International Journal of Water Resources and Arid Environments 7(2): 178-183, 2018
ISSN 2079-7079

Application of Linear Stochastic Models for Rainfall Data in West Darfur State, Sudan

Tariq Mahgoub Mohamed

Department of Civil Engineering, Jazan University, Jazan, KSA

Abstract: Using of linear stochastic models to simulate monthly rainfall is considered as one of the most important methods for planning different water resources systems. In this paper, linear stochastic models known as multiplicative seasonal autoregressive integrated moving average models, SARIMA, were used to simulate and forecast monthly rainfall at El Geneina gauging station, West Darfur, Sudan. For the analysis, monthly rainfall data during the period 1970 to 2010 were used. The data was obtained from the Sudan Meteorological Authority (SMA). It is observed that it is seasonal. The seasonality observed in Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) plots of monthly data was removed using first order seasonal differencing prior to the development of the model. Obviously, the SARIMA $(1,0,0)x(0,1,1)_{12}$ model was found to be most suitable for simulating monthly rainfall over the region. The model was found appropriate toforecast three years of monthly rainfall and assist decision makers toestablish priorities for water projects.

Key words: Sudan · El Geneina · Monthly rainfall · SARIMA models

INTRODUCTION

West Darfur is one of the states of the Darfur region, Sudan. It is geographically located in western Sudan, boarded by the North Darfur state to the north and northeast, Central Darfur state to the east and southeast, South Darfur state to the south and Chad to the west. The state has an area of 79,460 km^2 and an estimated population of about 1,308,000 according to 2008 PopulationCensus [1]. El Geneina is the capital of the state. The main water resources in the state are the rainfalls, seasonal streams (Wadis) and groundwater. The annual rainfall ranges from less than 200 mm on the northern border to more than 700 mm along the southern boundary. The length of rainy season fluctuates around six months i.e. from May to October. The seasonal streams are usually running during the rainy season. WadiAzum and wadikaja comprise the largest drainage system in the state. WadiAzum has catchment area about 40,000 km^2, with an estimated annual discharge of more than 487 million cubic meters [2]. Most of the agricultural production in West Darfur state is rain fed. The main cash crops grown in this state are sesame, groundnuts and tombac (chewing tobacco). Mangoes are produced in many locations along wadis. The state is also a major producer of livestock.

El Geneinagauge station is located at 13.29° N latitudes and 22.27° E longitudes and has elevation 805 meters above sea level (masl). The station is characterized by annual rainfall of 124 - 661 mm with an annual average of 427 mm and standard deviation of 120 mm during the last forty years. The monthly rainfall records show that most of the rainfalls in the period from May to October and reach its peak in August. According to the station records, the year 1995 received the highest amount of rainfall [661mm] followed by the year 2003 [653 mm]; and the lowest amount of rainfall was recorded in 1984 and 1987 [124 and 238 mm respectively], the years of drought. The annual number of rainy days (rainfall > 1 mm) is 90 days and the mean annual reference potential evapotranspiration (ET_o) using Penman / Monteith criterion for this station is about 2229 mm [3]. The climate in El Geneina is arid with mean annual temperature near 25.9°C [4].

This work is aimed at modeling monthly rainfall records obtained from El Geneina station by seasonal autoregressive integrated moving average (SARIMA) techniques.Rainfall is a seasonal phenomenon of period 12 months. SARIMA modeling has been extensively used to model hydrological seasonal time series. For instance, Nimarla *et al.* [5] fitted a SARIMA model of order $(0,1,1)x(0,1,1)_{12}$ for monthly rainfall in Tamilnadu in India.

Corresponding Author: Tariq Mahgoub Mohamed, Department of Civil Engineering, Jazan University, Jazan, KSA.
E-mail: tariqcivil73@yahoo.com.

Etuk *et al*. [6] applied a SARIMA $(0,0,0)x(0,1,1)_{12}$ model to monthly rainfall data for Wad Madani rainfall station in Sudan. Kibunja *et al*. [7] fitted a SARIMA $(1,0,1)x(1,0,0)_{12}$ to monthly rainfall in Mt. Kenya region, Kenya. Mohamed *et al*. [8] modeledmonthly flow for the Dinder River in Sudanusing a SARIMA $(2,0,0)x(0,1,1)_{12}$. Bazrafshan *et al*. [9] found that the application of SARIMA modelingwas suitable for the forecasting of hydrological drought in the Karkheh Basin.

