Original Research Article

SARIMA MODELLING OF THE FREQUENCY OF MONTHLY RAINFALL IN OSUN STATE, NIGERIA

Abstract

A Seasonal Autoregressive Integrated Moving Average (SARIMA) is proposed for Osun State monthly rainfall data, the analysis was based on probability time series modeling approach. The Seasonal Autoregressive Integrated Moving Average (SARIMA) model was estimated and the best fitted SARIMA model was used to obtain the rainfall pattern. The Plot of the original data shows that the time series is stationary and the Augmented Dickey-Fuller test did not suggest otherwise. The graph further displays evidence of seasonality and it was removed by seasonal differencing. The plots of the ACF and PACF show spikes at seasonal lags respectively, suggesting SARIMA (1, 0, 1) (2, 1, 1). Though the diagnostic check on the model favoured the fitted model, the Auto Regressive parameter was found to be statistically insignificant and this led to a reduced SARIMA (1, 0, 1) (1, 1, 1) model that best fit the data and was used to make forecast.

Keywords: Rainfall, Seasonality, Stationarity, SARIMA, Time Series

1.0 Introduction

The highly variable nature of rainfall as compared with the relatively stable nature of the temperature appears to have imbued more relevance to the former as the major component in the study of climate in a particular region. Then, there is need to understand the dynamical processes that determine changes that occur in climate system, though this has been very difficult and challenging to climate scientists till today[2].

The change has significantly contributed to the increase of global disasters caused by weather, climate and water related hazards as both developed and developing countries of the world are bearing the burden of repeated floods, temperature extremes and storms in which Nigeria is not left out. Water resources are essential renewable resources that are the basis for existence and development of a society. Proper utilization of these resources requires assessment and management of the quantity and quality of the water resources both partially and temporally. Water crises caused by shortages, floods and diminishing water quality, among others are increasing in all parts of the world. The growth of population demands for increased domestic water supplies and are the same time results with a higher consumption of water due to expansion in agriculture and industry[1]. Mismanagement and lack of knowledge about existing water resources and the changing climatic conditions have consequences of an imbalance of supply and demand of water.

2.0 Method

Rainfall data are time structured and time series analyses are often employed in the analysis of the data. The data were subjected to seasonal autoregressive integrated moving average (SARIMA). Modeling [3].

2.1 Seasonal Autoregressive Integrated Moving Average (SARIMA)

A time series is defined as a set of data collected sequentially in time. The measurements taking during an event in a time series are arranged in a proper chronological order. A time series contain of a single variable is termed as univariate. But if records of more than one variables are considered, its termed as multivariate.

2.2 Definitions

An ARIMA model is an algebraic statement that describes how a time series is statistically related to its own past. The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. One shorthand notation for the model is

$\text{ARIMA}\ (p, d, q) \times (P, D, Q)_S$,

With p = non-seasonal AR order, d = non-seasonal differencing, q = non-seasonal MA order, P = seasonal AR order, D = seasonal differencing, Q = seasonal MA order, and S = time span of repeating seasonal pattern.

Without differencing operations, the model could be written more formally as

$\Phi(B^S)\varphi(B)(x_t - \mu) = \Theta(B^S)\theta(B)w_t$

The non-seasonal components are:

AR: $\varphi(B) = 1 - \varphi_1 B - \ldots - \varphi_p B^p$

MA: $\theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q$

The seasonal components are:

Seasonal AR: $\Phi(B^S) = 1 - \Phi_1 B^S - \ldots - \Phi_P B^{PS}$

Seasonal MA: $\Theta(B^S) = 1 + \Theta_1 B^S + \ldots + \Theta_Q B^{QS}$

Note that on the left side of equation (1) the seasonal and non-seasonal AR components multiply each other, and on the right side of equation (1) the seasonal and non-seasonal MA components multiply each other[4].

