International Journal of Science and Advanced Innovative Research
ISSN: 2536-7315 (Print) 2536-7323 (Online)
Volume 5, Number 3, September 2020
http://www.casirmediapublishing.com

SARIMA MODELLING OF MONTHLY RAINFALL IN RIVERS STATE OF NIGERIA

[1]Love Cherukei Nnoka, [2]Ifeoma Better Lekara-Bayo & [2]Ette Harrison Etuk
[1]Department of Statistics, Elechi Amadi Polytechnic, Rumuola, Port Harcourt, Nigeria
[2]Department of Mathematics, Rivers State University, Port Harcourt, Nigeria
Email: **nnokalove@yahoo.com**

ABSTRACT

This work is about seasonal autoregressive integrated moving average (SARIMA) modeling of monthly rainfall of Rivers State of Nigeria from 1981 to 2016. The time plot shows seasonality of period 12 months as typical of rainfall data. Even though the Augmented Dickey Fuller test of unit root ceritifies the series as stationary the correlogram shows an undulating sinusoidal pattern of seasonality of period 12, as expected. The correlogram shows positive spikes at lag 12 and comparable spikes at the autocorrelation function (ACF) at lags 11 and 13 and spikes at lags 12 and 24 on the partial autocorrelation function (PACF). This suggests the involvement of a seasonal autoregressive order of 1 or 2 and a moving average non-seasonal order of 1 and a seasonal order of 1. This means the involvement of a SARIMA$(0,0,1)x(1,1,1)_{12}$ model. Other models worth testing are SARIMA$(0,0,1)x(2,1,1)_{12}$ model and a SARIMA$(1,0,1)x(0,1,1)_{12}$ model. By AIC and R^2 the lattest model was chosen. The correlogram of the residuals showed no significant spike, an evidence of model adequacy. The forecasts of 2017 were obtained finally.

Keywords: Rainfall, Rivers State, Nigeria, SARIMA modeling

INTRODUCTION AND LITERATURE REVIEW

Rivers State is one of the six states in the Niger Delta of the South South of Nigeria. The major occupation of dwellers of this state is agriculture (fishing and farming). The major livelihood of this people is rain dependent. The state would have benefited more in agriculture if farmers had access to reliable and efficient timely rainfall forecasts like these ones obtained here. This study would also benefit other sectors in the state that depend on reliable forecasts of climatic conditions such as tourism and industries. Data covering from 1981 to 2016 were obtained for this study. So many works have been done by different researchers on rainfall.

Chonge *et al.* (2015) used general autoregressive integrated moving average (ARIMA) family to fit a time series model to rainfall pattern in Uasin Gishu County of Kenya. Their result was that a SARIMA$(0,0,0)x(0,1,2)_{12}$ best fitted the Kapsoya historical rainfall data. Inderjeet and Sabita (2008) employed seasonal ARIMA model for prediction of temperature and rainfall on monthly scales for the state of Uttar Pradesh of India. They used periodic data to formulate the SARIMA model and in determination of model parameters. The performance evaluation of the model was carried out on the basis of correlation coefficient (R^2) and Root Mean Square Error (RMSE). Their result showed that the SARIMA approach provided reliable and satisfactory predictions for rainfall and temperature parameters on monthly scale.

Cowden *et al.* (2010) examined stochastic rainfall modeling in West Africa. They examined two stochastic rainfall models: Markov Models (MMs) and Large Scale Weakening (LARSWG). A first order Markov occurrence model with mixed with mixed exponential amount was selected as the best option for unconditional Markov models. They concluded that there was no clear advantage in selecting Markov models over the LARSWG model for Domestic Rainfall in West Africa.

Farajzadeh *et al.* (2012) assessed the modeling of monthly rainfall and run off of Urimia Lake Basin of Iran using Feed-Forward Network (FFNN) and Time Series Analysis Models (TSAM). They applied an ARIMA model to forecast the monthly rainfall in Urimia Basin and found that the estimated values of monthly rainfall through Feed-Forward Neural Network were close to ARIMA model with coefficient of correlation 0.62 and the Root Mean Square Error (RMSE) of 12.43.

