Journal of Mathematics and Statistics 8 (4): 500-505, 2012
ISSN 1549-3644
© 2012 Science Publications
doi:10.3844/jmssp.2012.500.505 Published Online 8 (4) 2012 (http://www.thescipub.com/jmss.toc)

SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODEL FOR PRECIPITATION TIME SERIES

[1]Xinghua Chang, [2]Meng Gao, [1]Yan Wang and [2]Xiyong Hou

[1]School of Mathematics and Information Sciences, Yantai University, Yantai, 264005, China
[2]Key Laboratory of Coastal Zone Environmental Processes,
Yantai Institute of Coastal Zone Research, CAS, Yantai, 264003, China

Received 2012-09-07, Revised 2013-01-10; Accepted 2013-02-20

ABSTRACT

Predicting the trend of precipitation is a difficult task in meteorology and environmental sciences. Statistical approaches from time series analysis provide an alternative way for precipitation prediction. The ARIMA model incorporating seasonal characteristics, which is referred to as seasonal ARIMA model was presented. The time series data is the monthly precipitation data in Yantai, China and the period is from 1961 to 2011. The model was denoted as SARIMA (1, 0, 1) $(0, 1, 1)_{12}$ in this study. We first analyzed the stability and correlation of the time series. Then we predicted the monthly precipitation for the coming three yesrs. The results showed that the model fitted the data well and the stochastic seasonal fluctuation was sucessfuly modeled. Seasonal ARIMA model was a proper method for modeling and predicting the time series of monthly percipitation.

Keywords: Time Series Analysis, SARIMA Model, Forecasting

1. INTRODUCTION

Autoregressive Integrated Moving Average Model (ARIMA), is a widely used time series analysis model in statistics. ARIMA model was firstly proposed by Box and Jenkins in the early 1970s, which is often termed as Box-Jenkins model or B-J model for simplicity (Stoffer and Dhumway, 2010). ARIMA is a kind of short-term prediction model in time series analysis. Because this method is relatively systematic, flexible and can grasp more original time series information, it is widely used in meteorology, engineering technology, Marine, economic statistics and prediction technology, (Kantz and Schreiber, 2004; Cryer and Chan, 2008).

The general ARIMA model is also applicable for non-stationary time series that have some clearly identifiable trends (Stoffer and Dhumway, 2010). We usually denote ARIMA model as ARIMA(p, d, q), where P and q are non-negative integers that correspond to the order of the autoregressive, integrated and moving average parts of the model, respectively. In addition to the general ARIMA model, namely non-seasonal ARIMA(p, d, q) model, we should also consider some periodical time series. The periodicity of periodical time series is usually due to seasonal changes (including monthly, quarterly and degree of weeks change) or some other natural reasons. We can build pure seasona A ARIMA(P,D,Q) model (He, 2004) with the time series date in different cycle and the same phase, the parameters P, D and Q are the relevant seasonal autoregressive parameter, seasonal integrated parameter and seasonal moving average parameter.

Considering the data relation, we can build a multiplication seasonal SARIMA(p, d, q)$(P, D, Q)_s$ model, (Wang *et al*., 2008). The model has been successfully applied in many subjects. In practical applications, the order of model SARIMA is usually not too large (Guo, 2009). If the period of time sereis equals to 12, it can be denoted as SARIMA(p, d, q)$(P, D, Q)_{12}$. In the adjustment of the season, this is a very convenient, steady model.

In this study, we will take the monthly precipitation time series as an example to build an seasonal ARIMA model and then forecast the precipitation in the next few

Corresponding Author: Xinghua Chang, School of Mathematics and the Information Sciences, Yantai University, Yantai, 264005, China

years. Specifically, in a seasonal ARIMA model, once we have smoothed the data and identified the parameters D and d, other parameters P, Q, P and q can be preliminarily identified from the ACF and PACF of the stationary processing series. Other related technologies were also used in the study.