2.3 Steps to SARIMA Modeling

The SARIMA modeling approach is concerned with finding a parsimonious seasonal ARIMA model that describes the underlying the generating processed of the observed time series. Box and Jenkins[5] established a three step modeling procedure: identification, estimation and diagnostic checking steps. The identification step is to tentatively choose one or more ARIMA/SARIMA model(s) using the estimated ACF and PACF plots. The ACF plot of the AR (Auto Regressive)/ SAR (Seasonal Auto Regressive) process shows an exponential decay while its PACF plot truncates at lag p/seasonal lag p and diminishes to zero afterwards. The ACF plot of the MA process truncates to zero after lag q/ seasonal lag q while its PACF decays exponentially to zero. The two processes: AR (p)/SAR(P) and MA (q)/SMA(Q), could be combined to form the ARMA (p, q)/SARMA (P, Q) process which has ACF and PACF that decays exponentially to zero. The maximum likelihood estimation method could be used in to estimate the parameters of the identified model(s) in the identification stage. The last diagnostic checking stage involves assessing the adequacy of the identified and fitted models through possible statistically significant test on the residuals to verify its consistency with the white noise process e.g. the Ljung-Box test[6]. Finally, the best fitting model would be selected among other satisfactory, competing models e.g. the information criteria statistics on the basis of the AIC[7] and BIC[8] rule of thumb (Models with the smallest information criterion is the best) and forecast is made with the model of best fit.

3.0 Results and Discussion

The data collected is a secondary data of “Osun state monthly rainfall” obtained from the National Beaurea of Statistics, Abuja from year 1981-2015. The behavior of the data was observed, after which the model was used to describe and forecast the data. The estimation of the expected models was carried out using the method of likelihood, using R software. Considering the plots of the ACF and PACF of the difference and non-difference series, From table 1 below, SARIMA $(1,0,1)\times(1,1,1)_{12}$ proved to be appropriate model with minimum Akaike information criterion (AIC) of 4721.14. This statistics provides an estimate of the proportion of the total variation in the series that is explained by the model.

3.1 IDENTIFICATION OF THE MODEL

In the process of identification, the aim is to identify the possible seasonal ARIMA model that describes the data at hand. We have already induced stationary in the series as shown in the figure 1, from that, we proceeds to obtain the Autocorrelation function ACF and the Partial Autocorrelation Function PACF. The results are shown in table 1 below. After several iterations, some models were suggested among which are presented in the table below: SARIMA$(1,0,1)(1,1,1)_{12}$, $(1,0,2)(1,1,1)_{12}$, $(1,0,1)(2,1,1)_{12}$, $(102)(1,1,2)_{12}$, $(2,0,1)(2,1,1)(1,2,1)_{12}$, $(1,0,1)(1,1,2)_{12}$ as presented in the table below;

Table 1: Summary of The Estimate of the Candidate components

CANDIDATE MODEL	COEFFICIENT	S.E	$Sigma^2$	LOG LIKELIHOOD	AIC
AR1 AR2 MA1 SAR1 SAR2 SMA1	1.1418 -0.1553 -1.0000 0.1062 0.0640 -0.9292	0.0200 0.0213 0.0131 0.0233 0.0231 0.0322	5697	-2354.59	4723.17
AR1 MA1 SAR1 SMA1	0.1897 -0.0374 0.0893 -0.9121	0.3071 0.3127 0.0593 0.0388	5775	-2355.57	4721.14
AR1 MA1 MA2 SAR1 SMA1	-0.3062 0.4603 0.0794 0.0910 -0.9135	0.8815 0.8803 0.1361 0.0594 0.0387	5771	-2355.5	4723.01
AR1 AR2 MA1 SAR1 SMA1 SMA2	0.0645 0.0176 0.0868 0.5672 -1.3955 0.4279	0.3733 0.3882 0.3992 0.3728 0.3971 0.3571	5738	-2355.1	4724.20
AR1 MA1 SAR1 SAR2 SMA1	0.2082 -0.0586 0.1020 0.0618 -0.9317	0.3220 0.3293 0.0591 0.0571 0.0420	5742	-2354.99	4721.97

AR1 MA1 SAR1 SMA1 SMA2	0.1698 -0.0181 0.5540 -1.3815 0.4152	0.3141 0.3184 0.3822 0.4055 0.3643	5739	-2355.1	4722.2

Comparing the SARIMA$(1,0,1)(1,1,1)_{12}$, $(1,0,2)(1,1,1)_{12}$, $(1,0,1)(2,1,1)_{12}$, $(102)(1,1,2)_{12}$, $(2,0,1)(2,1,1)(1,2,1)_{12}$, $(1,0,1)(1,1,2)_{12}$ models above in terms of the AIC, STD Error, log likelihood, square sigma estimated and coefficient respectively, clearly prefer SARIMA$(1,0,1)(1,1,1)_{12}$ model since It has smallest AIC.