Bari *et al.* (2015) built a SARIMA model using Box and Jenkins a(1976) method to forecast long-term rainfall ion Syihet City in Bangladesh. The used rainfall data from 1980-2010 of Syihet Station to build and check the model. The rainfall data from 1980-2006 were used to develop the model while the data from 2007-2010 were used for checking and forecasting. Their result showed that SARIMA$(0,0,1)x(1,1,1)_{12}$ was found most effective to predict future precipitation with 95% confidence interval. Jabrin *et al.* (2014) employed a SARIMA method to model and forecast rainfall pattern in Kano State, Nigeria. From their findings, the method of estimation and the model diagnostic revealed that the SARIMA$(0,0,0)x(1,1,1)_{12}$ adequately fitted the data.

Ogunrinde (2012) also used Box-Jenkins methodology to build a SARIMA model for time series of rainfall of Lagos State and also used SARIMA (2,0,0) model in applying time series to model rainfall in Maiduguri, North Eastern Nigeria. After some diagnostic tests, he found that an SARIMA (1,1,0) model provided a good fit for the rainfall data of Maiduguri and also found that the model was appropriate for the short-term forecast. Eni and Adeyeye (2015) worked on rainfall data from Warri town in Nigeria. They found that the model SARIMA$(1,1,1)\times(0,1,1)_{12}$ was adequate after meeting the criterion of model parsimony with the Residual Sum of Squares (RSS) value of 81.098, Akaike's Information Criterion (AIC) of 281 and Schwartz's Bayesian Criterion (SBC) value of 281.

MATERIALS AND METHOD

Data:

The data for this work are from 2018 Statistical Bulletin_Real Sector, page c5.1, from the Central Bank of Nigeria website www.cbn.org .

Sarima Modelling

Let $\{X_t\}$ be a time series. Suppose that it is stationary. It is said to follow an autoregressive moving average model of order p and q (denoted by ARMA(p, q)) if it satisfies the following equation

$$X_t - \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_p X_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} \quad \dots\dots(1)$$

where the α's and β's are constants such that (1) be both stationary and invertible and $\{\varepsilon_t\}$ is a white noise process.

Equation (1) may as well be written as

$$A(L)X_t = B(L)\varepsilon_t \quad \dots\dots(2)$$

where $A(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p$ and $B(L) = 1 + \beta_1 L + \beta_2 L^2 + \dots + \beta_q L^q$ and L is the backshift operator defined by $L^K X_t = X_{t-k}$.

If $\{X_t\}$ is not stationary according to Box and Jenkins (1976, 2004), a certain difference of $\{X_t\}$, $\nabla^d(X_t)$, might be, where $\nabla = 1 - L$ and d is a positive integer. Then if $\{X_t\}$ is replaced by $\nabla^d(X_t)$ in (1), the model becomes an autoregressive integrated moving average of order p, d, q, denoted by an ARIMA(p, d, q), in $\{X_t\}$. If the series is seasonal of period s, it is said to follow a seasonal autoregressive integrated moving average model of order (p, d, q)x(P, D, Q) (denoted by a SARIMA(p ,d, q)x(P, D, Q)s) if

$$A(L)\Phi(L^s) \nabla^d \nabla_s^D X_t = B(L)\Theta(L^s)\varepsilon_t \quad \ldots\ldots(3)$$

where $\Phi(L^s) = 1 - \phi_1 L^s - \phi_2 L^{2s} - \ldots - \phi_P L^{Ps}$ and $\Theta(L^s) = 1 + \theta_1 L^s + \theta_2 L^{2s} + \ldots + \theta_Q L^{Qs}$ and ∇_s^D is the seasonal difference operator such that $\nabla_s = 1\text{-}L^s$. Here P is the seasonal autoregressive order; Q is the seasonal moving average order; the ϕ's and the θ's are the seasonal autoregressive and the seasonal moving average parameters.

In practice, an autoregressive component of lag p is identified when the ACF cuts off at lag p and a moving average component of order q when the partial autocorrelation function cuts off at lag q. Stationarity of a time series is assured if the Augmented Dickey Fuller (ADF) unit root test is significant.

Computer Software

Eviews 10 shall be used throughout this work. It is based on the maximum likelihood estimation procedure.

