Application Research of ARIMA Model in Rainfall Prediction in Central Henan Province

Lulu Xu [1], Dexian Zhang [1], Xin Zhang [1]

[1]School of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China.

Abstract

The traditional rainfall prediction model is mainly based on grey theory and neural network prediction model, with simple method and low accuracy. In order to improve the prediction accuracy of rainfall, an ARIMA rainfall prediction model based on Cramer decomposition is proposed. Taking the rainfall data from 1996 to 2016 in central Henan Province as an example of application research. Among them, the data from 1996 to 2011 were used as the modeling samples, the data from 2012 to 2016 were used as the test samples. The results show that the accuracy of the prediction based on the improved ARIMA model is higher, the prediction accuracy is more than 95%,it provides a new method for rainfall forecast, and provides disaster prevention basis for relevant departments such as agriculture and water conservancy.

Keywords

Rainfall; prediction; ARIMA model; high precision.

1. Introduction

Rainfall is not only an important index to measure the degree of urban drought, but also an important parameter for disaster prediction analysis such as floods. Accurate prediction of rainfall can effectively help agriculture, water conservancy and other departments to prevent potential disasters, so as to achieve the purpose of reducing the degree of disaster. Therefore, it is very important to predict rainfall and improve the accuracy of rainfall prediction.

In recent years, domestic scholars have conducted extensive research on rainfall prediction, The rainfall prediction model based on grey theory , neural network] and, Markova model and so on are put forward. The specific model formulas and model parameters are given in most methods, the rainfall is assumed to be a stationary time series model, however, in practice, rainfall patterns tend to be non-stationary and no linear relation can be found. Many factors affecting the change of rainfall. There is no model for predicting rainfall accurately at present. An improved ARIMA rainfall prediction model proposed in this paper. Through the analysis of rainfall data for the last 15 years and the Cramer decomposition theorem, the non-stationary rainfall data is converted into a stable data sequence. Using the stationary data sequence to establish the model and select the optimal model parameters, so as to achieve the aim of predicting the next annual rainfall.

2. Model Principle

2.1 ARIMA Model

ARIMA model is called auto-regressive integral sliding average model by Box and Jenkins in the 1970s,It's also known as the box-Jenkins model, the box-Jenkins method.

Definition 1 Set sequence

$$\{X_t\} \text{ fit } X_t - a_1X_{t-1} - \cdots - a_pX_{t-p} = \varepsilon_t - b_1\varepsilon_{t-1} - \cdots - b_q\varepsilon_{t-q} \tag{1}$$

Among them, a_i, b_j is a constant, $\{\varepsilon_t\}$ is the white noise sequence, $\forall s < t$, $EX_s\varepsilon_t = 0$,this model is called auto-regressive sliding average model, that is ARMA (p, q) model.

Definition 2 Set time series

$$\{X_t\} \text{ fit } 1 - B^dX_t = W_t \tag{2}$$

Among them, W_t is a stationary and reversible ARMA (p, q) sequence, This model is called ARIMA (p, d, q) model [18].

Among them, definition 2 is the traditional ARIMA model. It combines the advantages of regression analysis and time series analysis, and it is a widely used prediction model. But ARIMA model is a variant of time series, so it is still based on the prediction of linear data in essence, but in reality, rainfall is a non-stationary random process with complexity and dynamics. Therefore, if the ARIMA model is completely used, the purpose of accurate rainfall prediction can not be achieved.

The analysis shows that, the variation trend of rainfall is non-linear, belonging to non-stationary time series, therefore, if the rainfall data can be preprocessed, transforming the non-stationary sequence into a stationary sequence, the prediction accuracy of the model can be greatly improved.

