Research on Geological Disaster Forecast Method Based on PCA Structure BP Model

Zhangbao Luan*
Guangzhou College of Commerce, Guangzhou 511363 China

*Corresponding author e-mail: 20031011@gcc.edu.cn

Abstract. The BP neural network prediction method constructed by PCA and the geological hazard prediction method based on the MM5 numerical model were used to establish geological hazard classification short-term objective forecast models. The calculation results show that these two objective forecast methods have a good fitting effect on historical samples. The independent sample's trial report effect is also good; based on the above two objective forecasting methods, through correction, the comprehensive forecast product is finally obtained.

Keywords: Principal Component Analysis; Neural Network; Numerical Model; Geological Disaster.

1. Introduction
Geological disasters refer to disasters related to geological processes such as mountain collapses, landslides, mudslides, ground subsidence, ground fissures, and ground subsidence that endanger people's lives and property safety caused by natural factors or human activities. In recent years, various geological disasters (except earthquakes) have caused an average of more than 1,000 deaths each year, and economic property losses have reached tens of billions. The problem of geological disasters has become an issue of great concern to people. According to statistics, more than 80% of geological disasters are related to precipitation. Rainfall not only saturates and softens the soft rock, increases its own weight, but also reduces the cohesion of the rock mass. Especially in the flood season, due to the influence of meteorological factors, sudden geological disasters such as collapses, landslides, and mudslides occur frequently. Decision-making meteorological service is an important decision basis for meteorological departments to improve the benefits of disaster prevention and mitigation for all levels of party committees and governments. It is the top priority of meteorological services. Although the effectiveness of decision-making services has been significantly improved compared with the past, the overall level of meteorological services is still far far from meeting the needs of party and government departments at all levels and the broad masses of the people, how to further improve the decision-making meteorological service capabilities and better serve as a good adviser for the local party and government departments is an important issue facing all meteorological workers. Continue to practice, explore and innovate in future work [1].

Fu Zongyu et al. (2020) based on the ensemble forecast data of the ECMWF model and used the joint probability method to establish two ensemble forecasting business products suitable for the two types of severe weather in winter in Beijing, China. Tests show that: when the forecast probability
reaches 10% and above, the actual situation may reach the level of early warning signals; this method has better forecasting performance for the northwestern part of Beijing, China, followed by the southeastern part of Beijing, China; for reaching the blue warning The signal standard area has a higher forecast hit rate[2].

Based on synoptic principles and synoptic analysis methods, Yu Zhiming et al. (2018) used Micaps, satellite cloud images, soundings, national basic meteorological observatories, and large buoy stations to analyze the Bohai Sea from 2001 to 2015 due to strong winds, fog, and fog. The weather cases of 78 shipwreck accidents caused by strong convection are analyzed. The results show that the winds in the Bohai Sea are dominated by northerly winds in winter and spring, followed by southerly winds. The northerly winds and cold air can be divided into three paths. The southerly windy ground pressure field can be divided into northeast low pressure type and North China topographic trough type. Heavy fog on the Bohai Sea mostly occurs in autumn and winter. Frequent advection fog enters and affects the Bohai Sea and coastal areas through three routes: east, southeast and southwest [3].

Hou Shumei et al. (2014) classified four types of strong winds along the coast of the Yellow Sea and the Bohai Sea from 2008 to 2011 according to the system. For different types of strong winds, they made statistics on the forecast accuracy of T639 numerical forecast products. The results show that the T639 numerical forecast has a better forecast accuracy rate for the coastal gales of the Yellow Sea and Bohai Sea, and the rate of underreporting is low; the forecasting ability of typhoons is biased; the forecast value is smaller than the actual situation, when the forecast has a cyclone or the forecasting time is longer When the range is large, the actual wind will increase by one to two orders of magnitude; the forecast of the start time of the gale is slightly earlier, and the forecast of the end time of the gale and the start and end times of the maximum wind speed are slightly later [4].

Nongjifu et al. (2008) carried out principal component analysis on the previous predictors of monthly average precipitation to achieve optimal compression of samples, thereby reducing the dimension of samples, and establishing a neural network based on principal component analysis for May average precipitation in northwestern Guangxi Prediction model. The calculation results show that the neural network model based on principal component analysis has better results than the multiple regression model in prediction [5].

2. Model and Data
At present, the domestic objective forecast methods for the grade of geological disasters mostly use linear statistical methods [6]. In this paper, the precipitation forecast field provided by the localized MM5 numerical model is coupled with the geological disaster classification model to form a geological disaster classification numerical prediction field; in addition, it also discusses the use of nonlinear artificial neural network statistical methods [7] to establish objective statistics Forecast tool. Based on the above objective forecast products, a comprehensive service platform for the graded forecast of geological disasters in Guangxi will be established. These studies have further enriched the technical methods of graded forecasting of geological disasters.