MATERIALS AND METHODS

Data: For this work, monthly rainfall data were obtained from the Sudan Meteorological Authority (SMA), for the period 1970– 2010 from El Geneinagauge station.

Modeling by SARIMA Methods: A stationary time series can be modeled in different ways: an autoregressive (AR) process, a moving average (MA) process, or an autoregressive and moving average (ARMA) process. However, an ARMA model can be used when the data are stationary, ARMA models can be extended to non-stationary series by allowing differencing of data series. These models are called autoregressive integrated moving average (ARIMA) models. A time series is said to be stationary if it has constant mean and variance.

The general non-seasonal ARIMA model is AR to order p and MA to order q and operates on dth difference of the time series X_t; thus a model of the ARIMA family is classified by three parameters (p, d, q) that can have zero or positive integral values. The general non-seasonal ARIMA model may be written as

$$\phi(B)\nabla^d X_t = \theta(B)\varepsilon_t \tag{1}$$

where:
B = The backward shift operator
$\phi(B)$ and $\theta(B)$ = Polynomials of order p and q, respectively.

$$\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots \phi_p B^p) \tag{2}$$

and

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots \theta_q B^q) \tag{3}$$

Often time series possess a seasonal component that repeats every s observations. For monthly observations s = 12 (12 in 1 year), for quarterly observations s = 4 (4 in 1 year). Box *et al*. [10] has generalized the ARIMA model to deal with seasonality and define a general multiplicative seasonal ARIMA model, which are commonly known as SARIMA models. In short notation the SARIMA model described as ARIMA (p, d, q) x $(P, D, Q)s$, which is mentioned below:

$$\phi_p(B)\Phi_p(B^s)\nabla^d\nabla_s^D(X_t) = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \tag{4}$$

where p is the order of non-seasonal autoregression, d the number of regular differencing, q the order of nonseasonal MA, P the order of seasonal autoregression, D the number of seasonal differencing, Q the order of seasonal MA, s is the length of season, Φ_p and Θ_p are the seasonal polynomials of order P and Q, respectively.

SARIMA model development consists of the following three steps: model identification, parameter estimation and diagnostic checking.The model that gives the minimum Akaike Information Criterion (AIC) and Hannan-Quinn Criterion (HQ) is selected as best fit model [11, 12]. In this work, the statistical and econometric software Eviews was used for all analytical work. It is based on the least squares optimization criterion.

Performance Evaluation: The following measures were used to evaluate the performance of the models:

- Mean Absolute Error:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_i - F_i| \tag{5}$$

- Coefficient of Determination:

$$R^2 = \left[\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(F_i - \bar{F})}{\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2 \sum_{i=1}^{n}(F_i - \bar{F})^2}}\right]^2 \tag{6}$$

- Coefficient of Efficiency:

$$E = 1 - \frac{\sum_{i=1}^{n}(Y_i - F_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \tag{7}$$

where, Y_i are the n observed flows, F_i are the n modeled flows, $\bar{Y}$ is the mean of the observed flows, $\bar{F}$ is the mean of the modeled flows.

RESULTS AND DISCUSSION

Model Identification: Model computation was made with monthly data from between January 1970 and December 2007. The dataset from January 2008 to December 2010 was considered in forecasting estimations of the model.