RESULTS AND DISCUSSION

The time plot of Figure 1 is typical of rainfall data as it shows seasonality of period 12 months, with the peaks around the middle of the year and the troughs around the end of the year and with no noticeable secular trend. The Augmented Dickey Fuller (ADF) unit root test of Table 1 gives an impression that the series is I (0) i.e. stationary. However the correlogram of Figure 2 shows a sinusoidal pattern of period 12 months with peaks at multiples of 12 lags and the troughs midway between the peaks. This shows clearly that the series is seasonal of period 12 months and no seasonal data is stationary.

A seasonal (i.e. 12 monthly) differencing is done to rid the series of the seasonality and make it stationary as evidenced by the ADF test of Table 2. The time plot of Figure 3 shows the differenced series devoid of any seasonality and trend. The correlogram of the differenced series of Figure 4 shows positive spikes at lag 12 and comparative spikes at lags 11 and 13 in the ACF. This indicates a seasonal moving average component with a non-seasonal component of order one and a seasonal component of order one. The PACF has spikes at lags 12 and 24 indicating a seasonal autoregressive component of order 2. Therefore the models entertained are as given in Table 3 and their AICs and R^2's. Based on Table 3 data, the model chosen was the SARIMA(1,0,1)x(0,1,1)$_{12}$ model given in Table 4 by

$\nabla_{12} X_t = -0.8810\nabla_{12} X_{t-1} + 0.8890\varepsilon_{t-1} - 0.9867\varepsilon_{t-12} - 0.9024\varepsilon_{t-13} + \varepsilon_t, \hat{\sigma}^2 = 6272$ ……………………………………………………………………………….. (4)

where X_t is the monthly rainfall data in Rivers State at time t. This model (4) is certified adequate by the correlogram of its residuals if Figure 5. All spikes are within 95% confidence limits.

Table 5 gives forecasts of 2017 based on model (4).