2.2 Improved ARIMA Model

It is known from the Cramer decomposition theorem, variances homogeneous non-stationary sequences can be decomposed into the following forms:

$$X_t = u_t + \varepsilon_t = \sum_{j=0}^{d}\beta_j t^j + \Theta(B)\varepsilon_t \tag{3}$$

Among them,$\{\varepsilon_t\}$ is zero-mean white noise sequence.The d-order difference of the white noise sequence is equivalent to the d-order derivative of the continuous variable, Therefore, the accuracy information contained in $\{X_t\}$ can be completely extracted by d-order difference.For example, polynomial $\sum_{j=0}^{d}\beta_j t^j$ can be transformed into a constant by d-order difference. The process is as follows:

$$\nabla^d\sum_{j=0}^{d}\beta_j t^j = c \tag{4}$$

The c is a constant. Expand first-order difference, there is:

$$\nabla X_t = X_t - X_{t-1} \tag{5}$$

Equivalent to,

$$X_t = \nabla X_t + X_{t-1} \tag{6}$$

From (4), (5), (6) we can see that the first order difference is essentially an auto-regressive process. It mainly uses $\{X_{t-1}\}$ as an independent variable to explain the change of $\{X_t\}$.Differential operations essentially use auto-regressive methods to extract the accuracy information.

Based on the above analysis, The structure of ARIMA prediction model based on Cramer decomposition theorem is :

$$\begin{cases} \varphi(B)\nabla^dX_t = \Theta(B)\varepsilon_t, \\ E(\varepsilon_t) = 0, \ \ Var(\varepsilon_t) = \sigma_\varepsilon^2, \\ E(\varepsilon_t\varepsilon_s) = 0, \ \ s \neq t, \\ E(x_s\varepsilon_t) = 0, \ \ \forall s < t. \end{cases} \tag{7}$$

Compared to the d-order homogeneous non-stationary sequence $\{X_t\}$,$\{\nabla^dX_t\}$is stationary sequences, therefore, it is suitable for the traditional ARIMA model prediction, that is:

$$\varphi(B)\nabla^dX_t = \Theta(B)\varepsilon_t \tag{8}$$

Among them

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p \quad (9)$$

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q \quad (10)$$

Among them, p represents the auto-regressive order, d represents the difference order, q represents the moving average order. Improved ARIMA model modeling process shown in Figure 1.

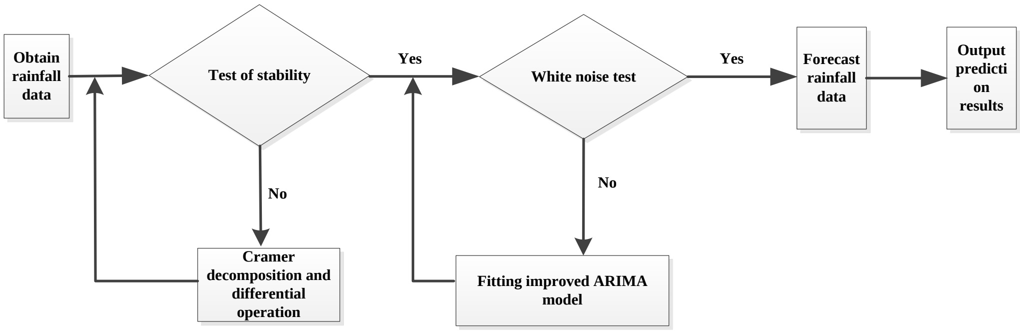

Figure 1. Improved model modeling flow chart

Modeling steps:

Step 1: determine the stability of the sequence.

Step 2: if the sequence is non-stationary, stabilize the original sequence, that is, according to Cramer decomposition theorem for processing and differential computing;

Step 3: white noise test is carried out for the stationary sequence after second steps, if the test is not satisfied, the stationary non-white noise sequence is fitted, and get the model parameters;

Step 4: through step 3 to obtain the model parameters to establish an improved prediction model, and use this model to predict the rainfall in the next year;

Step 5:output rainfall forecast result.

3. Analyses

In this paper, precipitation data from 1996 to 2010 in central Henan Province were used as modeling samples; data from 2011 to 2016 were used as test model data.

According to Definition 1and Definition 2 of the relevant theory in 2.1,using EViews 8 software to build prediction model, performing a first-order difference operation on the sequence to get the result shown in Figure 2.

