internal representations by error propagation. *Nature*, vol. 323, no. 6088, pp. 533-536.

Skinner, C. J. (1985): An introduction to multivariate statistical analysis by T. W. Anderson. *Journal of the Royal Statistical Society*, vol. 148, no. 2, pp. 164-165.

Straub, S.; Schroder, D. (1996): Identification of nonlinear dynamic systems with recurrent neural networks and Kalman filter methods. *International Symposium on Circuits and Systems*, vol. 3, pp. 341-344.

Tarvainen, M. P.; Cornforth, D. J.; Jelinek, H. F. (2014): Principal component analysis of heart rate variability data in assessing cardiac autonomic neuropathy. https://ieeexplore.ieee.org/document/6945157.

Wang, D.; Song, X.; Yin, W. (2015): Forecasting core business transformation risk using the optimal rough set and the neural network. *Journal of Forecasting*, vol. 34, no. 6, pp. 478-491.

Wang, G. Z.; Li, X.; Chen, J. B.; Wu, X. H. (2015): The rainstorm indirect economic loss assessment of multi departments based on IO model-A case study on the rainstorm on July, 21st in Beijing. *Journal of Catastrophology*, vol. 30, no. 2, pp. 94-99.

Wang, G. Z.; Wu, L. Y.; Chen, J. B. (2016): Intensity and economic loss assessment of the snow, low-temperature and frost disasters: a case study of Beijing City. *Natural Hazards*, vol. 84, no. 1, pp. 1-15.

Wei, H. J.; Yang, L.; Liu, R. J. (2011): Application of matter-element model in the typhoon loss evaluation of rubber plantation. *Guangdong Agricultural Sciences*, vol. 38, no. 3, pp. 168-171.

Wei, Q. C.; Zhang, Q. H. (1996): Determination of disaster losses and disaster grading. *Journal of Catastrophology*, vol. 11, no. 1, pp. 1-5.

Xie, L. W.; Zhong, J.; Cao, F. X. (2015): Evaluation of soil fertility in the succession of karst rocky desertification using principal component analysis. *Solid Earth*, vol. 6, no. 2, pp. 3333-3359.

Xu, X. Z.; Pan, W. Z.; Miao, Q. L. (2009): Disaster risk assessment of tornado in Jiangsu Province. *Transactions of Atmospheric Sciences*, vol. 32, no. 6, pp. 792-797.

Yang, S. X.; Yang, L. Q.; Zhang, H. Y. (2010): Research on the application of meteorological disaster losses assessment technology. *Meteorological and Environmental Research*, vol. 1, no. 4, pp. 97-101.

Yang, S. S. (1997): Study on the model grade division of natural disaster and comparison of disastrous conditions. *Journal of the Natural Disasters*, vol. 6, no. 1, pp. 8-13.

Zhao, H.; Niu, C.; Bai, X. (2016): A nonlinear optimization model for fishing vessel evacuation during typhoon emergencies. *Human & Ecological Risk Assessment: An International Journal*, vol. 23, no. 3, pp. 457-473.

Zhang, J.; Huang, C.; Liu, X. (2011): Risk assessment of land-falling tropical cyclones in guangdong province, china. *Human & Ecological Risk Assessment: An International Journal*, vol. 17, no. 3, pp. 732-744.

Zhang, Y. C.; Zhong, L. J. (2013): The research and analysis on typhoon disasters based on grayness correlation and regression analysis. *Journal of Tropical Meteorology*, vol. 29, no. 4, pp. 665-671.