Current Journal of Applied Science and Technology

39(8): 69-76, 2020; Article no.CJAST.55802
ISSN: 2457-1024
(Past name: British Journal of Applied Science & Technology, Past ISSN: 2231-0843, NLM ID: 101664541)

SARIMA Modelling and Forecasting of Monthly Rainfall Patterns for Coimbatore, Tamil Nadu, India

S. Kokilavani[1*], R. Pangayarselvi[2], S. P. Ramanathan[1], Ga. Dheebakaran[1] N. K. Sathyamoorthy[1], N. Maragatham[1] and R. Gowtham[1]

[1]*Agro Climate Research Centre, Tamil Nadu Agricultural University, Coimbatore-3, India.*
[2]*Agricultural College and Research Institute, Tamil Nadu Agricultural University, Killikulam, India.*

Authors' contributions

This work was carried out in collaboration among all authors. Author SK designed the study wrote the protocol and the first draft of the manuscript. Author RP performed the statistical analysis and author SPR managed the analyses of the study. Authors GD, NKS, NM and RG managed the literature searches. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/CJAST/2020/v39i830594
Editor(s):
(1) Dr. A. A. Hanafi-Bojd, Tehran University of Medical Sciences, Iran.
Reviewers:
(1) Suryakanti Dutta, India.
(2) Venkata Sanyasi Seshendra Kumar Karri, GITAM University, India.
Complete Peer review History: http://www.sdiarticle4.com/review-history/55802

Original Research Article

Received 05 February 2020
Accepted 11 April 2020
Published 05 May 2020

ABSTRACT

Weather forecasting is an important subject in the field of meteorology all over the world. The pattern and amount of rainfall are the essential factors that affect agricultural systems. The present paper describes an empirical study for modeling and forecasting the time series of monthly rainfall patterns for Coimbatore, Tamil Nadu. The Box-Jenkins Seasonal Autoregressive Integrated Moving Average (SARIMA) methodology has been adopted for model identification, diagnostic checking and forecasting for this region. The best SARIMA models were selected based on the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) and the minimum values of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The study has shown that the SARIMA $(0,0,0)(2,0,0)_{12}$ model was appropriate for analysing and forecasting the future rainfall patterns. The Root Means Square Error (RMSE) values were found to be 52.37 and proved that the above model was the best model for further forecasting the rainfall.

**Corresponding author: E-mail: kokilavani.s@tnau.ac.in;*

Keywords: Rainfall; SARIMA; time series analysis; residual analysis.

1. INTRODUCTION

The atmosphere and ocean have warmed, the amounts of snow and ice have diminished and sea level has risen in the recent past. When the temperature increases beyond 2.5°C, then 20 to 30 per cent of known animal and plant species would be at increased risk of extinction. If the global average temperature increase exceeded 3.5°C, models suggested that there would be extinctions of 40 to 70 per cent of known species [1]. India is one of the 27 countries identified as most vulnerable to the impact of global warming. [2,3] studied the changes in the frequency of rainy days, rainy days as well as heavy rainfall days using the daily rainfall data for the period 1901-2005 all over India. It is a fact that climate change is real and is happening across the world in different magnitudes. The long term mean seasonal and annual rainfall analysis showed that South West Monsoon (SWM) rainfall observed was 176.9 mm and North East Monsoon (NEM) was 336.9 mm with an annual rainfall of 674.8 mm at Coimbatore [4].

The agricultural practices and crop yields of India are heavily dependent on the climatic factors like rainfall. Out of 142 million ha cultivated land in India, 92 million ha (i.e. about 65%) are under the influence of rain-fed agriculture [5,6,3]. Unlike irrigated agriculture, rain-fed farming is usually diverse and risk prone. The monsoon season is the principal rain-bearing season and in fact, a substantial part of the annual rainfall over a large part of the country occurs in this season. Small variations in the timing and the quantity of monsoon rainfall have the potential to impact on agricultural output [7].

Rainfall is natural climatic phenomena whose prediction is challenging and demanding. Its forecasts of particular relevance to the agriculture sector, which contributes significantly to the economy of the nation [8,6]. On a worldwide scale, numerous attempts have been made to predict its behavioural pattern using various techniques. In the last few decades, time series forecasting has received tremendous attention of researchers.