Figure 1 : Frequency of Monthly Rainfall in Osun State

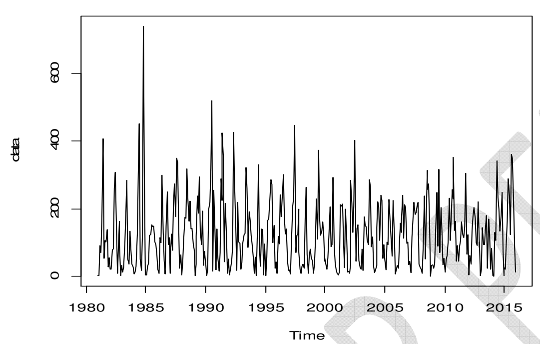


Figure 2: ACF Plot of the Frequency of Monthly Rainfall in Osun State

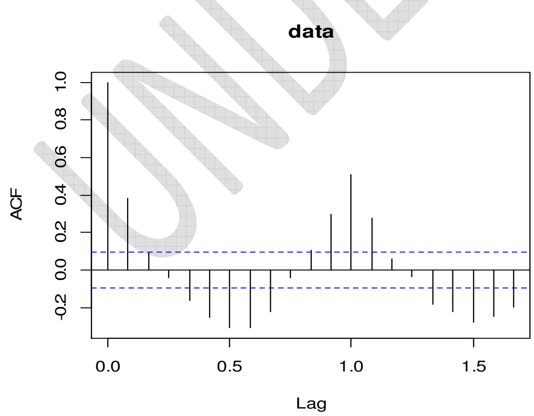
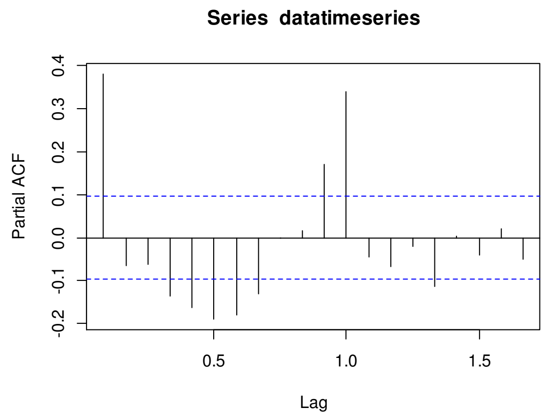


Figure 3: PACF Plot of the Frequency of Monthly Rainfall in Osun State



From the plots in Figure1 it could be seen that the time series plot displays a wave like pattern an evidence of seasonality and no trend is observed which implies that the time series might be stationary. The sinusoidal or periodic pattern in the ACF plot is again suggesting that the series has a strong seasonal effect also, the PACF plot is neither suggesting otherwise. In order to verify the stationarity claim of the visual displays, the Augmented Dickey-Fuller[10] test was performed

Table 2: Unit Root and Stationarity tests of Osun State Monthly Rainfall

Test	Test Statistics	Lag Order	p-value
Dickey-Fuller	-13.626	0	0.01

Table 2 above depicts the Augmented Dickey-Fuller Test, the hypothesis;

H_0: the series is unit root non stationary

Vs

H_1: the series is unit root stationary

The decisions involved rejecting H_0 if the p-value is less than the significance level of 0.05 and accepting H_0 if otherwise.

Small p-value of 0.01 less than 0.05 is in favour of the alternative hypothesis. Thus, strong evidence against the null hypothesis at 5% level of significance.

In order to eliminate the seasonal effect from the time series i subjected the data to a seasonal differencing and the data is re-examined visually.