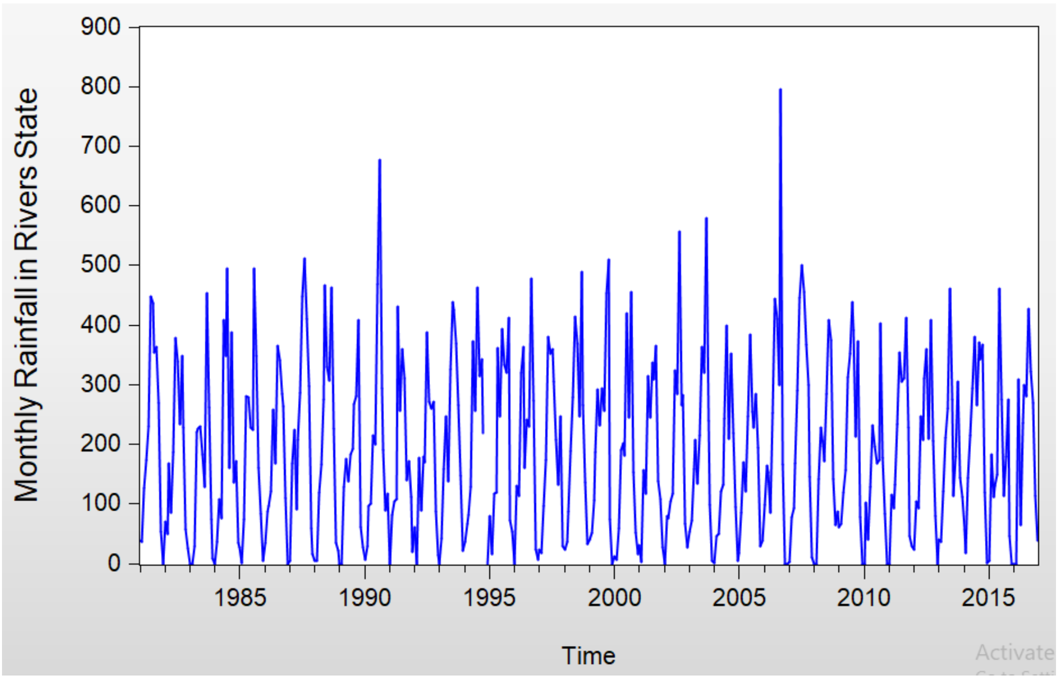


Figure 1: Time Plot of monthly rainfall in Rivers State

Table1: Augmented Dickey Fuller Test of the original data

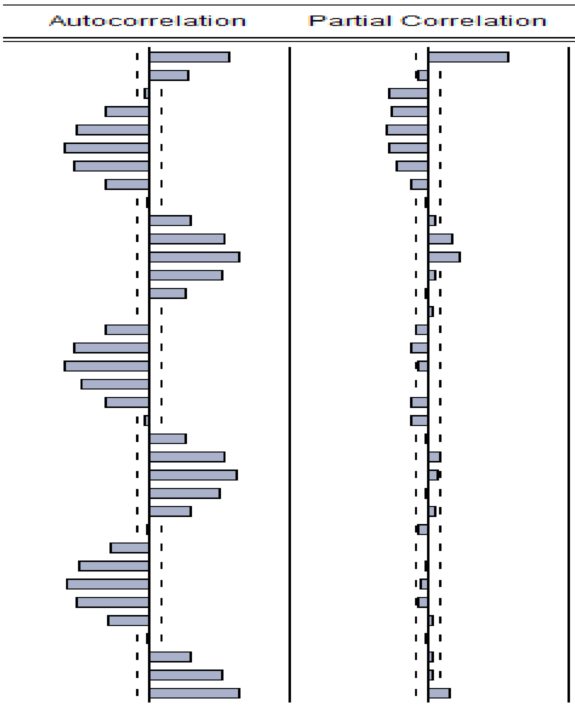
Null Hypothesis: MRRS has a unit root
Exogenous: Constant
Lag Length: 11 (Automatic - based on SIC, maxlag=17)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.276489	0.0000
Test critical values:	1% level	-3.446201	
	5% level	-2.868422	
	10% level	-2.570501	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(MRRS)
Method: Least Squares
Date: 09/10/20 Time: 05:09
Sample (adjusted): 1982M01 2016M12
Included observations: 407 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MRRS(-1)	-1.031597	0.195508	-5.276489	0.0000
D(MRRS(-1))	0.225089	0.184610	1.219268	0.2235
D(MRRS(-2))	0.215842	0.170521	1.265776	0.2063
D(MRRS(-3))	0.110208	0.155638	0.708105	0.4793
D(MRRS(-4))	0.073402	0.138639	0.529448	0.5968
D(MRRS(-5))	-0.016387	0.122085	-0.134228	0.8933
D(MRRS(-6))	-0.147100	0.106696	-1.378676	0.1688
D(MRRS(-7))	-0.278985	0.093303	-2.990095	0.0030
D(MRRS(-8))	-0.379518	0.081496	-4.656888	0.0000
D(MRRS(-9))	-0.378797	0.070081	-5.405114	0.0000
D(MRRS(-10))	-0.389018	0.060151	-6.467341	0.0000
D(MRRS(-11))	-0.243285	0.048991	-4.965894	0.0000
C	195.8641	37.44266	5.231040	0.0000
R-squared	0.537183	Mean dependent var		0.452334
Adjusted R-squared	0.523087	S.D. dependent var		131.8211
S.E. of regression	91.03413	Akaike info criterion		11.89177
Sum squared resid	3265162	Schwarz criterion		12.01981



	AC	PAC	Q-Stat	Prob
1	0.585	0.585	148.28	0.000
2	0.292	-0.075	185.36	0.000
3	-0.044	-0.280	186.20	0.000
4	-0.320	-0.263	230.94	0.000
5	-0.541	-0.312	359.02	0.000
6	-0.626	-0.288	531.00	0.000
7	-0.553	-0.233	665.76	0.000
8	-0.328	-0.138	713.36	0.000
9	-0.012	-0.023	713.43	0.000
10	0.297	0.048	752.49	0.000
11	0.561	0.184	892.14	0.000
12	0.672	0.222	1093.3	0.000
13	0.540	0.053	1223.7	0.000
14	0.277	-0.015	1258.1	0.000
15	-0.005	0.038	1258.1	0.000
16	-0.323	-0.084	1305.0	0.000
17	-0.557	-0.125	1444.8	0.000
18	-0.626	-0.071	1621.8	0.000
19	-0.513	-0.006	1741.1	0.000
20	-0.329	-0.127	1790.2	0.000
21	-0.048	-0.123	1791.2	0.000
22	0.273	-0.024	1825.3	0.000
23	0.553	0.085	1965.3	0.000
24	0.654	0.081	2161.5	0.000
25	0.524	-0.019	2287.7	0.000
26	0.307	0.054	2331.0	0.000
27	-0.021	-0.069	2331.2	0.000
28	-0.294	0.005	2371.2	0.000
29	-0.524	-0.025	2498.6	0.000
30	-0.609	-0.062	2671.1	0.000
31	-0.539	-0.072	2806.5	0.000
32	-0.298	0.040	2848.1	0.000
33	-0.018	-0.014	2848.2	0.000
34	0.311	0.036	2893.8	0.000
35	0.544	0.034	3033.2	0.000
36	0.665	0.163	3242.2	0.000