Time series models have been commonly used in a broad range of scientific applications.

Some of the major advantages of time series models include their systematic search capability for identification, estimation and diagnostic checking. Time series models, like the Autoregressive Integrated Moving Average (ARIMA), effectively consider serial linear correlation among observations, whereas Seasonal Autoregressive Integrated Moving Average (SARIMA) models can satisfactorily describe time series thatexhibit non-stationary behaviours both within and across seasons [9].

SARIMA models are the most general forecasting models with high degree of accuracy. An attempt has been made in the present paper to analyse and predict the monthly rainfall patterns for Coimbatore, Tamil Nadu using the SARIMA model.

2. MATERIALS AND METHODS

2.1 Data

In this study, the time series is the average monthly rainfall data of Coimbatore, Tamil Nadu from 1991-2018, obtained from Agro Climate Research Centre, Tamil Nadu Agricultural University. The data processing tool used for the study is R software.

2.2 Seasonal ARIMA Model

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

$$\Phi_P(B^s)\phi_p(B)\nabla_s^D\nabla^d x_t = \mu + \Theta_Q(B^s)\theta_q(B)a_t \quad (1)$$

Where,

$\Phi_P(B^s) = 1 - \Phi B^s - \Phi_2 B^{2s} - LL\Phi_P B^{ps}$ is the seasonal autoregressive operator of order P.

$\phi_P(B) = 1 - \phi_1 B - \phi_2 B^2 - LL\phi_P B^p$ is the regular autoregressive operator of order p.

$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - LL\Theta_Q B^{Qs}$ is the seasonal moving average operator of order Q.

$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - LL\theta_q B^q$ is the regular moving average operator of order q.

Where μ is the intercept term or mean term $\nabla^d = (1-B)^d; \nabla_s^D = (1-B^s)^D; B^k x_t = x_{t-k}$, a_t the non-stationary time series and is the usual Gaussian white noise process; s is the period of the time series and B is the backshift operator [10].

The monthly rainfall took for the study, s=12. Hence, the above equation (1) can be written as

$$\Phi_P(B^{12})\phi_p(B)\nabla_{12}^{D}\nabla^{d}x_t = \mu + \Theta_Q(B^{12})\theta_q(B)a_t$$

2.3 Model Identification

In time series analysis, the most crucial steps are to identify and build a model based on the available data. At this stage it is necessary to identify the values of (*p*,*d*,*q*) and (P,D,Q)*s*. The goal is to employ computationally simple techniques to narrow down the range of parsimonious models. The Box-Jenkins method is only suitable for stationary time series data.

For this purpose, one should construct a time plot of the data and inspect the graph for any anomalies [11]. Through careful examination of the plot, usually one could get an idea about whether the series contains a trend, seasonality, outliers; non-constant variances and other non-normal and non-stationary phenomena. This information would help to choose proper data transformation. If the variance grows with time, we should use variance-stabilizing transformations and difference. A series with non-constant variance often needs a logarithmic transformation.

The next step is to identify preliminary values of auto regressive 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), the partial autocorrelation function (PACF) are the most important elements [12]. 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* are necessary. The parameter *d* is the order of difference frequency changing from non-stationary time series to stationary time 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. Further one could choose parameters using minimum Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC).Once the model is tentatively established, the parameters and the corresponding standard errors can beestimated using statistical techniques.

2.4 Diagnostic Checking

Diagnostic checks have become a standard tool for the identification of adequate models before forecasting the data. The overall test for lack of fit for autoregressive moving average models proposed by [13] and a measure of lack of fit in time series models proposed by [14] are considered. The selected appropriate model is used for forecasting the monthly rainfall.

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 the estimated model.

3. RESULTS AND DISCUSSION

The mean monthly rainfall ranged from 7 mm (January) to 189.7 mm (October) (Table 1). From, April to November, the Co-efficient of Variation is less than 100 per cent and the dependability of rainfall for these months are higher compared to other months.

3.1 Nature of Time Series Data

The nature of the rainfall data from 1991 to 2018 conveyed the presence of the seasonality trend. Hence, forecasting rainfall for successive years using SARIMA model was employed in the present study (Fig. 1).