2.1. Meteorological Warning Indicators for Geological Disasters
The inducing effect of precipitation on geological disasters is not entirely dependent on the rainfall of the day, and is related to the rainfall of the previous period, but the degree of impact of the rainfall of each day in the previous period on this day is not the same. In order to reflect this difference, we have adopted the following calculation formula [8]:

$$P_z = P_0 + \sum_{i=1}^{10} \alpha_i P_i$$  \hspace{1cm} (1)

In formula (1), $P_z$ is the comprehensive rainfall of a certain day; $P_0$ is the rainfall of the day; $P_i$ is the rainfall of the previous $i$ day; $\alpha_i$ is the influence coefficient of the previous $i$ day. Since the influence of rainfall on a certain day on the inducing effect of geological disasters gradually weakens and
disappears with the extension of time, it is an attenuation process, so the attenuation coefficient $\lambda$ can be introduced, and it is assumed that:

$$\alpha_i = \lambda^i$$  \hspace{1cm} (2)

Through the calculation of 224 geological disasters from 1999 to 2015, $\lambda=0.7$ is obtained through the optimization method.

According to the topography, geomorphology, geological structure, and stratigraphic lithology of Guangxi, China, the susceptibility of geological disasters in Guangxi, China is divided into high-prone areas, moderately prone areas, and low-prone areas. Table 1 shows the critical value of comprehensive rainfall in different geological disaster-prone areas.

| The name of the susceptibility zone | Threshold of Induced Rainfall (unit: mm) |
|-----------------------------------|-----------------------------------------|
|                                    | Level 5 | level 4 | Level 3 |
| High-prone area of geological hazards | 120     | 80      | 50      |
| Areas prone to geological hazards   | 130     | 85      | 55      |
| Low-prone area of geological hazards | 150     | 90      | 60      |

2.2. Neural Network Prediction Method Based on Principal Components

The artificial neural network with the characteristics of nonlinear quasi-dynamic system is used for forecasting modeling. In view of the "over-fitting" phenomenon in the prediction modeling of the neural network method, in order to improve the generalization performance, we adopted the prediction modeling method of the principal component analysis to construct the neural network low-dimensional learning matrix, and reduce the dimensionality by condensing the forecast information Denoising, to achieve the purpose of improving the accuracy of forecasting [9].

2.3. Basic Principles and Methods of Forecast Modeling

2.3.1. Principal component analysis method (PCA). Principal component analysis (PCA) is a multivariate statistical analysis method that uses the idea of dimensionality reduction to transform the original multiple indicators into a few independent comprehensive indicators. This method is widely used in the analysis and research of natural sciences [10]. The basic idea is: Suppose there are n predictors for a certain forecast object, $X=(x_1, x_2...x_n)^T$, and n new comprehensive factor variables can be constructed through principal component analysis, $Z=(z_1, z_2...z_n)^T$. Each new comprehensive factor variable is a linear combination of cause subvariables, and the new factor variables are orthogonal to each other, that is, the correlation coefficient between each factor is zero, and the characteristic value of each new comprehensive factor variable clearly indicates its contribution to the cause subgroup.

The new comprehensive factor variables obtained by principal component analysis are used to construct the neural network learning matrix, which can intuitively remove the corresponding principal components with zero eigenvalues, because these principal components hardly contain the information of the original variables; only the eigenvalues are retained. A relatively large principal component with a high degree of correlation with the forecast; in addition, it can be found in actual calculations that generally the first few principal components have a greater variance contribution and have a better correlation with the forecast, because the principal components are orthogonal Therefore, the final determined learning matrix composed of several principal components with large eigenvalues and high correlation with the forecast quantity will not have the influence of redundant repetitive information noise. From the above analysis, it is not difficult to find that the use of principal component analysis to construct the neural network learning matrix can well retain the useful information of all the original factors, and has a good dimensionality reduction effect on the original predictor matrix.
2.3.2. *Neural network method.* Artificial neural network is a data processing system that can automatically extract the nonlinear relationship between a group of forecast variables and another group of independent variables. The establishment of the network is called the training process of artificial neural network, which is solved by recursive iterative adaptive algorithm. The non-linear relationship between the dependent variable and the independent variable. The trained neural network model can be used to estimate or predict predictor variables. So far, there have been many neural network models and corresponding learning methods. Among them, the back-propagation neural network model (BP model for short) is one of the more widely used [11]. This model inserts several hidden layers between the input layer and the output layer, and the neural network elements between adjacent layers. The connection weight coefficients are used to connect each other, and there is no connection between the neurons in each layer. The neural network model used in this article contains only one hidden layer, \( f(x) \) is the network activation function, and the sigmoid function is used:

\[
 f(x) = \frac{1}{1 + e^{-x}} 
\]  
(3)