Figure 2: Correlogram of the original data

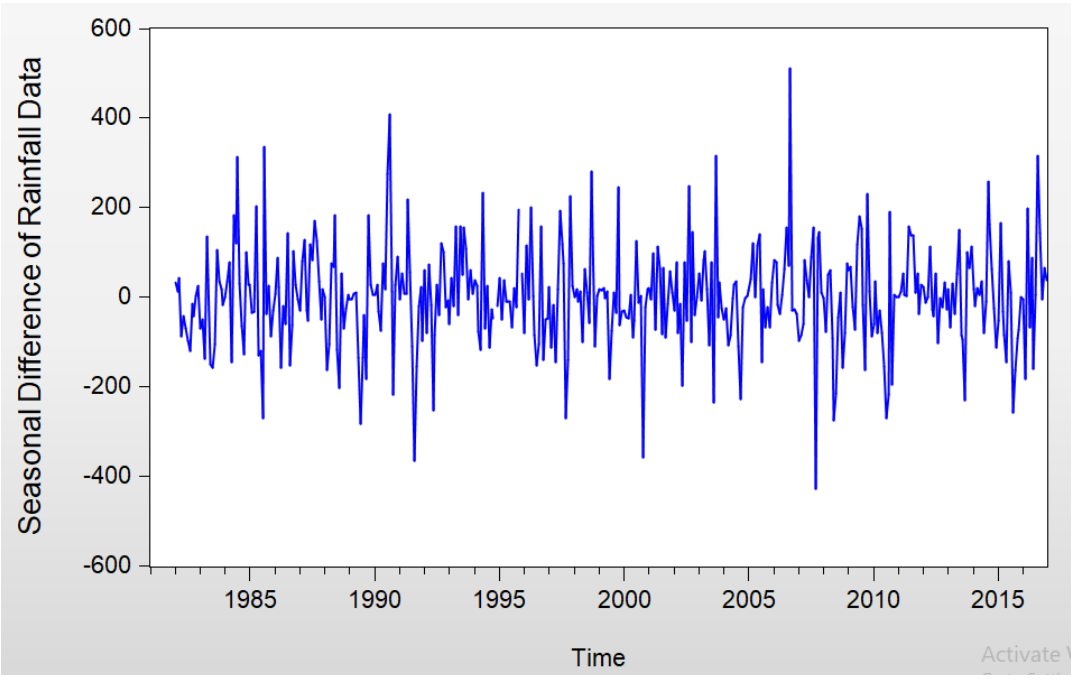


Figure 3: Time Plot of the seasonal difference of the original data

Table 2: Augmented Dickey Fuller Test for the seasonal difference

Null Hypothesis: SDMRRS has a unit root
Exogenous: Constant
Lag Length: 11 (Automatic - based on SIC, maxlag=17)