3.2 Stationarity

The time series plot showed that the data exhibited stationary and the quality of stationarity of the observation was further tested by Augmented Dickey-Fuller test (ADF), KPSS test, PP test (Table 2). The probability value of ADF and PP test were less than 0.05 and greater than 0.05 for KPSS test for the rainfall data [15]. Thus, the data set was considered to be stationary at 5 and 6 lag.

Table 1. Descriptive statistics for rainfall

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Mean	7.0	14.6	26.4	55.7	66.3	38.3	44.0	40.4	63.8	189.7	142.9	30.2
SD	15.5	30.9	41.9	48.4	61.2	25.4	31.9	33.4	56.8	96.9	103.3	42.2
CV	223.3	211.9	158.9	86.9	92.3	66.4	72.6	82.5	89.00	51.1	72.3	139.9
Min	0.0	0.0	0.0	0.8	6.5	6.8	5.1	0.8	0.0	30.0	3.4	0.0
Max	72.7	125.6	151.4	168.8	259.0	107.8	125.8	163.6	218.1	352.1	311.1	161.6

Table 2. Stationarity test for rainfall

Name of the test	Rainfall		
	Value	Lag	P value
ADF test	-7.821	6	0.01
KPSS test	0.0719	5	0.1
PP test	-15.073	5	0.01

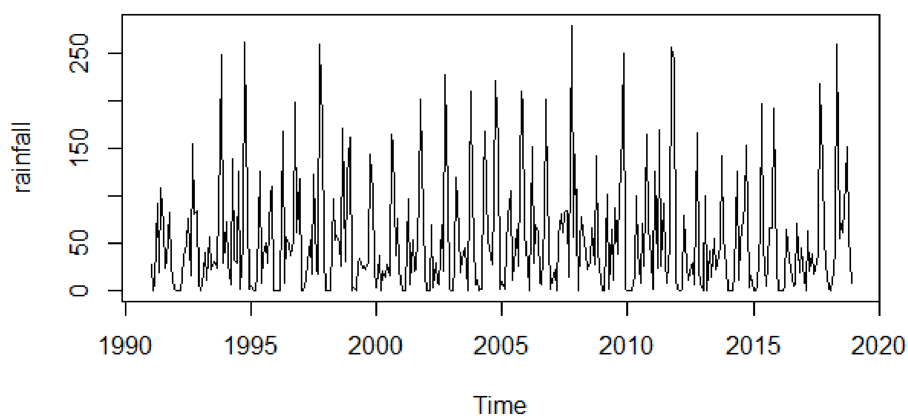

Fig. 1. Time series plot of monthly rainfall

3.3 Model Identification

The correlogram plot (ACF, PACF) for rainfall data set was given in Fig. 2. These plots were used to find the appropriate values of P, p for autoregressive model and Q, q for moving average model.

Various models were framed based on p,d, q, P,D,Q values. It was clearly observed from Table 3, the lowest AIC and BIC values were founded in SARIMA $(0,0,0)(2,0,0)_{12}$ Model. Moreover, the Root Mean Square Error (RMSE) values for the selected model were 52.37 and it was less as compared with other models. Hence, the SARIMA $(0,0,0)(2,0,0)_{12}$ model could be used to forecast future rainfall.

3.4 Diagnostic Checking of the SARIMA Model

This diagnostic check of the model residuals are used to check the adequacy of the fitted model. The normal time series plot of model residuals are shown in Fig. 3.

The residuals were further visualized by ACF and PACF plots. Fig. 4 shows the ACF and PACF pattern of SARIMA model residuals. The plot infers that residual lag values lied within the confident intervals.

3.5 Test for Auto Correlation

Ljung-Box test was used to check the autocorrelation property of the residuals and the results were presented in Table 4. This statistical test was used to find whether any group of the autocorrelation of a time series would differ from zero. The residual coefficients were tested statistically to be non-significant with Ljung-Box statistics. The probability value for both Ljung-Box and Box-Pierce test were more than 0.05 which shows non-significant nature. A non-significant value would conclude that the models were fitted well.

