Long-term intelligent calculation and prediction model for heavy precipitation satellite cloud Images

Xiaoyan Huang1*, Li He1, Huasheng Zhao1, Yushuang Wu1 and Ying Huang1

1 Guangxi Research Institute of Meteorological Disasters Mitigation, Nanning, Guangxi, 530022, China

*Corresponding author’s e-mail: gx_huangxy@163.com

Abstract. A nonlinear roiling prediction model for satellite image has been developed based on Shapley neural network using the ensemble prediction method similar to the numerical prediction model, due to lacking of the guidance of a nonlinear prediction theory for satellite image at present. Empirical Orthogonal Function (EOF) method is applied to the samples of infrared satellite image every 6 h in heavy rainfall processes, and time coefficients extracted are used as predictands. Since the changes of precipitation cloud system are governed by the physical quantity fields in cloud cluster, the physical quantities prediction products from numerical prediction model are used as predictors, and Shapley Neural Network Ensemble Prediction models are established for the corresponding time coefficients based on the technique of the reduction of data dimensionality for data interpretation. By integrating the predicted time coefficient and space vector, the future satellite image is obtained. Results show that the nonlinear prediction model can better forecast the main features of the development of heavy rainfall cloud cluster in future 24h.

1. Introduction

Satellite cloud image has been widely used in the meteorological field because of its high temporal and spatial resolution and wide coverage. It has become an important non-conventional meteorological data except for conventional data, including: surface meteorological observation data, air wind data, air temperature, pressure and humidity data. Although the research work on cloud image prediction has been carried out for a long time at home and abroad, it has made more research progress, but mainly based on the premise that the system remains stable, some linear methods are used to map the local feature matching based on the cloud image and the motion vector before and after the cloud image. The linear extrapolation of the relationship has certain deficiencies in the forecasting and forecasting of the cloud's development trend, which greatly limits the availability of these short-term cloud image predictions[1-4]. Moreover, so far, most of the prediction studies on the future development of cloud images are based on some short-term (1-3 hours) cloud image prediction studies. For the future cloud image changes that have exceeded the longer time interval of 3 hours, the current cloud image prediction theory and methods and prediction effects are not perfect, and there are few practical prediction models[5-9].

Since the formation of the cloud is a very complicated process, the change of the cloud image over time is affected by the combination of the internal and external atmospheric environment of the cloud image. Therefore, there is a nonlinear relationship between the cloud image and the atmospheric element fields, and the cloud image is always in constant displacement. In the process of deformation, expansion and contraction, and even splitting and fusion, the whole tracking process is much more
difficult than the tracking problem of general linear or near-linear objects in computer vision. Therefore, the traditional linear prediction method has certain limitations. In particular, changes in the future cloud image conditions beyond the longer time interval of 3 hours, relying only on linear extrapolation or physical processes and dynamic factors that do not involve changes in cloud image movement, may make it difficult to obtain good prediction results. Therefore, according to the whole life history process of the development of the cloud group until its extinction, it will be obviously affected by the physical quantity factors of the atmospheric environment field. And the irregular changes in the cloud state are more non-stationary and nonlinear. This study establishes a nonlinear statistical relationship model between the evolution process of the cloud and the influence factors of the environmental field. Predictive research on the future development of satellite cloud images aims to explore long-term forecasting methods for satellite cloud images.

2. Cloud design prediction and principle method

2.1 Design principle of cloud prediction
From the cloud observation data, we can clearly see that a large number of cloud movements, changes in production and consumption are very fast, especially some strong convective cloud systems, which may change significantly in a short time interval. The rapid change of this cloud system over time is closely related to the dynamics, heat and water vapor conditions within the cloud system, and is also inseparable from the atmospheric conditions around the cloud system. Therefore, in order to effectively predict the cloud system status of the cloud system at relatively long time intervals (such as 24 hours), it is necessary to consider the above-mentioned internal and external atmospheric physical factor conditions in the prediction model. On the one hand, it is necessary to reasonably describe the characteristics of the cloud image, and on the other hand, how to link these cloud system features with the future atmospheric conditions. To this end, this study proposes to make full use of numerical forecast product data to establish a cloud cluster prediction model with a long time interval. The calculation formula can be written as:

\[ \hat{S} = f(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) \]  

Among them, \( \hat{S} \) is the future cloud system status, \( \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n \) is the atmospheric physical quantity factor at the future moment closely related to the future cloud system status, and \( f \) is the nonlinear mapping relationship between the cloud system and the physical quantity factor.