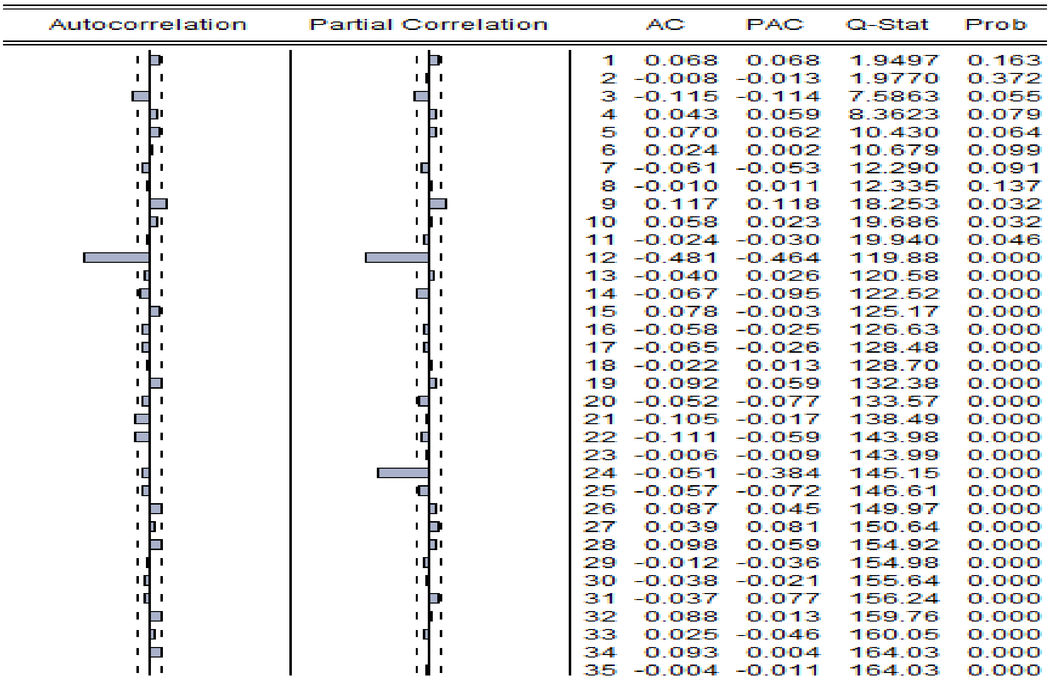
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-8.526865	0.0000
Test critical values:	1% level	-3.447214	
	5% level	-2.868868	
	10% level	-2.570740	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(SDMRRS)
Method: Least Squares
Date: 09/07/20 Time: 11:09
Sample (adjusted): 1983M01 2016M12
Included observations: 383 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SDMRRS(-1)	-1.253184	0.146969	-8.526865	0.0000
D(SDMRRS(-1))	0.302310	0.141999	2.128963	0.0339
D(SDMRRS(-2))	0.326841	0.137104	2.383896	0.0176
D(SDMRRS(-3))	0.284667	0.131701	2.161467	0.0313
D(SDMRRS(-4))	0.315305	0.124209	2.538494	0.0115
D(SDMRRS(-5))	0.353153	0.116228	3.038455	0.0025
D(SDMRRS(-6))	0.384127	0.108228	3.549223	0.0004
D(SDMRRS(-7))	0.358746	0.099261	3.614181	0.0003
D(SDMRRS(-8))	0.378306	0.088435	4.277787	0.0000
D(SDMRRS(-9))	0.453176	0.075204	6.025981	0.0000
D(SDMRRS(-10))	0.491519	0.062616	7.849689	0.0000
D(SDMRRS(-11))	0.490201	0.046233	10.60280	0.0000
C	0.224385	5.153264	0.043542	0.9653

R-squared	0.605770	Mean dependent var	0.037337
Adjusted R-squared	0.592984	S.D. dependent var	157.9331
S.E. of regression	100.7578	Akaike info criterion	12.09667
Sum squared resid	3756292.	Schwarz criterion	12.23068



Lag	AC	PAC	Q-Stat	Prob
1	0.068	0.068	1.9497	0.163
2	-0.008	-0.013	1.9770	0.372
3	-0.115	-0.114	7.5863	0.055
4	0.043	0.059	8.3623	0.079
5	0.070	0.062	10.430	0.064
6	0.024	0.002	10.679	0.099
7	-0.061	-0.053	12.290	0.091
8	-0.010	0.011	12.335	0.137
9	0.117	0.118	18.253	0.032
10	0.058	0.023	19.686	0.032
11	-0.024	-0.030	19.940	0.046
12	-0.481	-0.464	119.88	0.000
13	-0.040	0.026	120.58	0.000
14	-0.067	-0.095	122.52	0.000
15	0.078	-0.003	125.17	0.000
16	-0.058	-0.025	126.63	0.000
17	-0.065	-0.026	128.48	0.000
18	-0.022	0.013	128.70	0.000
19	0.092	0.059	132.38	0.000
20	-0.052	-0.077	133.57	0.000
21	-0.105	-0.017	138.49	0.000
22	-0.111	-0.059	143.98	0.000
23	-0.006	-0.009	143.99	0.000
24	-0.051	-0.384	145.15	0.000
25	-0.057	-0.072	146.61	0.000
26	0.087	0.045	149.97	0.000
27	0.039	0.081	150.64	0.000
28	0.098	0.059	154.92	0.000
29	-0.012	-0.036	154.98	0.000
30	-0.038	-0.021	155.64	0.000
31	-0.037	0.077	156.24	0.000
32	0.088	0.013	159.76	0.000
33	0.025	-0.046	160.05	0.000
34	0.093	0.004	164.03	0.000
35	-0.004	-0.011	164.03	0.000

