Application of Principal Component Analysis in Meteorological Forecast

Zhonghua Ling¹, Yushan Gao² and Qiang Chen¹, *

¹School of Mathematics and Information Science, Guangxi College of Education, Guangxi, China
²Nanning Number Two High School, Guangxi, China

*Corresponding author e-mail: zhenyan@gxnu.edu.cn

Abstract. Based on the principle of weather dynamics, this paper studies that the pressure gradient, the temperature gradient and the high-altitude guiding wind are the three major types of factors that affect strong winds, and forms three strong wind prediction factor groups. Multiple indexes are transformed into a few linearly independent comprehensive indexes by principal component analysis (PCA). After the factor index is standardized, the comprehensive index of gale forecast is obtained based on PCA. The test results show that the Beibu Gulf gale wind speed has a good corresponding relationship with the gale forecast comprehensive index (the first 24 hours), and the trend basically corresponds to the peak-valley value, which shows that the comprehensive index has a good indication significance for gale forecast. The comprehensive index 30 can be taken as the starting threshold for forecasting strong winds in Beibu Gulf, which has a good reference value for forecasting the presence or absence of strong winds and is a useful discussion for quantifying the forecasting method of weather conceptual model.

Keywords: Principal component analysis (PCA); Forecast method; Winter gale; Comprehensive index of gale forecast.

1. Introduction

Principal component analysis (PCA) is to try to recombine many original indexes with certain correlation into a new set of linearly independent comprehensive indexes to replace the original indexes. PCA is a multivariate statistic method to examine that correlation between multiple variables, and study how to reveal the internal structure of multiple variables through a few principal component. That is, a few principal components are derived from the original variables, so that they retain as much information as possible from the original variables and are linearly independent of each other. Usually, the mathematical treatment is to use the original P indexes as a linear combination as a new comprehensive index.

Li Yufeng (2018) [1] took the air quality of 6 monitoring points as the research object, and analyzed the main influencing factors of the ambient air quality of each monitoring point based on PCA and fuzzy mathematics analysis. The results show that: 1) The air quality is judged to be Grade I by fuzzy
comprehensive evaluation; 2) The first principal component is air suspended particles $PM_{2.5}$ and $PM_{10}$; the second principal component is $O_3$ or $NO_2$; 3) The comprehensive score of ambient air quality of each monitoring point in different months is higher in January, February and December, and the pollution degree is more serious.

Taking some index data of environmental pollution and economic growth from 2003 to 2016 in Beijing, China as an example, Wang Moyu (2018) [2] analyzed economic factors based on PCA, and studies the relationship between economic development and environmental changes. Generally speaking, Beijing's economic growth and environmental pollution are generally "N-shaped", which indicates that when the economy is at a higher level of development, it will in turn inhibit the continuous deterioration of the environment.

The macro-environmental parameters involved in lightning occurrence are many and complicated. Without losing a lot of information, Liu Xiaoyan (2019) [3] compared the characteristics of two lightning weather processes based on the PCA. Using lightning, radar and weather background data, the lightning activity samples from Xining, China, from June to August 2011 to 2013 were selected, and 26 environmental parameters from the same sounding data were extracted for PCA to deeply analyze the evolution and environmental conditions of the two lightning weather processes on July 29th, 2012 (Process I) and August 26th, 2013 (Process II). The results show that the parameters of atmospheric dynamics and temperature and humidity are most closely related to lightning activity in Xining, China. The cumulative probability of lightning intensity shows that the dynamic type is stronger than the temperature and humidity type, the average negative ground flash intensity is higher than -1.91 kA, and the average positive ground flash intensity is higher than 1.92 kA. Process I is characterized by westerly trough lightning weather, which is greatly affected by atmospheric temperature and humidity parameters. However, Process II belongs to subtropical high edge lightning weather, which is obviously affected by atmospheric dynamic parameters. The lightning activity of the two processes is mainly negative ground flash, which has obvious diurnal variation law and is distributed in a unimodal pattern. Process I is dominant in lightning frequency, but obviously weaker than Process II in intensity.