Figure 4: Plot of diff (1) of Monthly Rainfall in Osun State

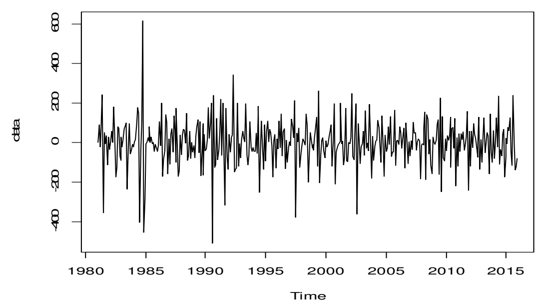


Figure 5: ACF Plot of diff (1) of the Frequency of Monthly Rainfall in Osun State

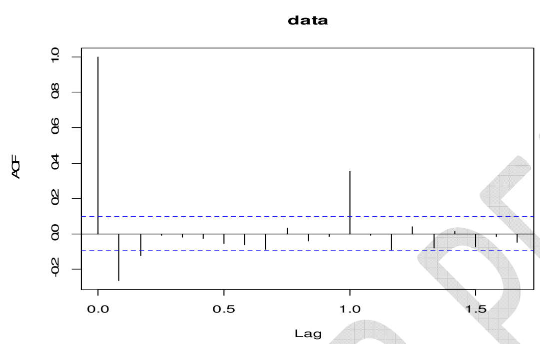


Figure 6: PACF Plot of diff (1) of the Frequency of Monthly Rainfall in Osun State

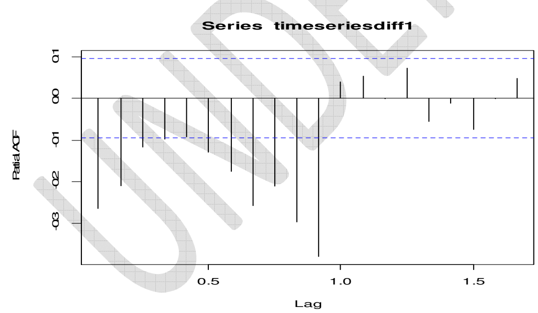


Table 3: Unit Root and Stationarity tests of Osun State Monthly Rainfall

Test	Test Statistics	Lag Order	p-value
Dickey-Fuller	-12.085	1	0.032

Table 3 above depicts the Augmented Dickey-Fuller Test, the hypothesis;

H_0: the series is unit root non stationary

Vs

H_1: the series is unit root stationary

The decisions involved rejecting H_0 if the p-value is less than the significance level of 0.05 and accepting H_0 if otherwise.

4.7 FORECASTING WITH THE FITTED MODEL

One of the objectives of fitting SARIMA model to data is to be able to forecast its future values. The model that best fits the data is SARIMA $(1,0,1)\times(1,1,1)_{12}$. Consider the general SARIMA case. The fitted model is therefore usd to forecast for 16 years, from 2016-2030.

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan 2016	15.36211	-82.25182	112.9768	-133.925716	164.6507
Feb 2016	30.07109	-68.75338	128.8956	-121.067900	181.2101
Mar 2016	87.03243	-11.83149	185.8963	-64.166884	238.2317
Apr 2016	135.47231	36.60574	234.3389	-15.731056	286.6757
May 2016	217.30619	118.43930	316.1731	66.102330	368.5101
Jun 2016	221.34064	122.47370	320.2076	70.136708	372.5446
Jul 2016	180.92120	82.05425	279.7881	29.717249	332.1251
Aug 2016	156.42678	57.55984	255.2937	5.222836	307.6307
Sep 2016	215.29227	116.42532	314.1592	64.088317	366.4962
Oct 2016	202.80709	103.94012	301.6741	51.603108	354.0111
Nov 2016	68.80167	-30.06542	167.6688	-82.402493	220.0058
Dec 2016	53.45305	-45.41474	152.3208	-97.752177	204.6583
Jan 2017	22.87171	-77.59039	123.3338	-130.771816	176.5152