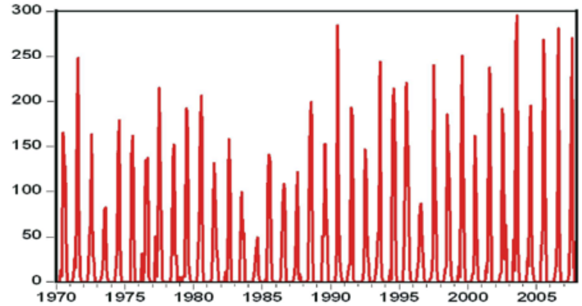
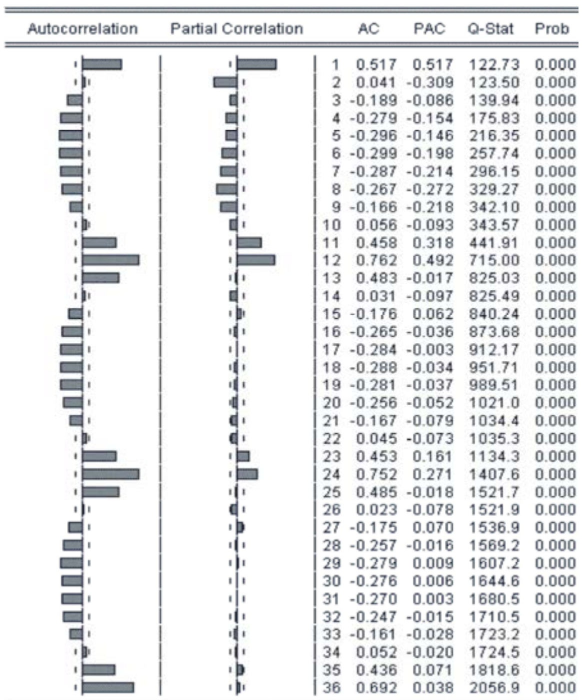


Fig. 1: Monthly rainfall data for El Geneina station [1970-2007] in mm.



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.517	0.517	122.73	0.000
		2	0.041	-0.309	123.50	0.000
		3	-0.189	-0.086	139.94	0.000
		4	-0.279	-0.154	175.83	0.000
		5	-0.296	-0.146	216.35	0.000
		6	-0.299	-0.198	257.74	0.000
		7	-0.287	-0.214	296.15	0.000
		8	-0.267	-0.272	329.27	0.000
		9	-0.166	-0.218	342.10	0.000
		10	0.056	-0.093	343.57	0.000
		11	0.458	0.318	441.91	0.000
		12	0.762	0.492	715.00	0.000
		13	0.483	-0.017	825.03	0.000
		14	0.031	-0.097	825.49	0.000
		15	-0.176	0.062	840.24	0.000
		16	-0.265	-0.036	873.68	0.000
		17	-0.284	-0.003	912.17	0.000
		18	-0.288	-0.034	951.71	0.000
		19	-0.281	-0.037	989.51	0.000
		20	-0.256	-0.052	1021.0	0.000
		21	-0.167	-0.079	1034.4	0.000
		22	0.045	-0.073	1035.3	0.000
		23	0.453	0.161	1134.3	0.000
		24	0.752	0.271	1407.6	0.000
		25	0.485	-0.018	1521.7	0.000
		26	0.023	-0.078	1521.9	0.000
		27	-0.175	0.070	1536.9	0.000
		28	-0.257	-0.016	1569.2	0.000
		29	-0.279	0.009	1607.2	0.000
		30	-0.276	0.006	1644.6	0.000
		31	-0.270	0.003	1680.5	0.000
		32	-0.247	-0.015	1710.5	0.000
		33	-0.161	-0.028	1723.2	0.000
		34	0.052	-0.020	1724.5	0.000
		35	0.436	0.071	1818.6	0.000
		36	0.692	0.038	2056.9	0.000

Fig. 2: ACF and PACF Plots for El Geneina Station Monthly Rainfall Series

The graphical presentation of the monthly rainfall data, Figure 1, shows clearly that there is a seasonal cycle in the series. Figure 2 shows the sample autocorrelation function and partial autocorrelation for the monthly data. The seasonal autocorrelation relationships are shown quite prominently, which proves that the series is non-stationary. Non-stationarity is also confirmed by the Augmented Dickey-Fuller (ADF) unit root test on the monthly rainfall data in Table 1. The table displays results of the test: statistic value -1.076 greater than critical vales –2.570, –1.941, –1.616 all at 1%, 5% and 10%, respectively. This indicates that the series is non-stationary and also confirm that the data needs differencing in order to be stationary. Etuk *et al*. [13] also observed monthly rainfall in Sudan asnon-stationary. This non-stationarity stems from the seasonal nature of the series.