At present, the prediction of strong winds at sea mainly depends on conceptual models and statistical prediction methods. Statistical forecasting methods are widely used in regression, similarity and other forecasting methods [4]. These statistical forecasting methods basically belong to linear sequential processing techniques. However, due to the complexity of the influencing variables of strong wind changes, the stability of deterministic forecasting is poor. The conceptual model prediction method has a clear analysis idea of principle of weather dynamics, but it is more qualitative circulation classification and index groups, which makes it difficult for forecasters to popularize and apply it due to its large randomness in actual use. This paper attempts to use PCA to transform the original multiple indexes into a few independent new index groups. After all kinds of factor indexes are standardized, based on the characteristic values, the factor weight value and the comprehensive index of strong wind forecast are calculated, so as to quantitatively discuss the weather conceptual model forecast method, so as to achieve the purpose of improving the accuracy of strong wind forecast.

2. Prediction factor processing

2.1. Analysis of principle of weather dynamics
When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question. Should authors use tables or figures from other Publications, they must ask the corresponding publishers to grant them the right to publish this material in their paper.

According to principle of weather dynamics, the actual wind in the atmosphere is the sum of geostrophic wind and geostrophic deviation, namely:

$$V = V_g + D$$  \hspace{1cm} (1)

Where $V$ is the actual wind (m/s), $V_g$ is the geostrophic wind (m/s) and $D$ is the geostrophic deviation (m/s).
The analysis shows that $V_g$ is related to the pressure gradient. However, the geostrophic deviation $D$ is mainly related to temperature gradient and high altitude wind speed. Therefore, the pressure gradient, temperature gradient and high-altitude wind speed are used as prediction factors for offshore gale.

### 2.2. Selection of forecast factors

This paper selects a case of strong winds above 8m/s in Beibu Gulf during the winter of 2001-2011, and takes Weizhou Island as the representative station. The total number of samples is 245, of which 225 samples are used as statistical calculation and the other 20 samples are used as independent samples for forecast analysis. According to principle of weather dynamics and many years of practical experience in forecasting, 14 forecasting factors (see Table 1), including air pressure gradient, air temperature gradient, pressure variation, 500hPa wind speed, etc., are selected as principal component variables for forecasting strong winds at sea through screening by correlation analysis methods. These forecast factors have good correlation with the forecast quantity, that is, the strong wind after 24 hours. The absolute value of correlation coefficient is between 0.31 and 0.68, and all reach the correlation significance level of 0.05, which indicates that the selected forecast factors have good correlation with the forecast quantity.

#### Table 1. Selection of Forecast Factors

| No. | 1    | 2    | 3    | 4    | 5    | 6            | 7     |
|-----|------|------|------|------|------|-------------|-------|
| Factor | Chengdu and Haikou | Chengdu and Haikou | Hankou and Haikou | Hankou and Haikou | Changsha and Haikou | Changsha and Haikou | Guilin and Haikou |
|      | $\Delta P_1$ | $\Delta H_1$ | $\Delta P_2$ | $\Delta H_2$ | $\Delta P_3$ | $\Delta H_3$ | $\Delta P_4$ |

| No. | 8    | 9    | 10   | 11   | 12   | 13          | 14   |
|-----|------|------|------|------|------|-------------|------|
| Factor | Guilin and Haikou | Chengdu and Haikou | Hankou and Haikou | Changsha and Haikou | Chengdu | Hankou | Changsha |
|      | $\Delta H_4$ | $\Delta T_1$ | $\Delta T_2$ | $\Delta T_3$ | $V_1$ | $V_2$ | $V_3$ |

($\Delta P$ is that surface pressure gradient difference, $\Delta H$ is 850hPa, $\Delta T$ is 850hPa, $V$ is the northerly component of 500hPa)

### 2.3. Factor standardization treatment

After the factor system is established, because the dimensions of each factor are not unified and there is no comparability between factors, all data must be dimensionless [5-6] to standardize and compare each data. The calculation method is as follows:

1. **Positive effect index**

   The standard values for the i-th sample of the factor are:

   $P_y = \frac{P_y^' - \min(P_y^')}{\max(P_y^') - \min(P_y^')}$  \hspace{1cm} (2)

2. **Negative effect index**

   $P_y = \frac{\max(P_y) - P_y^'}{\max(P_y^') - \min(P_y^')}$  \hspace{1cm} (3)