Table 3. SARIMA models for rainfall data

Model	Maximum temperature		RMSE
	AIC	BIC	
SARIMA(2,0,2)(1,0,1)$_{12}$	3695.06	3652.16	54.39
SARIMA(1,0,0)(1,0,0)$_{12}$	3671.65	3686.8	56.25
SARIMA(0,0,0)(1,0,0)$_{12}$	3670.27	3681.65	56.28
SARIMA(0,0,0)(2,0,0)$_{12}$	3627.21	3642.36	52.37

Table 4. Ljung-box test statistic for fitted models

Test	Rainfall		
	Q Statistic	lag	P value
Ljung-Box	21.739	20	0.2438
Box-Pierce	20.799	20	0.2897

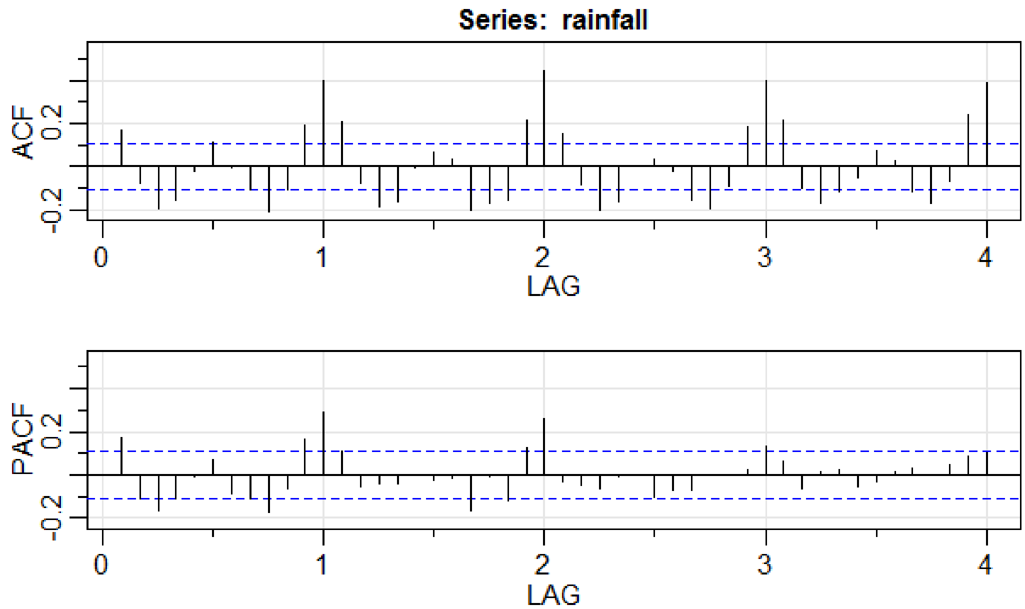

Fig. 2. ACF and PACF plot of rainfall

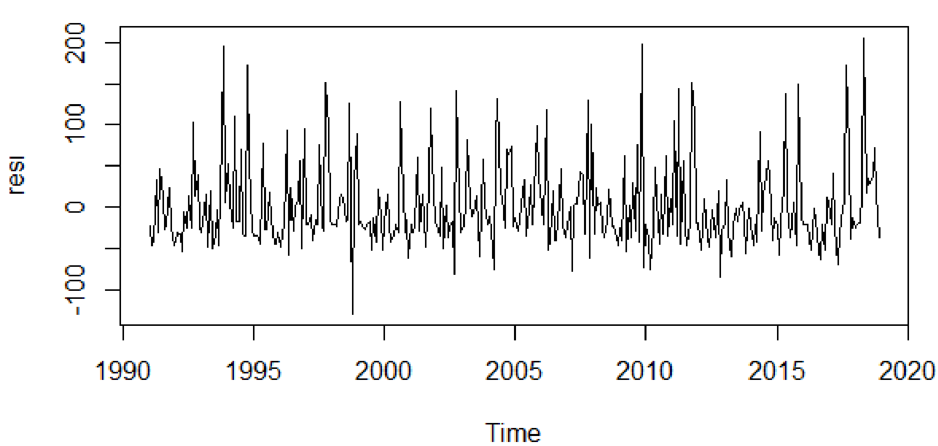

Fig. 3. Time series plot of residuals of the selected model