2.4. *The Establishment of the Forecast Model*

A total of 160 historical examples of geological disasters from 1999 to 2015 were selected as samples, 150 samples were used for forecast model modeling, and 10 samples were used for forecast verification. Through the census of related fields, a total of 16 predictors were obtained, including: the previous 6 hours, the previous 9 days, 24 hours by 24 hours, the mesoscale numerical model MM5, the next 24 hours precipitation, a total of 11 factors, the wind direction and wind speed of 850Hpa in the previous day. There are 5 factors in total, air pressure, dew point and temperature.

According to the principal component analysis method, the above-mentioned factors are subjected to principal component analysis. Table 2 shows the eigenvalues of each principal component calculated by using the preliminarily selected predictor group as the principal component and the correlation coefficient between each principal component and the forecast quantity. The eigenvalue of the principal component is relatively large, and the high correlation coefficient with the forecast is considered as the selection criterion to determine the input of the neural network learning matrix. A three-layer feedforward network calculation method is adopted, the number of hidden nodes is 3, and the learning factor and power factor are respectively 0.9 and 0.7. Use the formula given above to perform repeated learning and training on the standardized learning matrix. When the learning matrix is trained for 5000 times, the error function tends to be stable. The connection weight coefficients and thresholds determined by the network model can be used to obtain the corresponding forecast model, and the determined parameters and network structure can be used as future independent samples and actual forecast models.

| Serial number | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   |
|---------------|-----|-----|-----|-----|-----|-----|-----|-----|
| Eigenvalues   | 5.30| 2.34| 2.04| 1.81| 1.13| 1.04| 0.90| 0.82|
| Correlation coefficient | 0.37 | -0.18 | -0.11 | -0.28 | -0.24 | -0.16 | -0.27 | 0.26 |
| Serial number | 9   | 10  | 11  | 12  | 13  | 14  | 15  | 16  |
| Eigenvalues   | 0.77| 0.74| 0.68| 0.57| 0.24| 0.13| 0.12| 0.08|
| Correlation coefficient | 0.34 | -0.13 | -0.19 | 0.06 | -0.07 | 0.15 | 0.15 | -0.03 |
2.5. Forecast Verification

It can be seen from Figure 1 that the objective prediction method based on principal component analysis and neural network has a better fitting effect on historical samples. The general statistical forecasting method has good fitting effect, but the actual forecasting effect is often not ideal. Therefore, in order to examine the actual forecasting ability of the forecasting method, we conduct forecast verification on independent samples. Judging from the test results in Table 3, for the prediction of geological hazards above level 3, the neural network prediction method based on the principal component has a certain forecasting ability, and has high application value in the actual geological hazard forecasting work.

Table 3. Comparison of forecast results of independent samples of two forecast models

| Sample | Live   | Forecast |
|--------|--------|----------|
| 1      | Level 2| 3.4      |
| 2      | Level 5| 4.1      |
| 3      | Level 3| 5.6      |
| 4      | Level 2| 3.4      |
| 5      | Level 4| 4.4      |
| 6      | Level 1| 4.1      |
| 7      | Level 3| 3.5      |
| 8      | Level 2| 2.4      |

3. Forecast of Geological Hazards Based on MM5 Numerical Model

In order to better carry out the forecast of geological disasters induced by precipitation in Guangxi, the MM5 mesoscale numerical model (15*15km) was introduced, and the modification of model parameters and the pre-processing of data for the Guangxi region were completed. The provided precipitation forecast field is coupled with the geological disaster classification model to form an objective initial field for geological disaster classification forecast.

3.1. The Localization Improvement of the Coupling Model for the Numerical Prediction of Geological Disasters in Guangxi

3.1.1. Handling of model background field and side boundary conditions. The background fields used in the Guangxi mesoscale numerical model include T213 and NCEP. Generally, T213 is the background field used during the daily operation of the model; the NCEP reanalysis data is used as the use of the mesoscale model to affect the heavy rains and tropical cyclones in Guangxi. The background field used in the case of back-calculation.

3.1.2. Optimization of the mode parameter setting scheme. The introduction of mesoscale numerical models to establish Guangxi’s mesoscale numerical forecasting system must optimize the model
according to the specific weather and climate characteristics, topography and landform characteristics of Guangxi, and repeat the calculation range, time step, vertical coordinate, and physical process and calling method. Comparing testing and optimization, reconfiguring the parameter settings of the model to improve the model’s ability to forecast various weather in Guangxi.