2. MATERIALS AND METHODS

2.1. Seasonal ARIMA Model

The general form of seasonal model SARIMA(p, d, q) (P, D, Q)s is given by:

$$\Phi_p(B^s)\varphi(B)\nabla_s^D\nabla^d x_t = \Theta_Q(B^s)\theta(B)w_t \quad (1)$$

where, $\{w_t\}$ is the nonstationary time series, $\{w_t\}$ is the usual Gaussian white noise process. s is the period of the time series. The ordinary autoregressive and moving average components are represented by polynomials $\phi(B)$ and $\theta(B)$ of orders p and q. The seasonal autoregressive and moving average components are $\Phi_p(B^s)$ and $\Theta_Q(B^s)$, where P and Q are their orders. ∇_d and ∇_s^D are ordinary and seasonal difference components. B is the backshift operator. The expressions are shown as follows:

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$$
$$\Phi_p(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_p B^{ps}$$
$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$$
$$\Theta_Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs}$$
$$\nabla^d = (1-B)^d$$
$$\nabla_s^D = (1-B^s)^D$$
$$B^k x_t = x_{t-k}$$

In this study, we concentrate on monthly precipitation time series. If the seasonal period of the series s = 12. It is clear that we may then rewrite Equation (1) as:

$$\Phi_p(B^{12})\varphi(B)\nabla_{12}^D\nabla^d x_t = \Theta_Q(B^{12})\theta(B)w_t \quad (2)$$

2.2. Model Identification

In the tentative specification phase, namely model identification, the goal is to employ computationally simple techniques to narrow down the range of parsimonious models. The B-J method is only suitable for stationary time series data. In such case, we should possibly observe time series graph and transform the data appropriately.

First, we should construct a time plot of the data and inspect the graph for any anomalies (Cryer and Chan, 2008). If the variance grows with time, it will be necessary stabilize the variance. The next step is to identify preliminary values of autoregressive order P, the order of differencing d, the moving average order q and their corresponding seasonal parameters P, D and Q. Here, the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) are the most important elements (Stoffer and Dhumway, 2010). The ACF measures the amount of linear dependence between observations in a time series that are separated by a lag q. The PACF helps to determine how many autoregressive terms p is necessary. The parameter d is the order of difference frequency from non-stationary time series to stationary series. Furthermore, a time series plot and ACF of data will typically suggest whether any differencing is needed. If differencing is called for, the time plot will show some kind of linear trend.

When preliminary values of D and d have been fixed, the next step is to check the ACF and PACF of $\nabla_{12}^D\nabla^d x_t$ to determine the values of P, Q, P and q. We can further choose parameters using Akaike's Information Criterion (AIC) to determine the values of parameters (Stoffer and Dhumway, 2010).

2.3. Parameters Estimation

Once the model is tentatively established, the parameters and the corresponding standard errors can be estimated using statistical techniques, such as Maximum Likelihood (ML), least square estimation method and Yule-Walker.

2.4. Diagnostic Checking

Generally, this step includes the analysis of the residuals as well as model comparisons. If the model fits well, the standardized residuals should behave as an independent and identically distributed sequence with mean zero and variance one (Cryer and Chan, 2008). A standardized residuals plot or a Q-Q plot can help in identifying the normality (Stoffer and Dhumway, 2010). The model should pass the parametric test and diagnostic check.

2.5. Fitting and Prediction

Once a model has been identified and all the parameters have been estimated, we can predict future values of a time series with this model.