163	Feb 2017	33.48983	-67.02443	134.0041	-120.233470	187.2131
164	Mar 2017	84.42676	-16.09105	184.9446	-69.301961	238.1555
165	Apr 2017	136.37262	35.85437	236.8909	-17.356773	290.1020
166	May 2017	207.18579	106.66748	307.7041	53.456301	360.9153
167	Jun 2017	220.21331	119.69499	320.7316	66.483810	373.9428
168	Jul 2017	182.39962	81.88129	282.9179	28.670109	336.1291
169	Aug 2017	149.68948	49.17116	250.2078	-4.040030	303.4190
170	Sep 2017	212.07930	111.56098	312.5976	58.349791	365.8088
171	Oct 2017	199.83608	99.31773	300.3544	46.106535	353.5656
172	Nov 2017	67.77333	-32.74517	168.2918	-85.956453	221.5031
173	Dec 2017	58.44459	-42.07480	158.9640	-95.286561	212.1758
174	Jan 2018	23.24208	-78.49374	124.9779	-132.349433	178.8336
175	Feb 2018	34.57036	-67.20853	136.3493	-121.087027	190.2277
176	Mar 2018	83.61879	-18.16338	185.4010	-72.043613	239.2812
177	Apr 2018	135.06850	33.28591	236.8511	-20.594546	290.7316
178	May 2018	202.12505	100.34239	303.9077	46.461902	357.7882
179	Jun 2018	219.42295	117.64028	321.2056	63.759786	375.0861
180	Jul 2018	186.30029	84.51762	288.0830	30.637129	341.9635
181	Aug 2018	136.99492	35.21224	238.7776	-18.668248	292.6581
182	Sep 2018	203.84346	102.06079	305.6261	48.180291	359.5066
183	Oct 2018	199.05576	97.27305	300.8385	43.392541	354.7190
184	Nov 2018	66.53447	-35.24841	168.3174	-89.129019	222.1980
185	Dec 2018	61.65131	-40.13266	163.4353	-94.013840	217.3165

186	Jan 2019	23.94533	-78.40262	126.2933	-132.582346	180.4730
187	Feb 2019	35.09976	-67.27187	137.4714	-121.464135	191.6637
188	Mar 2019	83.59608	-18.77772	185.9699	-72.971136	240.1633
189	Apr 2019	135.20616	32.83206	237.5803	-21.361512	291.7738
190	May 2019	201.22227	98.84812	303.5964	44.654520	357.7900
191	Jun 2019	219.49075	117.11659	321.8649	62.922989	376.0585
192	Jul 2019	186.99776	84.62360	289.3719	30.430001	343.5655
193	Aug 2019	135.52501	33.15085	237.8992	-21.042751	292.0928
194	Sep 2019	203.03491	100.66076	305.4091	46.467149	359.6027
195	Oct 2019	199.01391	96.63972	301.3881	42.446099	355.5817
196	Nov 2019	66.56296	-35.81142	168.9374	-90.005152	223.1311
197	Dec 2019	62.48961	-39.88596	164.8652	-94.080322	219.0595
198	Jan 2020	24.25371	-78.62394	127.1314	-133.084081	181.5915
199	Feb 2020	35.43322	-67.46663	138.3331	-121.938523	192.8050
200	Mar 2020	83.76050	-19.14149	186.6625	-73.614511	241.1355
201	Apr 2020	135.35698	32.45470	238.2593	-22.018483	292.7324
202	May 2020	201.04265	98.14032	303.9450	43.667115	358.4182
203	Jun 2020	219.66535	116.76301	322.5677	62.289804	377.0409
204	Jul 2020	187.51743	84.61510	290.4198	30.141889	344.8930
205	Aug 2020	134.83001	31.92767	237.7323	-22.545539	292.2056
206	Sep 2020	202.67424	99.77190	305.5766	45.298687	360.0498
207	Oct 2020	199.17805	96.27568	302.0804	41.802451	356.5537
208	Nov 2020	66.70668	-36.19591	169.6093	-90.669246	224.0826

Dec 2020	62.98184	-39.92203	165.8857	-94.396044	220.3597

In the process of identification, the aim is to identify the possible seasonal ARIMA model that describes the data at hand. I have already induced stationary in the series as shown in the figure 7 from that, i proceed to obtain the Autocorrelation function ACF and the Partial Autocorrelation Function PACF. After several iterations, some models were suggested.