2.2 Nonlinear ensemble prediction method for cloud group prediction
Considering that the movement or evolution process of the cloud system is mainly affected by the comprehensive environmental conditions of the cloud and the surrounding atmosphere, the change characteristics are mainly non-stationary nonlinear changes. To this end, the predictive model (1) needs to adopt a more reasonable nonlinear statistical forecasting modeling method.

At present, in the atmospheric discipline, the statistical weather forecasting method is based on the correlation between forecasting factors and forecasting quantities to establish forecasting equations for independent variables and dependent variables. However, these statistical forecasting methods, whether linear or nonlinear, have a certain degree of uncertainty in their single forecasting equations. In order to improve this uncertainty, the integrated forecasting method is often used in meteorology. It is mainly for the same forecasting object, using different forecasting methods to establish several forecasting equations, and giving different (or equivalent) weights of each equation for integration to arrive at the final forecasting conclusion. In order to improve the long-term forecasting difficulty of cloud clusters, this research project uses a combination of Shapley value and neural network to design a new nonlinear statistical ensemble forecasting method to improve the forecasting skills and forecast stability of a single statistical forecasting equations.
The Shapley value method of cooperative countermeasures is used to determine the weight coefficient of the neural network integrated forecasting individual, and the integrated individual difference degree is enhanced [10-12]. The Shapley value provides a good way to determine the distribution ratio of each participant in a n-person bargaining game. In order to improve the generalization performance of the neural network integrated forecasting model, the Shapley value method is adopted, and the sum of the squares of the combined forecasting effectiveness is used as the cooperation criterion. The Shapley value of the cooperative strategy is distributed among the integrated forecasting individuals to determine the individual forecasting individuals. In the integrated forecasting weight coefficient, the difference of the integrated forecasting individual is improved, and a new machine learning objective integrated forecasting model affecting the satellite cloud image change is established. Specifically, \( \Phi(\upsilon)=(\varphi_1(\upsilon), \varphi_2(\upsilon), \ldots, \varphi_m(\upsilon)) \) can be set as the Shapley value of the cooperative \( m \) neural network individual countermeasure \( \Gamma=[M, \upsilon] \), which can be proved

\[
\sum_{j=1}^{n} \varphi_j(\upsilon) = v(M) \quad (2)
\]

In the above formula, the sum \( \varphi_j(\upsilon) \) of the average contributions of the individual neural network prediction individuals in the combined prediction method is equal to the total result of the cooperation, and the weighting coefficient of each individual neural network prediction individual in the combined forecasting method should be determined according to the average contribution in cooperation. Considering the effectiveness \( v(M) \) of the combined prediction, the normalization process \( \varphi_j(\upsilon) \) will be performed as follows, and the weight coefficient of each individual neural network prediction individual can be obtained.

\[
l_i = \varphi_i(\upsilon) / \sum_{j=1}^{m} \varphi_j(\upsilon) \quad (3)
\]

Obviously, \( l_i \geq 0, \sum_{j=1}^{m} l_i = 1, i = 1, 2, \ldots, m \). Equation (3) is the calculation formula of the combined prediction weight coefficient obtained. Using this formula to calculate the weight coefficients of the integrated forecasting individuals in the neural network integrated forecasting model of satellite cloud image changes, and then rationally allocate them, which can improve the degree of difference between individuals, so that the individual forecasting models can be fully considered to improve the forecast. The contribution of efficiency, thus a superior combination forecast.

### 3. Cloud image ensemble prediction experiment

#### 3.1 Forecast modeling data and processing methods

In this study, when establishing the cloud cluster nonlinear ensemble prediction model, it is based on the cloud top brightness temperature data of FY-2E in 2013-2016, and the projection method adopts equal latitude and longitude projection. Taking into account the relationship between the generation and development of the cloud system and the precipitation process, the cloud image of the precipitation process with a daily average precipitation of more than 25 mm in the South China Regional Meteorological Observatory was selected as a model for forecasting modeling. The spatial extent of the cloud map is taken from South China: 15°~30°N, 100°~120°E, the cloud image pixel is 401×301, and the resolution is 0.05°×0.05°. According to the selection criteria of the above cloud data, the precipitation process was selected from 2013 to 2015, and a total of 196 cloud data were used as prediction modeling samples every 6 hours. With the same standard, in 2016, a strong precipitation process was selected, and a total of 86 cloud images were used as forecast test samples for the forecast model.