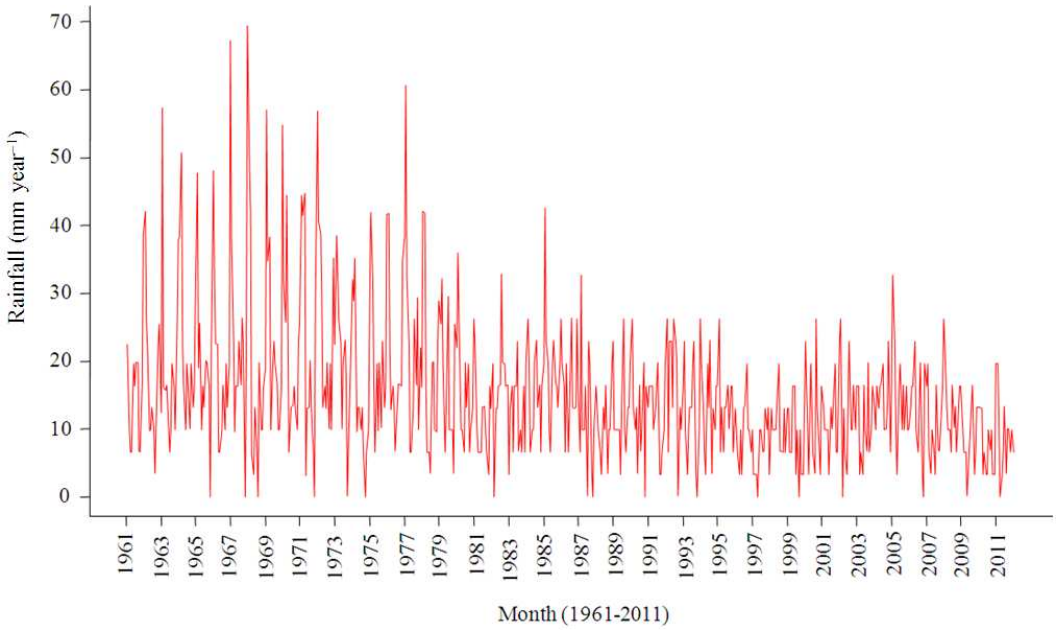

Fig. 1. Time sereis of monthly precipitation data for Yantai station

Table 1. Basic statistics for Yantai monthly precipitation data (in mm)

No of observation	Mean	St Dev	Median	Min	Max
612	15.292	10.419	13.119	0	69.427

2.6. Data

In this study, the time series is the monthly percipitation data from Yantai, a coastal city in China. The annual mean temperature in Yantai is 12°C and the annual precipitation is 620 mm. The data processing tool is the free statistical software R. Time series plot is shown in **Fig. 1**. The descriptive statistics for our data are summarized in **Table 1**.

3. RESULTS

The ACF and PACF of the original data $\{x_t\}$, t = 1,2,…,612, are shown in **Fig. 2**. The ACF and **Fig. 1** show a seasonal fluctuation occur every 12 month, resulting in s = 12 (Wang, 2008; Momani and Naill, 2009). Concentrating on the ACF of original data, we note a slow decreasing trend in the ACF peaks at seasonal lags, h = 1s, 2s, 3s, 4s, where s = 12. It indicates a nonstationary behavior and suggests a seasonal difference.

Figure 3 shows the ACF and PACF of the de-seasonalized precipitation data. The ACF decreases to zero exponentially indicating a stationary behavior (Stoffer and Dhumway, 2010; Han *et al*., 2008). Then the SARIMA (p, 0, q)(P, 1, Q)$_{12}$ model could be fitted to the de-seasonalized data. From ACF of the stationary series, we can see the ACF peak at h =1s; while for PACF, it peaks at h =1s, 2s,…,6s. This phenomenon means that the ACF is cutting off after lag 1s and the PACF is tailing off in the seasonal lags. So we can build two models: (i) an SAR model of order Q = 1, or (ii) an SARMA of orders P = 1,2,…,6 and Q = 1. The characteristic of graph turns out model (i) is much better. Inspecting the ACF and PACF at lags h = 1, 2,…,11, it appears that either: (a) ACF and PACF are both tailing off; (b) PACF cuts off at lag 1, ACF tails off; (c) ACF cuts off at lag 1, PACF tails off.

The result indicates that we should consider the following models and choose a better model based on AIC, AICc and BIC criteria. The optional models and the correlation values are shown in **Table 2**. Obviously, model SARIMA (1,0,1) (0,1,1)$_{12}$ has the smallest value of AIC, AICc and BIC and then we temporarily have a model SARIMA (1,0,1) (0,1,1)$_{12}$. As a rule of thumb in SARIMA modeling, we need to minimize the sum squared of residuals (RSS) and the number of model parameters. We had considered this message when calculating the related values (Stoffer and Dhumway, 2010).

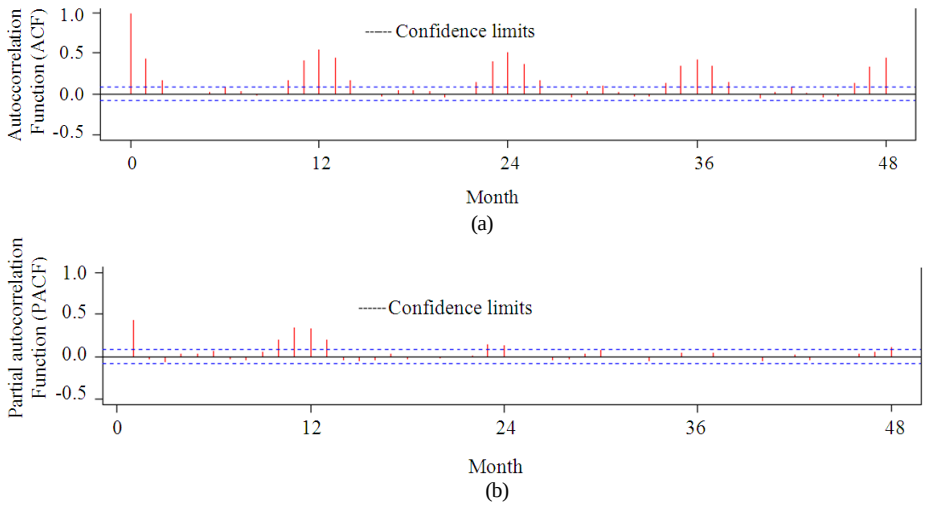

Fig. 2. (a) Autocorrelation (ACF) and (b) Partial Autocorrelation (PACF) for original time series of monthly precipitation data