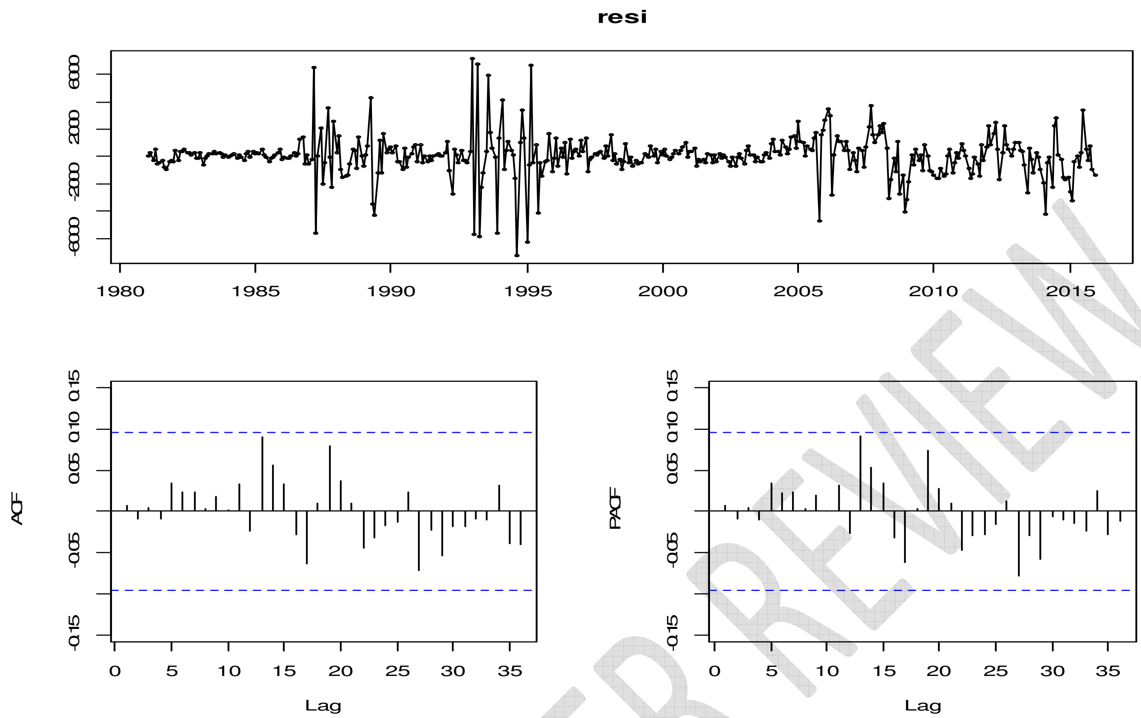
The estimation of the expected models was carried out using the method of likelihood, using R software. Considering the plots of the ACF and PACF of the difference and non difference series, from figure 3 and 4 above, SARIMA (1, 0, 1) × $(1, 1, 1)_{12}$ proved to be appropriate model with minimum Akaike information criterion (AIC) of 4721.14. This statistics provides an estimate of the proportion of the total variation in the series that is explained by the model.