In Equations (2) and (3), $P_y$ is the normalized factor value, $P_y^'$ is the original factor value, max is the maximum value, and min is the minimum value.
2.4. **Recombination of factor groups by PCA**

Based on the idea of dimension reduction, Principal Component Analysis (PCA) is a multivariate statistical analysis method that transforms the original multiple indexes into a few independent comprehensive indexes [5], which is widely used in the prediction and analysis of atmospheric science. Its basic idea is: set a certain prediction object have n prediction factors, \( X = (x_1, x_2, \cdots, x_n)^T \), and n new comprehensive factor variables, \( Z = (z_1, z_2, \cdots, z_n)^T \) can be constructed by PCA. 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 eigenvalue of each new comprehensive factor variable clearly indicates its contribution to the cause subgroup.

In this paper, EOF method is used to reduce the dimension, and 14 factors with high correlation are further concentrated. This method has the following advantages: (1) EOF expansion obtains mutually orthogonal vectors; (2) The contribution of each vector to the original field can be quantitatively expressed by its eigenvalues; (3) After analysis, the main characteristics of the original field can be expressed by a few principal components with a small number, and because these quantities are orthogonal, there will be no redundant repeated information.

PCA is used to make PCA on the above prediction factors. Table 2 is the calculated eigenvalues of each principal component and the correlation coefficients between each principal component and the forecast quantity, with the primary forecast factor group as the principal component. The new index group is determined by taking the relatively large principal component eigenvalue and considering the good correlation between forecast quantities as the selection standard. As can be seen from Table 2, relatively speaking, the absolute value of correlation coefficient between the first, second, third and fourth principal components and the forecast quantity is the largest, and the corresponding eigenvalues are also relatively large. Therefore, these four principal components are used to construct independent new index groups, which are respectively recorded as \( P_1, P_2, P_3 \) and \( P_4 \).

3. **Calculation of comprehensive index of gale forecast**

3.1. **Determination of weight and construction of comprehensive index for gale forecast**

\[
H_i = 100 \times \sum_{j=1}^{q} W_j P_{ij} \\
W_i = \frac{Q}{\sum_{j=1}^{q} Q_j}
\]

(4) (5)

Where \( H_i \) is the comprehensive index of gale forecast; \( W_i \) is the weight of factor \( P_i \); \( Q \) is the eigenvalue.

**Table 2.** Eigenvalues of Principal Components and Correlation Coefficients between Principal Components and Forecast Quantities

| No. | 1     | 2     | 3     | 4     | 5     | 6     | 7     |
|-----|-------|-------|-------|-------|-------|-------|-------|
| Eigenvalue | 6.73  | 2.95  | 1.39  | 1.06  | 0.78  | 0.65  | 0.54  |
| Correlation coefficient | 0.58  | -0.47 | -0.19 | 0.22  | 0.01  | 0.15  | 0.01  |
| No. | 8     | 9     | 10    | 11    | 12    | 13    | 14    |
| Eigenvalue | 0.44  | 0.11  | 0.06  | 0.01  | 0.00  | 0.00  | 0.00  |
| Correlation coefficient | 0.16  | 0.03  | -0.04 | 0.21  | -0.21 | -0.12 | 0.03  |
Table 3. Weights of New Index Groups

|       | P1 | P2 | P3 | P4 |
|-------|----|----|----|----|
| Weight| 0.45 | 0.15 | 0.12 | 0.08 |

According to Equation (4), 225 samples are used to calculate the weight of each new index group (see Table 3). From Table 2 and the standardized values of each sample, the gale pre-index comprehensive index H of 20 samples that did not participate in statistical analysis is calculated according to Equation (5) (see Figure 1).

Figure 1. Analysis of the wind speed and the comprehensive index of gale forecast (the first 24h).

4. Analysis of Calculation Results

As can be seen from Figure 1, there is a good correspondence between the gale wind speed in Beibu Gulf, Guangxi, China and the gale prediction comprehensive index (the first 24 hours), and the trend basically corresponds to the peak-valley value, which shows that the comprehensive index has a good indication significance for gale prediction. There are strong winds above 8m/s in Beibu Gulf, and most (70%) have a comprehensive index value of more than 30 in the first 24 hours. The comprehensive index value of 30 can be taken as the starting threshold for forecasting strong winds in Beibu Gulf.