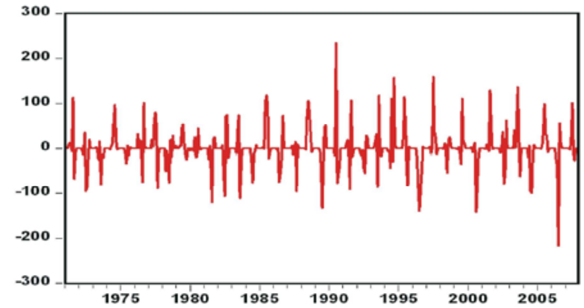


Fig. 3: Monthly rainfall data after one seasonal differencing

A seasonal differencing of the monthly data, as shown in Figure 3, yields a series has a flat trend. The ADF test was done again on the seasonally differenced data. The results of the test: statistic value -9.311 less than critical vales -2.570, -1.941, -1.616 all at 1%, 5% and 10% respectively. This adjudges that the differenced series is stationary. Figure 4 shows the ACF and PACF plots of the data after we take seasonal difference. It appears that most of the seasonality is gone now and the data became stable. The autocorrelation structure in Figure 4 suggests two SARIMA models. The suggested models, the Akaike Information Criterion (AIC) and the Hannan-Quinn Criterion (HQ) values are shown in Table 2. The model that gives the minimum AIC) and (HQ) criterion is selected as best fit model. Obviously, model SARIMA $(1,0,0)x(0,1,1)_{12}$ has the smallest values of AIC andHQ, then one would temporarily have a model SARIMA $(1,0,0)\times(0,1,1)_{12}$.

Parameter Estimation: After the identification of model using the AIC and HQ criteria, estimation of parameters is done. The value of the parameters, standard errors, t-

Table1: ADF-unit root test for El Geneina monthly Rainfall

Station	Variable	ADF Test	Level of Confidence	Critical Value	Probability	Result
El Geneina	Monthly Rainfall	–1.076	1%	–2.570	0.2551	Non-stationary
			5%	–1.941		
			10%	–1.616		

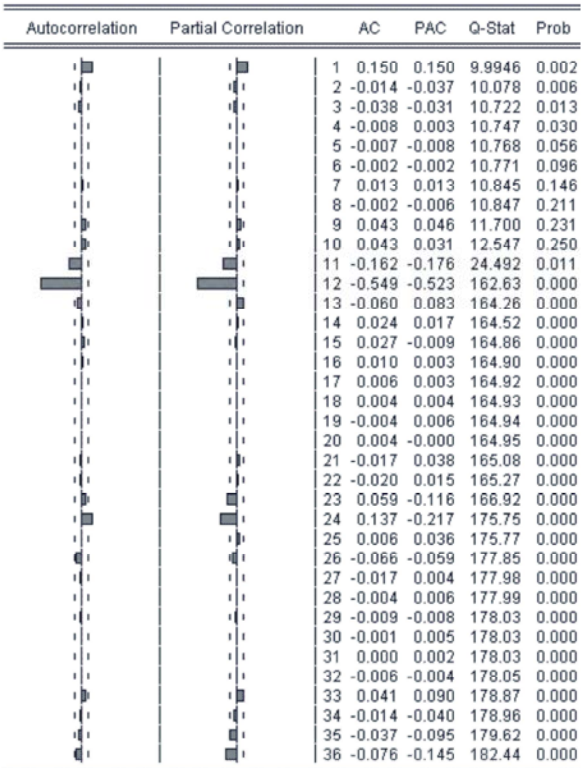


Fig. 4: ACF and PACF Plots for El Geneina Station after one Seasonal Difference

Table 2: Comparison of the Suggested Models

Variable	Station	Model	AIC	HQ
Monthly	El Geneina	SARIMA(0,0,0)x(0,1,1)$_{12}$	9.7009	9.7046
Rainfall		SARIMA(1,0,0)x(0,1,1)$_{12}$	9.6946	9.7019

Table 3: Estimation of SARIMA (1,0,0)x(0,1,1) $_{12}$ Model

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.114548	0.047361	2.418637	0.0160
MA(12)	-0.886760	0.026198	-33.84858	0.0000
R-squared	0.449307	Mean dependent var		0.109029
Adjusted R-squared	0.448058	S.D. dependent var		41.39993
S.E. of regression	30.75717	Akaike info criterion		9.694628
Sum squared resid	417187.6	Schwarz criterion		9.713109
Log likelihood	-2145.360	Hannan-Quinn criter.		9.701917
Durbin-Watson stat	1.996101			
Inverted AR Roots	.11			
Inverted MA Roots	.99	.86-.50i	.86+.50i	.50-.86i
	.50+.86i	.00+.99i	-.00-.99i	-.50+.86i
	-.50-.86i	-.86+.50i	-.86-.50i	-.99

statistic and p-values are listed in Table 3. The result proved that the parameters are significant since their p-values are smaller than 0.05. The selected model is tested for stationarity and invertibility using the inverted AR and MA roots. As reported in Table 3 all the absolute values of the inverted AR and MA roots are smaller than one, hence the selected model is stationary and invertible.