Null Hypothesis: D(DY1) has a unit root
Exogenous: Constant
Lag Length: 3 (Automatic - based on SIC, maxlag=3)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.422278	0.0006
Test critical values:	1% level	-4.297073	
	5% level	-3.212696	
	10% level	-2.747676	

*MacKinnon (1996) one-sided p-values.
Warning: Probabilities and critical values calculated for 20 observations and may not be accurate for a sample size of 10

Figure 2. First-order difference result graph

As shown in Figure 2, after the first-order difference of the original sequence, the Prob data is far less than 0.05, so the sequence after the first-order difference is stable. From the first-order difference results, the model range can be assumed, and the possible cases are fitted respectively, the results are shown in Figure. 3.

Date: 06/28/18 Time: 16:06
Sample (adjusted): 2000 2011
Included observations: 12 after adjustments
Convergence achieved after 361 iterations
MA Backcast: OFF (Roots of MA process too large)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.011064	0.360904	-0.030655	0.9805
AR(1)	-0.346394	10.57610	-0.032753	0.9792
AR(2)	0.008176	10.41461	0.000785	0.9995
AR(3)	-0.320958	17.58874	-0.018248	0.9884
MA(1)	-2.340945	77.67541	-0.030138	0.9808
MA(2)	-3.877657	86.02889	-0.045074	0.9713
MA(3)	6.715670	50.16968	0.133859	0.9153
MA(4)	1.227373	158.4141	0.007748	0.9951
MA(5)	-1.956140	33.14994	-0.059009	0.9625
MA(6)	0.455494	142.8849	0.003188	0.9980
MA(7)	-0.901476	40.98730	-0.021994	0.9860

R-squared	0.990337	Mean dependent var	-0.011272
Adjusted R-squared	0.893710	S.D. dependent var	0.266653
S.E. of regression	0.086934	Akaike info criterion	-2.698900
Sum squared resid	0.007558	Schwarz criterion	-2.254402
Log likelihood	27.19340	Hannan-Quinn criter.	-2.863469
F-statistic	10.24909	Durbin-Watson stat	1.787798
Prob(F-statistic)	0.238818		

Inverted AR Roots	.24+.58i	.24-.58i	-.83	
Inverted MA Roots	2.85	1.17	.83	-.01+.49i
	-.01-.49i	-.79	-1.70	

Estimated MA process is noninvertible

Figure 3. Model fit results

It can be seen from the results of Figure. 3 that AR(2) and MA(6) are closest to 0 in the t-test result, and the original sequence of the model undergoes first-order difference. Therefore, it can be concluded that the final model type is ARIMA(2,1,6) model. And from the results of Figure 3, the least squares estimation result of the model is:

$$X_t = -0.011064 + 0.008176X_{t-2} + 0.455494X_{t-6} \tag{11}$$

Among them, t=1996, 1997, 1998...the estimated variance of the error term is: $\sigma = 0.086934$

The EViews 8 software is used to build the predictive model, and the model is trained using the modeling samples, and the test sample is used to predict the model. The rainfall prediction results and observations are shown in Figure 4.In order to illustrate the validity of the model, the prediction results of the rainfall prediction model based on artificial neural network are introduced for comparison. The results are shown in Figure 4 and Table 1.

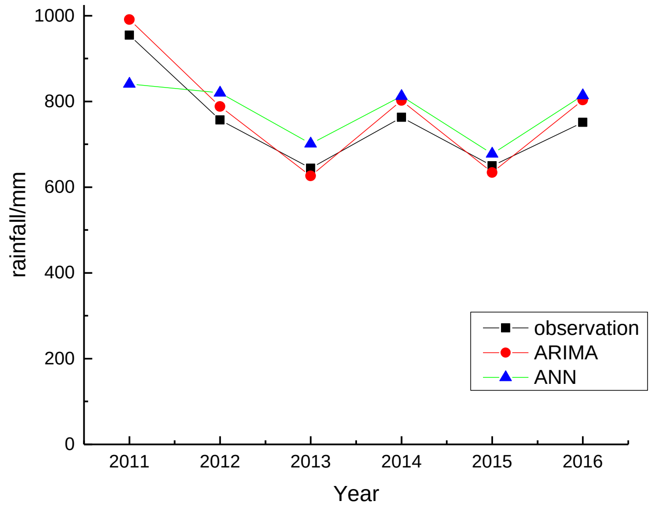

Figure 4. Comparison of prediction results

Table 1. Precision comparison of model prediction results

Model name	Percentage of predicted mean error
Improved ARIMA（2,1,6）	5.046%
Neural Network	8.018%