In addition, 20 samples of Beibu Gulf wind were randomly selected for comparative analysis with the comprehensive index of gale forecast (the first 24 hours) (see Table 4). The results show that there is still a good correspondence between wind speed and the comprehensive index of gale forecast (the first 24 hours). However, if the comprehensive index value of 30 is taken as the starting threshold for forecasting strong winds in Beibu Gulf, the number of empty reports is more, 4 times, while the number of missing reports is only one, which indicates that the index can be used as the starting threshold for forecasting strong winds. If the accuracy rate is improved, it is also necessary to eliminate empty reports in combination with circulation classification.

Table 4. Comparative Analysis of Comprehensive Index of Wind Speed and Gale Forecast (First 24h)

| Times | Wind number | Index | Wind number | Index |
|-------|-------------|-------|-------------|-------|
|       | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|       | 4.5 | 5.3 | 9.7 | 10.5 | 7.7 | 12.3 | 4.4 | 11.6 | 10.8 | 10.0 |
| Index | 25.2 | 32.3 | 30.0 | 35.1 | 35.4 | 42.6 | 24.6 | 40.4 | 38.2 | 27.5 |
| Times | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Wind number | 5.2 | 8.6 | 10.5 | 6.3 | 4.8 | 17.7 | 7.2 | 12.3 | 9.6 | 11.6 |
| Index | 34.3 | 38.3 | 32.5 | 35.4 | 22.3 | 45.2 | 25.3 | 44.3 | 37.5 | 32.5 |
5. Conclusion

(1) Firstly, according to the principle of weather dynamics, three types of factors affecting the gale in Beibu Gulf are obtained, namely, air pressure gradient, air temperature gradient and high-altitude guiding wind. According to these three types, strong wind prediction factor groups are formed as the main components of strong wind prediction at sea. The original multi-index is transformed into a few independent comprehensive indexes by PCA.

(2) After standardization of various factor indexes, the comprehensive index of gale forecast is obtained based on PCA;

(3) The test results show that there is a good correspondence between the wind speed of Beibu Gulf and the comprehensive index of gale prediction (the first 24 hours), and the trend basically corresponds to the peak and valley values, which shows that the comprehensive index has a good index significance for gale prediction. The comprehensive index 30 can be taken as the starting threshold for forecasting strong winds in Beibu Gulf, which has a good reference value for forecasting the presence or absence of strong winds and is a useful discussion for quantifying the forecasting method of weather conceptual model:

(4) The index can be used as the threshold for reporting the presence or absence of strong winds, but there are many empty reports. If the accuracy is to be improved, it is necessary to eliminate empty reports in combination with circulation classification.

Acknowledgments

This work was financially supported by the Basic Ability Improvement Projects for Young and Middle-aged Teachers from Colleges of Guangxi in 2019 (NO.2019KY1679, 2019KY1681), the Subproject of Bagui Scholars Innovation Team of Guangxi College of Education and the Guangxi College of Education (No. A2017001).

References

[1] Yu-feng Li, Fuzhou City Air Quality Assessment Based on the Principal Component Analysis and Fuzzy Comprehensive Evaluation, J. Journal of Wuyi University. Vol.37, No.9 (2018) 39-44.

[2] Moyu Wang, Research on the Relationship between Beijing’s Economic Development and Environmental Pollution Based on Principal Component Analysis, J. Electronic Technology & Software Engineering. 23 (2018) 182-183.

[3] Xiao-yan Liu, Comparation on Characteristics of Two Lightning Processes Based on the Principal Component Analysis, J. Meteorological and Environmental Sciences. Vol.42, No.4 (2019) 119-126.

[4] Rong Wang, Analysis and forecasting of wind field character on the north of South China Sea borders near-ocean area, J. Marine Forecasts. Vol.28, No.2 (2011) 1-8.

[5] Ji-fu Nong, The Rainfall Forecast Model of PCA-RBF Neural Networks Based On MATLAB, J. Journal of Tropical Meteorology. Vol.24, No.6 (2008) 713-717.