FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 7 No. 6, December, 2023, pp 24 - 31
DOI: https://doi.org/10.33003/fjs-2023-0706-2095

ON THE PERFORMANCE OF SARIMA AND SARIMAX MODEL IN FORECASTING MONTHLY AVERAGE RAINFALL IN KOGI STATE, NIGERIA

***Ibrahim A. and Musa A. O.**

Department of Statistics, Confluence University of Science and Technology, Osara Kogi state

*Corresponding authors' email: ibrahima@custech.edu.ng

ABSTRACT

Forecasting monthly rainfall is very important in Kogi state for better approach to flood management and also plays a pivotal role in agriculture which remains a significant factor in Nigeria's economy. Advanced time series univariate models such as Seasonal Autoregressive Integrated Moving Average (SARIMA) models are usually employed in modelling and forecasting rainfall in Nigeria due to their non-linear pattern and spatiotemporal variation. Few studies have attempted to investigate the influence of other climatic factors in modelling and prediction of rainfall pattern. This study examines the performance of a univariate seasonal ARIMA and seasonal ARIMA which uses monthly temperature and relative humidity as exogenous factors otherwise known as SARIMAX model in forecasting monthly average rainfall in Lokoja, the capital of Kogi state. The study uses monthly data on rainfall, temperature and relative humidity spanning from 2010 to 2022 obtained from Nigeria Meteorological Agency NiMet, Lokoja station. The series were appropriately differenced to attain stationarity. The plots of autocorrelation function (ACF) and partial autocorrelation function (PACF) were used to select some tentative models whose parameters would be estimated. The most suitable SARIMA model [SARIMA $(1,0,0) \times (0,1,1)_{12}$] was chosen based on maximum Coefficient of Determination R^2, and the minimum Akaike information criterion (AIC). However, SARIMAX model outperformed SARIMA model based on the criteria earlier highlighted. SARIMAX model was therefore recommended for modelling and forecasting monthly average rainfall in Kogi state.

Keywords: SARIMA, SARIMAX, exogenous, ACF, PACF, Rainfall, Temperature, Humidity

INTRODUCTION

Weather refers to the atmospheric conditions of a particular place at a specific time, which can vary on a daily, weekly, monthly, or yearly basis. Changes in key climate factors such as rainfall, temperature, and humidity can have serious impacts on human lives and crop yields. Excessive or insufficient amounts of these factors can lead to decreased crop yields. Most of the changes in climate pose serious challenges to the planet which usually result in flooding, drought, poor agricultural productivity, variation of ground and surface water (Oruonye, 2014). Climate change is a pressing global issue that has caused severe flooding and drought, resulting in loss of life and property. Researchers around the world are working to understand and mitigate the impacts of climate change, but more work is needed to address this urgent problem. Kogi state which is situated in North central Nigeria is regarded as a confluence state as it is sitting in the confluence of rivers. The state is among the states worst hit by the recent flooding disaster in the country. Several lives were lost in the state to the flood and hundreds of riverine communities including the state capital Lokoja were hugely affected by the flood water. Also, commuters who ply lokoja high way have been seriously affected by the flood. It is therefore highly imperative to critically examine rainfall situation in kogi state by developing an appropriate robust model capable of predicting the future pattern of rainfall in the state. This will enable policy makers to take proactive measures to curtail the devastating effects of excel rainfall. Weather forecasting is the application of scientific knowledge to predict future atmospheric conditions in a given area. Weather forecasters collect and analyze historical weather data to identify patterns and predict how these patterns will evolve in the future. They use mathematical models to make these predictions. The field of weather forecasting has evolved over time, incorporating a range of methods to predict future conditions, from simple ensemble probabilistic models to more complex approaches like the Autoregressive Integrated Moving Average (ARIMA) model. ARIMA models are widely used univariate time series models that decompose data into different components, such as seasonal variations, long-term trends, and residual errors. This allows forecasters to analyze the relationships between various variables and make more accurate predictions. Seasonal autoregressive integrated moving average model shortened as SARIMA models are usually employed when there is obvious seasonal component in the time series data. SARIMAX model is an extension of SARIMA model with exogenous variables when it is imperative to account for the influence of certain external variables.

Previous researches have used univariate time series models to forecast weather patterns, including rainfall and temperature, by only considering past values of the variables being studied. These models were used by researchers such as Samuel and Adam (2020), Okorie et al. (2015), Mahsin *et al.* (2012), Seyid *et al.* (2011), Jibril *et al.* (2019), Emmanuel and Bakari (2015), Peng *et al.* (2018), and Wiredu *et al.* (2013).This study attempts to investigate the forecast performance of SARIMA model and SARIMAX model with temperature and humidity as exogenous variables which by intuition have influence on the response variable (rainfall). The performance of the models will be ascertained using different statistical measures such as AIC, BIC, R^2 and so on.