As can be seen from Figure 4 and Table 1,The accuracy of the improved ARIMA rainfall prediction model is higher than that of the artificial neural network rainfall prediction model. The average error of the artificial neural network prediction model is 8.018%, while the improved ARIMA prediction model is only 5.046%, and the prediction accuracy is close to 95%.Moreover, the improved ARIMA prediction model predicts that rainfall trends are the same as actual observations. It can be seen that the improved ARIMA rainfall prediction model, through the Cramer decomposition theorem, first transforms the non-stationary sequence into a stationary sequence, which makes the model more reasonable to extract the regular information in the data, and the final modeling data sequence is obtained by differential operation, and the accuracy of the prediction is higher according to this sequence.

4. Conclusions

In this paper, rainfall prediction as the main body of the study, a rainfall prediction model based on improved ARIMA is constructed, taking the rainfall data from 1996 to 2010 in central Henan Province as modeling sample data and rainfall data from 2011 to 2016 as model verification data. The results show that the prediction model based on improved ARIMA has high prediction accuracy and accurate trend change. This model provides a new method for rainfall prediction, and provides new technical support for the prevention and control of water conservancy and agriculture departments.

References

[1] Ren ye, Xu Shuqin. Application of grey neural network combination model in rainfall prediction in Qing'an County [J]. water saving irrigation, 2012, (9): 24-25,29.

[2] Hai Qing, based on the annual rainfall prediction model of grey system, explores the development and management of [J]. water resources, 2017, (3): 61-63.

[3] Cui Lei, Chi Dao Cai, martial arts. Grey prediction method for predicting rainfall by wavelet denoising [J]. Journal of Liaoning Technical University (NATURAL SCIENCE), 2009,29 (5): 853-856.

[4] Ge Cailian, Cai Huanjie, Wang Jian, Zhang Lulu. Rainfall prediction research based on BP neural network [J]. water saving irrigation, 2010, (11): 7-10.

[5] Zhang Shuai, Wei Zhengying, Zhang Yubin. Application of recurrent neural network in rainfall prediction [J]. water saving irrigation, 2017, (5): 63-66,71.

[6] Feng Yi, Wu Boqiang, Cui Ling Zhou. Prediction of Typhoon Rainfall Based on BP neural network [J]. soil and water conservation research, 2012,19 (3): 289-293.

[7] Ren Gang Hong, Du Kun, Zhou Ming, et al. Annual rainfall prediction based on cascaded neural network [J]. civil engineering and environmental engineering, 2016, S2 (38): 137-141.

[8] Wang Tao, Qian Hui, Li Peiyue. Application of weighted Markov chain to rainfall prediction in Yinchuan area [J]. south to North Water Diversion and water conservancy science and technology, 2010,8 (1): 78-81.

[9] Wang Yan, Mao Mingzhi, Fan Jing, et al. Application of the weighted Markov chain determined by the optimal segmentation method in rainfall prediction [J]. statistics and decision making, 2009, (11): 17-18.

[10] Zhong Yuan Jian, Li Jing, Wang long. Application of Improved Markov chain rainfall prediction model [J]. Journal of University of Jinan (NATURAL SCIENCE EDITION), 2009,23 (4): 402-405.

[11] Li Shujuan. Application of fuzzy Markov chain in rainfall prediction in Yuanzhou District [J]. agricultural science research, 2016,3 (37): 18-21.
[12] sun Zhu Ling. Time series and multivariate statistical analysis [M]. Shanghai: Shanghai Jiao Tong University press, 2016:25-28.