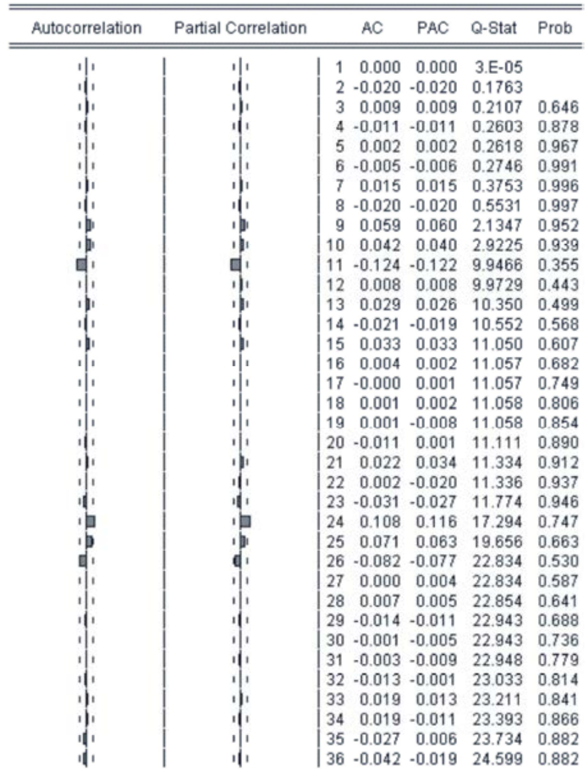


Fig. 5: ACF and PACF Plots for SARIMA (1,0,0)x(0,1,1) 12 Residual

Diagnostic Check: Once an appropriate model is selected and its parameters are estimated, the Box-Jenkins methodology requires examining the residuals of the model to verify that the model is an adequate one for the time series. An adequate model should have uncorrelated residuals. This is the minimal condition. Different tests were carried out on the residual series. The tests are summarized briefly in the following paragraphs.

ACF and PACF of Residuals: The ACF and PACF of residuals of the model SARIMA (1,0,0)x(0,1,1)$_{12}$ are shown in Figure 5. Most of the values of the RACF and RPACF lie within confidence limits. The figure indicates no significant correlation between the residuals.

The Ljung-Box Test: The Ljung-Box test is used for checking independence of residual. From Figure 5, the goodness of fit values for the autocorrelations of residuals from the model up to lag 36 was = 0.05. The result proves the acceptance of the null hypothesis of model adequacy at the 5% significance level and the set of autocorrelations of residuals was considered white noise.

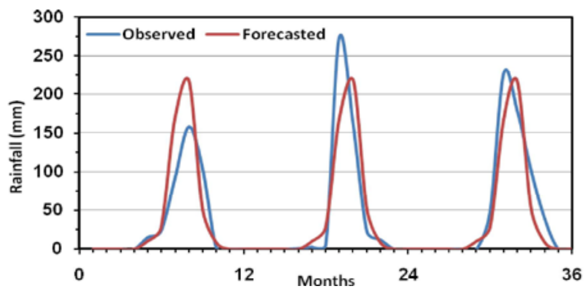


Fig. 6: Forecasting of monthly rainfall using developed SARIMA model (1, 0, 0) × (0, 1, 1)$_{12}$, (2008–2010)

Table 4: The Breusch-Godfrey Serial Correlation LM Test

F-statistic	0.792869	Prob. F(12,429)	0.6580
Obs*R-squared	8.870571	Prob. Chi-Square(12)	0.7139

Table 5:Forecasting accuracy statistic

Statistic measures	Value
MAE	18
R^2	0.81
E	80%

The Breusch-Godfrey Serial Correlation LM Test: The Breusch-Godfrey serial correlation LM test accepts the hypothesis of no serial correlation in the residuals, as shown in Table 4. Hence, the model is adequate.