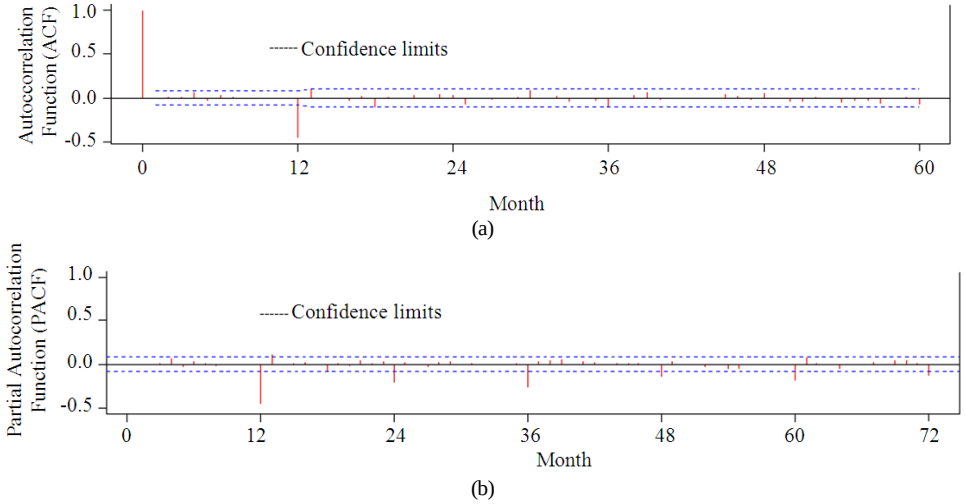

Fig. 3. (a) Autocorrelation (ACF) and (b) Partial Autocorrelation (PACF) for first order seasonal differencing and de-seasonal original precipitation data

The model parameters are estimated using Maximum Likelihood Estimation. The related parameters are shown in **Table 3**, where s.e stands for the standard deviation. It can be observed the parameters of model SARIMA(1, 0, 1)(0, 1, 1)$_{12}$ are all significant. Then we plug the related parameter into the Equation 2 and 3 the fitted model in this case is:

$$\varphi(B)\nabla_{12}^{1}x_t = \Theta_1(B^{12})\theta(B)w_t \quad (3)$$

The diagnostics for the model SARIMA(1, 0, 1)(0, 1, 1)$_{12}$ is displayed in **Fig. 4 and 5**. The standardized residual shows no obvious patterns, although there are a few suspicious values and unusual values (Kantz and Schreiber, 2004). The model fits well although a small amount of autocorrelation still remains. Moreover, we use the Ljung-Box test to examine the independence of the residuals. The p-values of Q-statistic for the first 12 lags of the model are shown in **Fig. 5**.

Finally, predictions based on the fitted model for the next three years are shown in **Fig. 5**. The model SARIMA (1, 0, 1) (0, 1, 1)$_{12}$ could be written as Equation 4:

$$(1-\varphi_1 B)\nabla_{12}x_t = (1+\Theta_1 B^{12})(1+\theta_1 B)w_t \quad (4)$$