Figure 4: Correlogram of the seasonal difference

Table 3: Statistics of two suggested models

SARIMA Model	AIC	R-Squared
SARIMA(0,0.1)(1,1,1)$_{12}$	11.71313	0.498724
SARIMA(1,0,1)(0, 1,1)$_{12}$	**11.71225**	**0.512513**
SARIMA(0,0,1)(2,1,1)$_{12}$	11.71571	0.496031

International Journal of Science and Advanced Innovative Research
ISSN: 2536-7315 (Print) 2536-7323 (Online)
Volume 5, Number 3, September 2020
http://www.casirmediapublishing.com

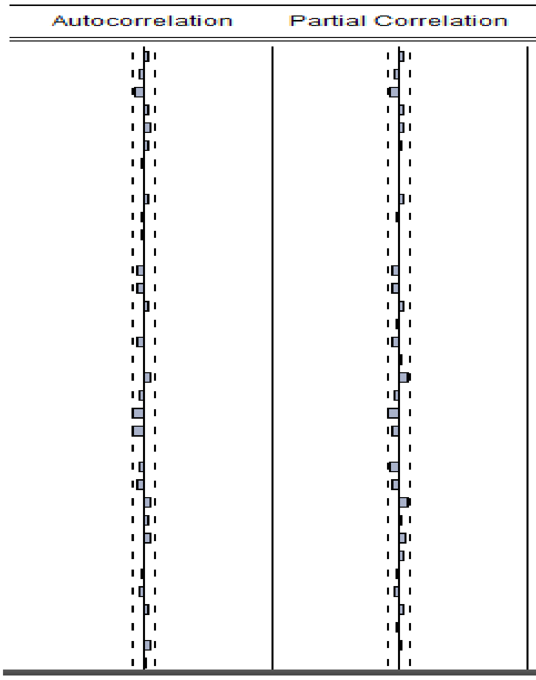
Table 4: Estimation of the SARIMA(1,0,1)x(0,1,1)$_{12}$ model

Dependent Variable: SDMRRS
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 09/08/20 Time: 05:32
Sample: 1982M01 2016M12
Included observations: 418
Failure to improve objective (non-zero gradients) after 22 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.881035	0.137660	-6.400094	0.0000
AR(12)	-0.031717	0.039994	-0.793054	0.4282
MA(1)	0.889036	0.132649	6.702186	0.0000
MA(12)	-0.986653	0.047710	-20.68014	0.0000
MA(13)	-0.902382	0.136847	-6.594117	0.0000
SIGMASQ	6272.387	1325.353	4.732616	0.0000

R-squared	0.512513	Mean dependent var	-0.227512
Adjusted R-squared	0.506596	S.D. dependent var	113.5677
S.E. of regression	79.77300	Akaike info criterion	11.71225
Sum squared resid	2621858.	Schwarz criterion	11.77017
Log likelihood	-2441.860	Hannan-Quinn criter.	11.73515
Durbin-Watson stat	1.944933		

Inverted AR Roots	.68+.19i	.68-.19i	.48-.52i	.48+.52i
	.14+.70i	.14-.70i	-.25-.69i	-.25+.69i
	-.61+.48i	-.61-.48i	-.88-.12i	-.88+.12i
Inverted MA Roots	1.00	.87-.50i	.87+.50i	.50+.87i
	.50-.87i	.00+1.00i	.00-1.00i	-.50+.87i
	-.50-.87i	-.86+.50i	-.86-.50i	-.94
	-.97			