Forecasting of Monthly Rainfall: SARIMA model can also be used for forecasting future values based on the historical data. The SARIMA (1,0,0)×(0,1,1)$_{12}$ model was tested for its validity to forecast 36 observations obtained for the years 2008 to 2010 for the station. The observed rainfall was found to be closely aligned to the forecasted values, Figure 6.

Forecasting Accuracy: If the fitted SARIMA (1,0,0)×(0,1,1)$_{12}$ model has to perform well in forecasting, the forecast error will be relatively small. To check goodness of the prediction, Mean Absolute Error (MAE), coefficient of determination (R^2) and Nash-Sutcliffe efficiency criteria (E) were used. Table 5 illustrates all of the statistic measures. From the statistics measurement, Table 5, it is observed that the model has lowervalue of MAE. The coefficient of determination (R^2) value of 0.81 and Nash-Sutcliffe efficiency criteria (E) value of 80% showed the very good performance of the model.

CONCLUSION

In this work, linear stochastic model known as Seasonal Autoregressive Integrated Moving Average model, SARIMA, was used to simulate monthly rainfall for El Geneina gauge station, Sudan. It has been demonstrated that the monthly rainfall follow a SARIMA (1,0,0)x(0,1,1)$_{12}$ model. This model may be used as the basis for forecasting rainfall in this region. The fitting of SARIMA models to rainfall time series could result in a better tool, which can be used for water resource planning. SARIMA model has the ability to predict accurately the future monthly rainfall for all gauge stations in Sudan.

REFERENCES

1. Central Bureau of Statistics, 2009. Statistical Year Book for the Year 2009,the Republic of Sudan.
2. Barsi, B.I., 2010. Water resources of wadi systems in Darfur, International Seminar on Challenges in Applications of Integrated Water Resources Management, Water and Environment Centre, Sana'a University, Yemen.
3. Le Houérou, H.N., 2009. Biomaclimatology and Biography of Africa. Verlag Berlin Heidelberg, Germany.
4. Elagib, N.A. and M.G. Mansell, 2000. Recent trends and anomalies in mean seasonal and annual temperatures over Sudan. Journal of Arid Environments, 45(3): 263-288.
5. Nirmala, M. and S.M. Sundaram, 2010. A seasonal arima model for forecasting monthly rainfall in Tamilnadu. National Journal on Advances in Building Sciences and Mechanics, 1(2): 43-47.
6. Etuk, E.H. and T.M. Mohamed, 2014. A seasonal ARIMA model for forecasting monthly rainfall in Gezira Scheme, Sudan. Journal of Advanced Studies in Agricultural, Biological and Environmental Sciences (JABE), 1(1): 1-9.
7. Kibunja, H.W., J.M. Kihoro, G.O. Orwa and W.O. Yodah, 2014. Forecasting precipitation using SARIMA model: A Case Study of Mt. Kenya Region, Mathematical Theory and Modeling, 4(11): 50-58.
8. Mohamed, T.M. and E.H. Etuk, 2015. Simulation of monthly flow for the Dinder River, Sudan, Journal of Basic And Applied Research International, 14(4): 199-205.
9. Bazrafshan, O., A. Salajegheh, J. Bazrafshan, M. Mahdavi and A.F. Marj, 2015. Hydrological drought forecasting using ARIMA models (a case study: Karkheh Basin), Ecopersia, 3(3): 1099-1117.
10. Box, G.E.P., G.M. Jenkins and G.C. Reissel, 1994. Time Series Analysis Forecasting and Control, 3rd edition. Prentice Hall, USA.

11. Akaike, H., 1974. A New look at the statistical model identification. IEEE Trans Automatic Control, AC-19:716–723.
12. Hannan, E.J. and B.G. Quinn, 1979. The determination of the order of an autoregression, Journal of the Royal Statistical Society, Series B, (41): 190-195.
13. Etuk, E.H. and T.M. Mohamed, 2014. Time series analysis of monthly rainfall data for the Gadaref rainfall station, Sudan, by SARIMAmethods. International Journal of Scientific Research in Knowledge, 2(7): 320-327.