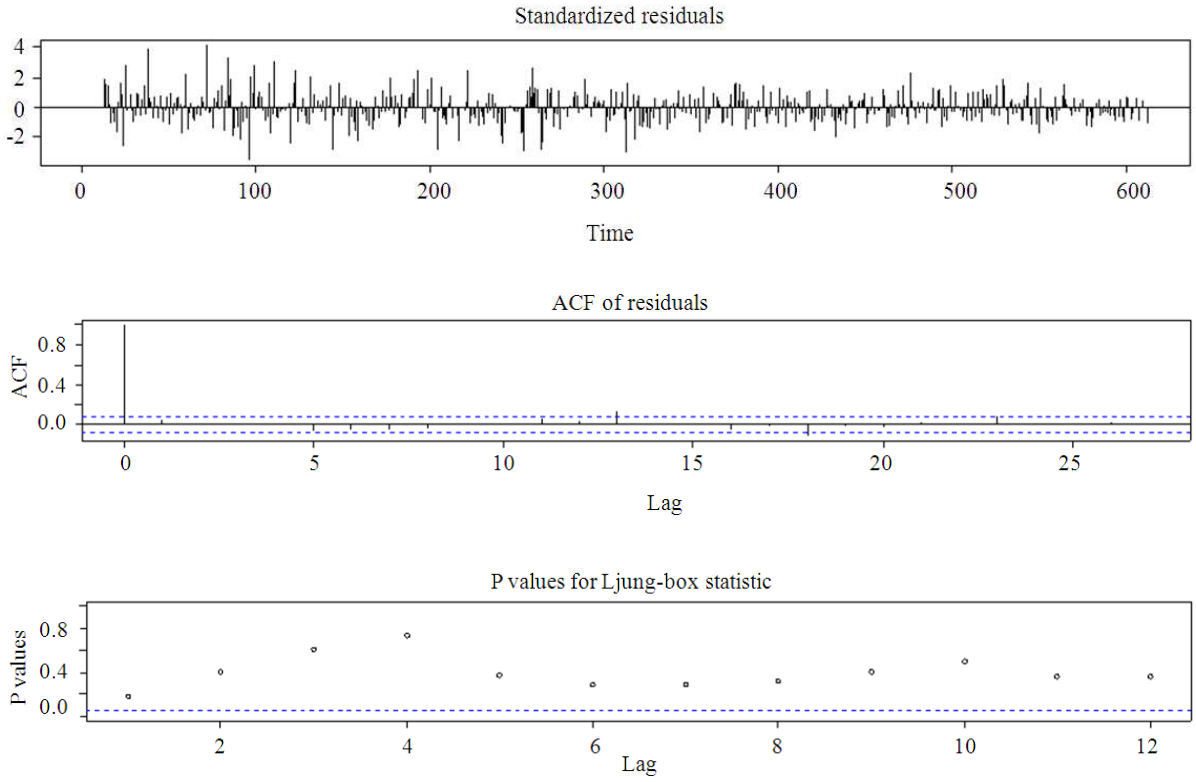

Fig. 4. Diagnostic for the SARIMA (1, 0, 1) $(0,1,1)_{12}$ model

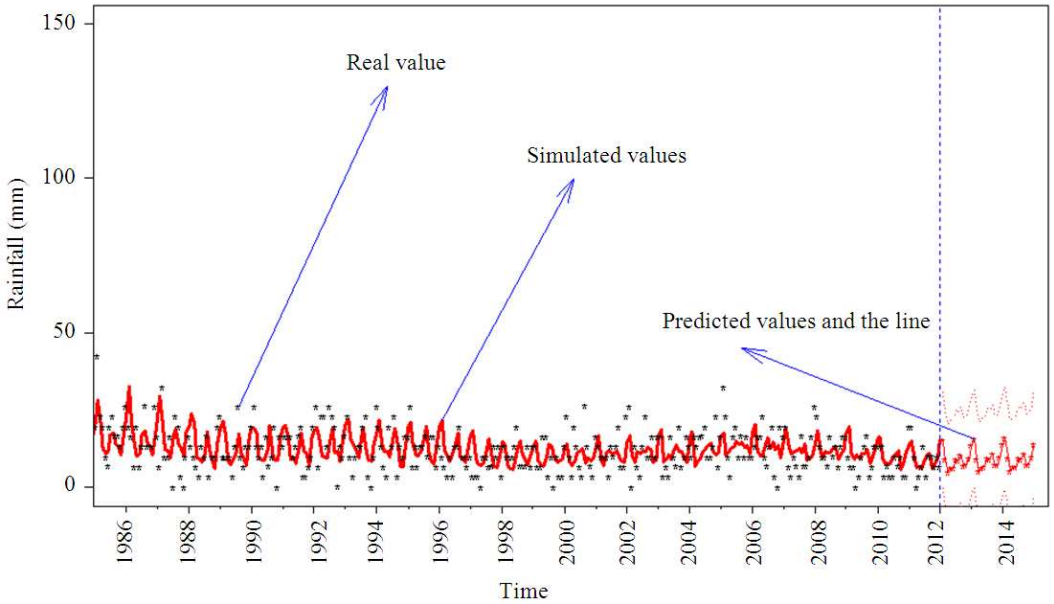

Fig. 5. Predicted and real values of monthly precipitation

Table 2. Optional models and the related standards values

Models	AIC	AICc	BIC
SARIMA (0,0,1) $(0,1,1)_{12}$	5.114070	5.117445	4.135720
SARIMA (1,0,1) $(0,1,1)_{12}$	5.098611	5.102041	4.127479
SARIMA (0,0,0) $(0,1,1)_{12}$	5.123903	5.127235	4.138337
SARIMA (1,0,1) $(0,1,1)_{12}$	5.112628	5.116003	4.134278