Table 4 fitted SARIMA (1,0,1)$(1,1,1)_{12}$ model

$Sigma^2$	LOG LIKELIHOOD	AIC
5771	-2355.5	4723.01

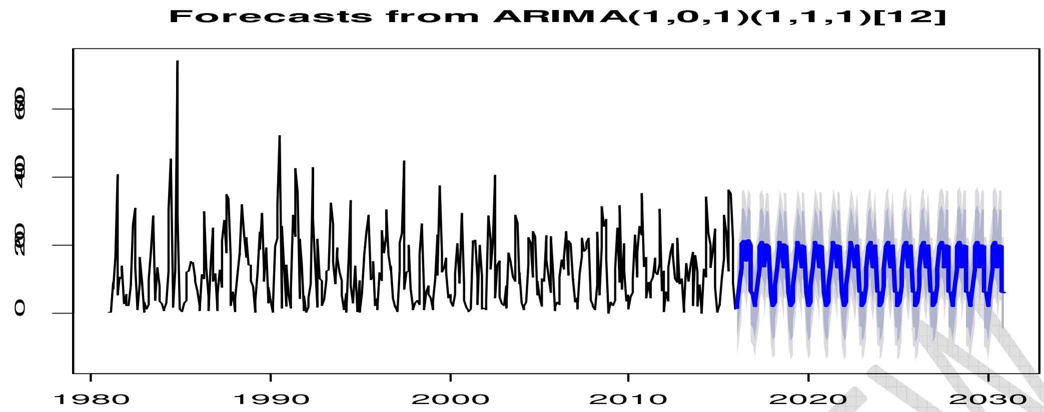
The model validation is concerned with checking the residual of the model to determine if the model contains any systematic pattern which can be removed to improve on the selected model may appear to be the best among a number of models considered it become necessary to do diagnostic checking to verify that the model is adequate.

Figure 7 Model Verification Plot



The plots Figure 7 above comprise of the time plot of the residuals, ACF and the PACF plot of the residuals respectively. The plot clearly shows that the residuals appear to be randomly scattered, no evidence exists that the error terms are correlated with one another as well as no evidence of existence of an outlier. The residuals or errors are therefore conceived of as an independently and identically distributed (i.i.d) sequence with a constant variance and a zero mean. The ACF and the PACF plot of the residuals show no evidence of a significant spike indicating that the residuals seem to be uncorrelated. Therefore, the SARIMA $(1,0,1)(1,1,1)_{12}$ model appears to fit well so we can use this model to make forecasts.

Figure 8 Forecasting Plot



4. CONCLUSION

In this study, the frequency, not the amount of monthly rainfall from 1981 to 2015 obtained from the National Beaurea of Statistics, Abuja is analysed using seasonal time series modeling approach. The plot of the original data shows that the time series is stationary and has evidence of seasonality. The augmented Dickey Fuller test confirmed the stationarity claim. Seasonal differencing was done to remove the seasonal effect. SARIMA modeling of the data was upheld after duly following the conventional three steps of identification. This resulted in obtaining SARIMA $(1,0,1)(2,1,1)_{12}$. However, the seasonal auto regressive parameter was found to be statistically insignificant and this consequently led to a new SARIMA $(1,0,1)(1,1,1)_{12}$ that best fit the data and was used to make forecast.

References

1. IPCC (2011). Intergovernmental Panel on Climate change. Third Assessment Report: Climate change 2001. WG1: The scientific basis, summary for policymakers, Geneva, Switzerland.
2. Adejuwon, J.O., Balogun, E.E. and Adejuwon, S.A. (1990). On the annual and seasonal pattern of rainfall fluctuations in sub-Saharan West Africa. International Journal of Climatology, 10: 839–848.

3. Adejuwon, J.O., Balogun, E.E. and Adejuwon, S.A. (1990). On the annual and seasonal pattern of rainfall fluctuations in sub-Saharan West Africa. International Journal of Climatology, 10: 839–848.
4. Box G.E.P. and Jenkins G.M. Time series Analysis, Forecasting and Control . San Francisco: Holden Day, 1976.
5. J. Lee, "Univariate time series modeling and forecasting (Box-Jenkins Method)", Econ 413, lecture 4.
6. Ljung G. M. and Box G. E. P. On a Measure of Lack of Fit in Time Series Models. Biometrika. - 1978. - Vol. vol. 65. - pp. 297.
7. Akaike H. A New Look at the Statistical Model Identification. Annals of Statistics: IEEE Trans. Automat. Contr., 1974. - Vols. Ac-19. - pp. 716-723.
8. Schwarz G. Estimating the Dimension of a Model. Annals of Statistics. - 1978. - Vol. 6. - pp. 461-464.
9. R Development Core Team R: A language and environment for statistical computing // R Foundation for Statistical Computing. - Vienna, Austria: [s.n.], 2014. - 3-900051-07-0.
10. Dickey D. A. and Fuller W.A. Autoregressive Time Series with a Unit Root. Journal of American Statistical Association. - 1979. - Vol. 74. - pp. 427-431.