First, the model of the model cloud image is decomposed using the Natural Orthogonal Expansion (EOF) method:

\[
X = VZ
\]

(4)
That is, the cloud image at any one time can be expressed as:

$$X(x, y, t) = \sum_{i=1}^{n} v(x, y) \cdot \xi_i(t)$$  \hspace{1cm} (5)

Where $v(x, y)$ is the principal component of the EOF decomposition, $\xi_i(t)$ is the expansion time coefficient corresponding to each principal component, and $i$ is the order of the natural orthogonal expansion. Table 1 shows the variance and cumulative variance contribution of the first 15 principal component expansions after the cloud model modeling sample is developed for EOF. It can be seen from Table 1 that the cumulative contribution of the first 15 principal components is nearly 80.97%, that is, the first 15 principal components can better reflect the main features of the cloud image. For this reason, the time series coefficient $T_1, T_2, T_3, \ldots, T_{15}$ corresponding to the first 15 principal components is used as the 15 forecast components to establish a nonlinear ensemble prediction model. Then, according to formula (5), the predicted time coefficient is combined with the corresponding space vector to obtain a forecast cloud image at a future time.

Table 1. Variance and cumulative variance contribution of the first 15 principal component expansions.

| item              | PC1  | PC 2 | PC 3 | PC 4 | PC 5 | PC 6 | PC 7 | PC 8 |
|-------------------|------|------|------|------|------|------|------|------|
| Variance contribution | 32.68 | 13.04 | 8.13 | 5.10 | 3.93 | 3.04 | 2.89 | 2.67 |
| Cumulative variance contribution | 32.68 | 45.72 | 53.85 | 58.95 | 62.88 | 65.92 | 68.81 | 71.48 |
| item              | PC9  | PC10 | PC11 | PC12 | PC13 | PC14 | PC15 |
| Variance contribution | 2.00  | 1.65 | 1.45 | 1.29 | 1.21 | 1.02 | 0.87 |
| Cumulative variance contribution | 73.48 | 75.12 | 76.58 | 77.86 | 79.08 | 80.09 | 80.97 |

3.2 Analysis of the results of cloud image prediction

In establishing the above-mentioned ensemble prediction model of 15 time coefficient forecast components, firstly, according to the global reanalysis data of ERA Interim of the European Centre for Medium-Range Weather Forecasting (ECWMF), the height field, wind field, vertical velocity field and relative interval are 6 hours apart. Correlation factors of 24 physical quantity forecast fields such as humidity, water vapor flux and K index are correlated with the significance level of 0.05 and the correlation coefficient $\geq 0.3$, which is the input of the nonlinear ensemble prediction model of each forecast component. Using the ensemble prediction individual generation method of Section 3 Shapley-Neural Network ensemble prediction modeling, an ensemble prediction model of 15 time coefficient forecast components is established respectively. Using 15 time coefficient sets to predict the time coefficient of the next moment obtained by the model prediction, and then synthesizing with the corresponding space vector, we can get the prediction result of each cloud image of the forecast sample. Similarly, it is also possible to make corresponding forecast results for a total of 86 cloud images for the 2016 precipitation process. Further statistical calculation of the correlation coefficient between the predicted cloud image and the cloud top brightness temperature of the live cloud image (Figure 1), it can be seen that the correlation coefficient between the 86 predicted cloud images and the live cloud images exceeds 50%, accounting for 45.35% of the total. The highest correlation coefficient is 0.85, and these samples are better predicted. The correlation is 60-70%, accounting for 18.6% of the total. However, the prediction effect of the sample is relatively poor, that is, the correlation is less than 20%, which only accounts for 12.9% of the total. Our actual observation time in the cloud image has reached 24 hours, and such forecasting time is of great significance for the actual heavy precipitation forecast.
4. Conclusions
In order to make better use of a large number of satellite cloud image observation data to improve the forecasting ability of heavy rainfall and heavy precipitation, the most important solution is to improve the prediction accuracy of the cloud system changes of heavy precipitation and increase the forecasting time for the changes of future cloud systems. In order to explore a new cloud image prediction method that can be used for the prediction of heavy precipitation, the project is based on the characteristics of the satellite image of strong precipitation, the nonlinearity, time-varying and non-stationary characteristics of the cloud. Starting from the atmospheric physical quantity forecasting factor with significant influence on the change, the cloud image change trend ensemble forecast for the next 24 hours of forecasting aging is carried out. The important feature of the project's modeling method is that in the cloud image forecasting modeling process, the system dimension reduction method is used to extract the features of the strong precipitation cloud map, and the time coefficient is used as the forecast component of the forecast modeling to establish the main extracted from the cloud map. The nonlinear mapping relationship between eigenvectors and numerical prediction model products makes each forecast component prediction model have a good physical basis. On the mathematical forecasting modeling method, a Shapley-Neural network ensemble prediction method similar to the numerical forecasting model ensemble prediction method is designed. This method is different from the traditional statistical forecasting method in the atmospheric discipline. It is a nonlinear statistical ensemble forecasting modeling method based on intelligent computing method. From the actual heavy precipitation satellite cloud image prediction test results of this study, it shows that the method has good stability and universal applicability when carrying out actual forecasting of each forecast component, which provides an objective and practical new method for carrying out actual cloud image prediction.