	AC	PAC	Q-Stat	Prob
1	0.027	0.027	0.3101	0.578
2	-0.039	-0.040	0.9542	0.621
3	-0.075	-0.073	3.3127	0.346
4	0.035	0.037	3.8217	0.431
5	0.042	0.034	4.5581	0.472
6	0.029	0.024	4.9126	0.555
7	-0.011	-0.005	4.9682	0.664
8	-0.003	0.004	4.9712	0.761
9	0.039	0.040	5.6392	0.775
10	-0.011	-0.018	5.6906	0.841
11	-0.009	-0.007	5.7276	0.891
12	-0.002	0.003	5.7302	0.929
13	-0.056	-0.062	7.1136	0.896
14	-0.050	-0.051	8.2057	0.878
15	0.035	0.034	8.7434	0.891
16	0.004	-0.008	8.7489	0.923
17	-0.059	-0.061	10.295	0.891
18	-0.004	0.010	10.304	0.922
19	0.064	0.068	12.132	0.880
20	-0.033	-0.046	12.602	0.894
21	-0.093	-0.089	16.461	0.743
22	-0.084	-0.064	19.604	0.608
23	0.002	-0.004	19.606	0.666
24	-0.040	-0.070	20.316	0.679
25	-0.056	-0.064	21.729	0.651
26	0.054	0.075	23.044	0.630
27	0.033	0.023	23.539	0.656
28	0.056	0.051	24.956	0.630
29	0.006	0.034	24.971	0.680
30	-0.021	-0.009	25.175	0.716
31	-0.037	-0.039	25.789	0.731
32	0.036	0.035	26.374	0.747
33	-0.005	-0.011	26.386	0.786
34	0.045	0.020	27.328	0.784
35	0.024	0.005	27.581	0.810

Figure 5: Correlogram of the residuals of the SARIMA(1,0,1)X(0,1,1)$_{12}$ model

Table 5: 2017 forecasts

Time	Forecast
January 2017	18.4
February 2017	60.9
March 2017	112.0
April 2017	162.7
May 2017	249.2
June 2017	305.7
July 2017	332.4
August 2017	330.8
September 2017	355.9
October 2017	256.2
November 2017	82.3
December 2017	30.9
January 2018	17.5

CONCLUSION

The model (4) gives an adequate representation of the data. By it forecasts have been made for the year 2017 and January 2018. Planning may be done on its basis by agriculturists, tourists, planners and administrators. We recommend that further research could focus on investigating the effectiveness of the fitted SARIMA model by comparing it to weather conditions of other states which were not considered in this work.

REFERENCES

Bari, S. H., Hussain, R. and Sourov, M. M. "Forecasting Monthly Precipitation in Sylhet City Using Arima Model." *Civil and Environment Research 7(1), 78-82 (2015).*

Box, G. E. P. and Jenkins, G. M. "Models For Forecasting Seasonal and Non-Seasonal Time Series in Spectral Analysis Of Time Series. *New York Wiley (1976).*

Box, G. E. P. and Jenkins, G.M. "Time Series Analysis Forecasting and Control." *New Jersey: Prentice Hall International Inc. (2004).*

Chonge, M., Nyongesa, K., Omukaba, M,. Makokha, L. and Tireito, F. "Time Series Model Of Rainfall Pattern of Uasin Gishu County. "*Journal of Mathematics, 11(5), 77-84 (2015).*

Cowden, R. J., Watnika, D. W. and Milhucic, R. J. "Stochastic Rainfall Modelling in West Africa: Parsimonious Approaches For Domestic Harvesting Assessment." *Journal of Hydrology, 36(1), 64-77 (2010).*

Farajzadeh, J., Fard, A. F. and Loffi, S. "Modeling Rainfall and Runoff of Urima Lake Basin Using Fee-forward National Network and Time Series". *Water Resources Industry, 7, 138-148 (2012).*

Inderjeet, K. and Sabita, M. S. "Seasonal Arima Model For Forecasting on Monthly Rainfall and Temerature." *Journal of Environment Reserch and Development, 3(2), 52-59.*

Jabrin, A. S., Abubakar, G. M., Kargi, I. A., Musa, Y. and Gulumbe, S. U. "Modelling and Forecsting Rainfall Pattern in Kano State, Northwest of Nigeria." *Journal of Pure and Applied Sciences, 64, 879-886 (2014).*

Ogunrinade, E. D. "Time Series Analysis of Rainfall South-West Part of Nigeria." *Journal of Pure and Applied Sciences, 22(2, 70-75.*