Table 3. Estimates of the model parameters

	Model parameter				
Model	ϕ_1	θ_1	Φ_1	RSS	K
SARIMA (1,0,1) $(0,1,1)_{12}$	0.9709	-0.9219	-0.8223	36398.08	4
s.c.	0.0245	0.0376	0.0257		

Or:

$$(1-\varphi_1 B)(1-B^{12})x_t = (1+\Theta_1 B^{12})(1+\theta_1 B)w_t$$

The equation can be multiplied and written in the following form that is used in forecasting, the values of the correlation coefficient as shown in **Table 3**, Equation 5:

$$x_t = \varphi_1 x_{t-1} + x_{t-12} - \varphi_1 x_{t-13} + w_t + \theta_1 w_{t-1} + \theta_1 w_{t-12} + \Theta_1 \theta_1 w_{t-13} \quad (5)$$

Finally, The comparison between the real values and the fitted value is shown in **Fig. 5**. The vertical dotted line separates the data from the predictions.

4. DISSCUSSION

Because of many stochastic environmental factors, such as temperature, geographic location and climate, the model state of precipitation is a complicated dynamical system. The time series model in study does not model the extreme values well. Further extensions of study may be undertaken by considering an intervention time series analysis such as Autoregressive conditional heteroskedasticity model to model the pheneonemon of extremums.

5. CONCLUSION

In this study, an ARIMA model that incorporates the seasonality of time series was presented. Using the time series of monthly precipitation in Yantai, we build a seasonal SARIMA (1, 0, 1) $(0, 1, 1)_{12}$. It was found that the model fitted the data well and the stochastic seasonal fluctuation was sucessfuly modeled except some extreme values. The predictions based on this model indicate that the percipitation in the next three years will decrease. The decreasing trend is consistent with that obtained in our previous study (Gao and Hou, 2012). This changing trend reminds us to make proper strategies of water resource management in response to water shortage.

6. ACKNOWLEDGEMENT

This study was partly supported by National Natural Sciences Foundation of China (10971011; 31000197) and Knowledge innovation project of CAS (KZCX2-EW-QN209).

7. REFERENCES

Cryer, J. D. and K.S. Chan, 2008. Time Series Analysis with Application in R. 2nd Edn., Springer, New York, ISBN-10: 0387759581, pp: 491.

Kantz, H. and T. Schreiber, 2004. Nonlinear Time Series Analysis. 2nd Edn., Cambridge University Press, Cambridge, ISBN-10: 0521529026, pp: 369.

Gao, M. and X.Y. Hou, 2012. Trends and multifractal analyses of precipitation data from shandong peninsula, China. Am. J. Environ. Sci., 8: 271-279. DOI: 10.3844/ajessp.2012.271.279

Guo, Z.W., 2009. The adjustment method and research progress based on the ARIMA model. Chinese J. Hosp. Stat., 161: 65-69.

Han, P., P.X. Wang and Y.J. Wang, 2008. Drought forecasting based on the standardized precipitation index at different temporal scales using ARIMA models. Agric. Res. Arid Areas, 26: 212-218.

He, S.Y., 2004. Applied Time Series Analysis. 1st Edn., Peking University Press, Beijing.

Momani, M. and P.E. Naill, 2009. Time series analysis model for rainfall data in Jordan: Case study for using time series analysis. Am. J. Environ. Sci., 5: 599-604. DOI: 10.3844/ajessp.2009.599.604

Stoffer, D.S. and R.H. Dhumway, 2010. Time Series Analysis and its Application. 3rd Edn., Springer, New York, ISBN-10: 1441978658, pp: 596.

Wang, J., Y.H. Du and X.T. Zhang, 2008. Theory and Application with Seasonal Time Series. 1st Edn., Nankai University Press, Chinese.

Wang, Y., 2008. Applied Time Series Analysis. 1st Edn., China Renmin University Press, Beijing.