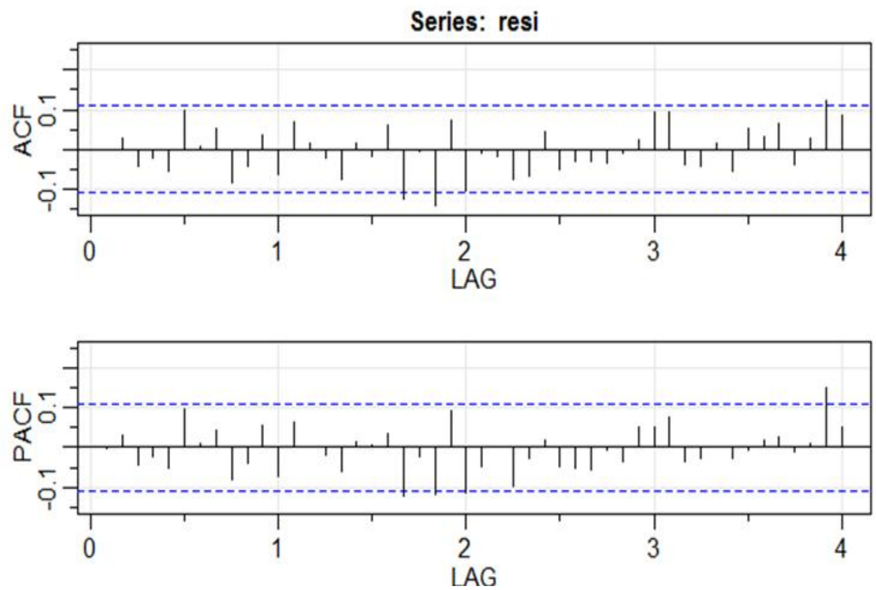

Fig. 4. ACF and PACF plot of fitted model residual

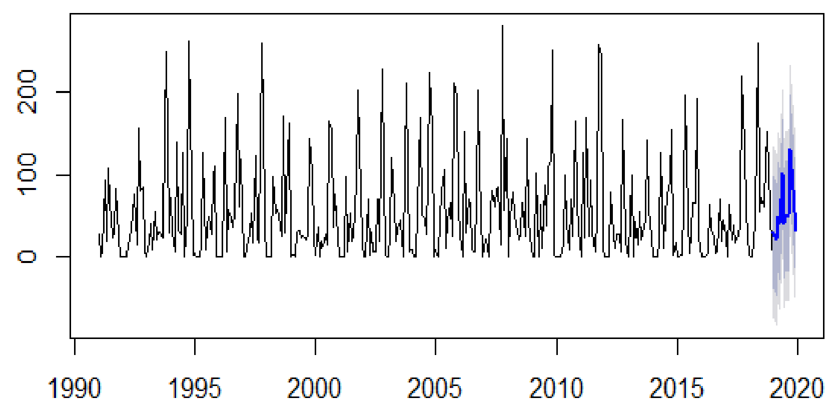

Fig. 5. Forecasted rainfall for 2019-2020

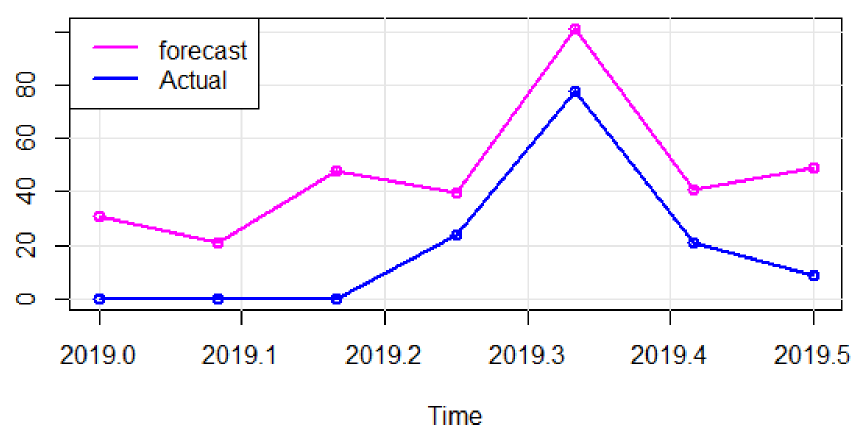

Fig. 6. Actual value Vs forecasted values

3.6 Forecasted Values

Thus, the selected models were considered to be the best model and now used to forecast the rainfall for next two years 2019 and 2020. Fig. 5 shows the forecasted series lies with the original time series data set. The forecasted values were indicated with the confidence limits.