3.1.3. This numerical model is a non-static balance model, using two nested grids, semi-implicit semi-Lagrangian integration scheme. The grid distance is 15 kilometers, and the vertical 20 floors.

3.1.4. Substitute the numerical precipitation forecast field into the geological hazard classification model to make the classified geological hazard forecasts by counties in Guangxi. Practice has proved that the use of the nested model of numerical prediction of geological disasters in Guangxi improves the temporal and spatial resolution of the forecast and has a better forecast effect.

3.2. Forecast Example Analysis
Since June 17, 2005, under the combined influence of the warm and humid air currents in the southwest and the weak cold air, an obvious process of heavy rainfall has occurred in most areas of northern Guangxi and central Guangxi. The MM5 mesoscale numerical model accurately predicted the precipitation process on the 17th. We substituted the precipitation forecast conclusions of the MM5 mesoscale numerical model into the Guangxi geological disaster forecast model, and obtained the conclusion from 20:00 to 18 June, 2005. Geological disaster forecast at 20 o'clock (see Figure 2).

According to the information provided by the Guangxi Civil Affairs Department, the geological disaster in Guangxi on June 18, 2005: Diaoshanjiao Village, Gaotian Town, Yangshuo County
There were many large cracks in the mountain. Two houses collapsed in the landslide of Huangtuba Village in Gaotian Town, and the landslide in Dongpingshan Village, Yangshuo, Jinbao Township.
Three houses cracked; multiple landslides and landslides occurred in Jinxiu County; landslides occurred in Shangfang and Pintun, Southeast Village, Longming Town, Tiandeng County. All geological disasters occur within the scope of our geological disaster forecasting. The forecast is satisfactory.

![Figure 2. MM5 mesoscale numerical model grade forecast field on June 18, 2005](image-url)
4. Conclusion

The principal component analysis method constructs the neural network learning matrix, which can condense the useful information of many predictors, play a significant role in dimensionality reduction, and make the network structure smaller. At the same time, due to the orthogonality of the principal components, the repetitive information and noise between the input nodes of the learning matrix can be reduced.

The calculation results of an example show that the main component neural network prediction model fits the historical samples better, and the prediction effect for independent samples is also better. This shows that the main component neural network prediction model significantly improves the forecast generalization ability.

The geological hazard grade forecast method based on the MM5 numerical model has certain forecasting capabilities.

References

[1] Wang Jinlan et al., Thinking about decision-making meteorological services triggered by extreme rainstorms, Policy Research & Exploration, 10(2016) 26-27.
[2] Fu Zongyu et al., Research on the Application of Joint Probability Method in Beijing Disaster Weather Forecast, JOURNAL OF METEOROLOGY AND ENVIRONMENT, 2020, pp. 1-9.
[3] YU Zhi-ming, WANG Si-yao, MA Dong-liang, Study on weather classification and forecasting index of marine meteorological disasters in the Bohai sea, Journal of Meteorology and Environment, 2018, pp. 106−111.
[4] Hou Shumei, etc., the effect of T639 numerical forecast products on the forecast of strong winds along the coast of the Yellow Sea and Bohai Sea, MARINE FORECASTS, 2014, pp. 48-31.
[5] Nong Jifu et al., BP neural network long-term prediction model based on principal component analysis, Journal of Guangxi Teachers Education University(Natural Science Edition), pp. 46-51.
[6] Chen Yuying, Zhao Guangping, Research on the weather forecast and early warning model of geological disasters in flood season in Ningxia, Science of Catastrophes, 2005, pp. 25-28.
[7] GRIORGIO C, GIOGIO G. Coupling Fuzzy Modeling and Neural Networks for River Flood Prediction. IEEE Transactions on Systems, Man, and Cybernetic- Part C: Applications and Reviews, 2005, pp. 382-388.
[8] Shan Jiusheng, Liu Xiufen, Wei Li, Analysis on the characteristics of precipitation that induced landslides in Jiangxi, Meteorology, 2004, pp, 13-15.
[9] SOLLICH P, INTRATOR N, Classification of seismic signals by integrating ensembles of neural networks, IEEE Transactions Signal Processing, 1998, pp. 1194-1021.
[10] Jin Long, Chen Ning, etc., Research and comparison of integrated forecasting methods based on artificial neural network, Acta Meteorological Sciences, 1999, pp. 198-207.
[11] Xiong Lihua, Guo Shengliang, Application of neural network method in real-time flood forecasting, Hydropower Energy Science, 2002, pp. 28-31.