Acknowledgments
This study was supported by the National Natural Science Foundation of China (Grant No. 41765002,41575051) and the Guangxi Key National Natural Science Foundation (Grant No. 2017GXNSFDA198030).

References
[1] Wang Dengyan. Morphology Feature and Extrapolation Forecast of MCS[J]. Meteorological Monthly. 2000,26(8): 22-24.
[2] Zhang Ren, Wang Hai, Sun Zhaoxi, et al. Fuzzy Inference Cloud Classification of Bispectral Satellite Cloud Image[J]. Journal of Disaster Prevention and Mitigation Engineering. 2004,24(3): 257-263.
[3] Zhang Ren, Liu Kefeng, Sun Zhaozhen, et al. Discrimination and Proximity Prediction of Cloud Movement in Satellite Cloud Image[J]. Journal of Basic Science and Engineering, 2004, 12(Supplement): 141-145.

[4] Wang Wei, Liu Juan, Meng Zhibin. Dynamic Tracing Prediction of Convective Clouds Based on Time Series Remote Sensing Satellite Cloud Image[J]. Chinese Journal of Electronics, 2014, 42(4): 804-808.

[5] Thomasm Hamill, Thomasm Nehrkomj. A short-term cloud forecast scheme using cross correlation [J]. Weather and forecasting, 1993, 8(4): 401-411.

[6] Bai Jie, Wang Hongqing, Tao Zuyu. Identification and Tracking of Strong Convective Clouds in GMS Infrared Satellite Cloud Image[J]. Journal of Tropical Meteorology, 1997, 13(2): 158-167.

[7] Wang Lei, Huang Peiqiang, Shi Hanqing. Using Satellite Data to Make Weather Forecasting of Weather System[J]. Journal of Meteorology, 1999, 19(3): 270-275.

[8] Huang Yong, Kong Qingxin, Zheng Lanzhi. Convective Cloud Prediction Based on Maximum Cross Correlation[J]. Meteorological Science, 2005, 25(4): 399-404.

[9] Gong Ke, Ye Dalu, Ge Chenghui. Motion Vector Method for Satellite Cloud Image Prediction[J]. Journal of Image and Graphics, 2000, 5(4): 349-352.

[10] Shapley Lloyd S. A value for n-person games [A]. In Contributions to the Theory of Games, Vol. II (by H.W. Kuhn and A.W. Tucker, editors). USA : Princeton University Press, 1953, 307-317.

[11] Chen Jing, Dai Xiaoping, Chen Xiang, et al. Calculation method of agricultural water saving compensation based on improved Shapley value method[J]. Journal of Hydraulic Engineering, 2011, 42(6): 7-14.

[12] Chen Qiming, Chen Huayou. Shapley value method for determining the weight coefficient of a class of combined forecasting models[J]. Journal of Anhui University (Natural Science), 2012, 36(2): 29-34.