3.7 Testing Data Set with Forecasted Values

Now the forecasted value by the selected model was compared with the actual observed value. Fig. 6 shows the performance of the selected model by plotting the forecasted series and actual series in a single plot. Having fitted the models to actual data, they are used to forecast one step ahead of the observed time series. From the above results it could be concluded that from various models SARIMA(0,0,0)(2,0,0)$_{12}$ model is the best model for forecasting rainfall.

4. CONCLUSION

The study has shown that the SARIMA(0,0,0)(2,0,0)$_{12}$ model was appropriate for analysing and forecasting the future rainfall patterns.The Root Means Square Error (RMSE) values was found to be 52.37 and proved that the above model was the best model for further forecasting the rainfall.

ACKNOWLEDGEMENT

The lead authour of the paper acknowledge the Head of the Department, Agro Climate Research Centre for the sharing the data and complete the research paper in a successful manner.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Intergovernmental Panel of Climate Change. Climate Change, Synthesis Report, Summary for Policymakers; 2014. Available:http://www.ipcc.ch/pdf/assessment-report/ar5/syr/AR5_SYR_FINAL_SPM.pdf, 2014
2. Guhathakurta P, Sreejithand OP, Menon PA. Impact of climate change on extreme rainfall events and flood risk in India. J Earth Syst. Sci. 2011;120(3):359-373.
3. Subbaiah Naidu KCHV. SARIMA modeling and forecasting of seasonal rainfall patterns in India. Int J Math Trends Technol (IJMTT). 2016;38(1):15-22.
4. Kokilavani S, Pangayarselvi R, Panneerselvam S, Dheebakaran Ga. Trend analysis of rainfall variability in western agro climatic zone of Tamil Nadu. Curr. World Environ. 2017;12(1):181-187.
5. Revathy R, Balamurali S. Distinguishing SARIMA with extensive neural network model for forecasting sugarcane productivity. In 2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS) IEEE. 2019;1-5.
6. Reddy PCS, Sureshbabu A. An applied time series forecasting model for yield prediction of agricultural crop. in International Conference on Soft Computing and Signal Processing. Springer, Singapore. 2019;177-187.
7. Graham A, Pathak Mishra E. Time series analysis model to forecast rainfall for Allahabad region. J. Pharmacognosy and Phytochemistry. 2017;6(5):1418-1421.
8. Esack Edwin Raj KV, Rajkumar R. Modelling the impact of climate on tea yield variability in South India. 2019;1981-2015.
9. Wang S, Feng J, Liu G. Application of seasonal time series model in the precipitation forecast. Mathem. & Comp. Modelling. 2013;58:677-683.
10. Subbiah Naidu KCHV. SARIMA Modeling and Forecasting of Seasonal Rainfall Patterns in India. Intl. J. Mathe. Trends and Tech. 2016;38(1):15-22.
11. Cryer JD, Chan KS. Time series analysis with application in R". 2nd edn. Springer, New York. 2008;491. [ISBN-10: 0387759581]
12. Stoffer DS. Dhumway RH. Time series analysis and its application. 3rd edn, Springer, New York. 2010;596. [ISBN-10: 1441978658]
13. Box GEP, Pierce DA. Distribution of residual autocorrelation in autoregressive

integrated moving average models. J. Amer. Stat. Assoc. 1970;64:1509-1526.

14. Ljung GM, Box GEP. On a measure of lack of fit in time series models. Biometrika. 1978;65:297-303.

15. Tadesse KB, Dinka MO. Application of SARIMA model to forecasting monthly flows in Waterval River, South Africa. Journal of water and land development. 2017;35(1):229-236.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/55802