Data Description

This study made use of secondary data on monthly average rainfall, temperature and relative humidity spanning from 2010 to 2022 obtained from Nigeria Meteorological Agency NiMet, Lokoja station**.** The data were analyzed using Stata and E-views software.

MATERIALS AND METHODS

Autoregressive Integrated Moving Average (ARIMA) Model

A stochastic process X_t is regarded as an ARIMA (p,d,q) if $\Delta^d X_t = (1-B)^d X_t$ is ARMA (p, q). The model is expressed in compact form as

$$\theta(B)(1-B)^d X_t = \vartheta(B)\varepsilon_t \quad (1)$$

ε_t is said to follow a white noise process. The lag operator is as defined below

$$B^k X_t = X_{t-k} \quad (2)$$

We define the autoregressive (AR) and moving average (MA) as

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ . \ . \ . \ \theta_P B^P \quad (3)$$

$$\vartheta(B) = 1 - \vartheta_1 B - \vartheta_2 B^2 - \ . \ . \ . \ \vartheta_q B^q \quad (4)$$

$\theta(B) \neq 0$ **for** $|\theta| < 1$, the process X_t attains stationarity if $d = 0$, hence the process reduces to ARMA (p,q) process.

Seasonal Autoregressive Integrated Moving Average (SARIMA) Model

SARIMA model is used when a time series exhibits a seasonal pattern. In this type of model, the periodic component of the series repeats itself at regular and constant interval. It is therefore an extension of ARIMA model. This model is often denoted as ARIMA (p,d,q)(P,D,Q)s which is expressed in lag form as given below

$$\theta(B)\omega(B)^S(1-B)^d(1-B)^D X_t = \vartheta(B)\psi(B)^S \quad (5)$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ . \ . \ . \ \theta_P B^P \quad (6)$$

$$\omega(B)^S = 1 - \omega_1 B^S - \omega_2 B^{2S} - . \ . \ . \omega_P B^{PS} \quad (7)$$

$$\vartheta(B) = 1 - \vartheta_1 B - \vartheta_2 B^2 - \ . \ . \ . \ \vartheta_q B^q \quad (8)$$

$$\psi(B)^S = 1 - \psi_1 B^S - \psi_2 B^{2S} - . \ . \ . \psi_Q B^{QS} \quad (9)$$

p, d and q denote the AR non-seasonal order, differencing and MA order respectively. P, D and Q are AR seasonal order, differencing and MA seasonal order respectively. denotes time series data at period t (kogi state monthly average rainfall in mm)

ε_t denotes white noise error at period t.

B denotes backward shift operator $B_k\, Xt = X_{t-k}$

S denotes the order of seasonal (s = 12 monthly data)

Sarimax Models

The Seasonal Autoregressive Integrated Moving Average with exogenous factors (SARIMAX) Model.

$$\theta(B)\omega(B)^S(1-B)^d(1-B)^D X_t = \vartheta(B)\psi(B)^S\varepsilon_t + \sum_{i=1}^{n}\beta_i\, y_{it} \quad (10)$$

β_i is the coefficient of the exogenous variables y_i. Other parameters are as previously defined.

Model estimation

Conditional least squares will be used to estimate the parameters of the model. The following information criterion with their respective statistics will be employed to select the parsimonious models.

$$\text{Akaike Information Criterion (AIC)} = nlog\left(\frac{RSS}{n}\right) + 2k \quad (11)$$

$$\text{Corrected Akaike Information Criterion (AICc)} = AIC + \frac{2(K+1)(K+2)}{n-k-2} \quad (12)$$

Bayessian Information Criterion

$$\text{(BIC)} = \{(\hat{\sigma}_e^2)\} + k\{\ln(n)\} \quad (13)$$

RSS is the fitted model estimated residual, n is the sample size of sample residual and k is the total number of estimated parameters in the fitted model while $\hat{\sigma}_e^2$ represents the error variance.

Unit Root Test

The test of unit root is done to know whether the original series is stationary or not augmented Dickey Fuller (ADP) and the Phillip Perron test of unit root with their respective statistics are as given

$$X_t = \alpha_0 + \rho_1 Xt - 1 + \sum_{j=2}^{\rho-1}\beta_j \nabla Xt - 1 + \varepsilon_t \quad (14)$$

$$Z_p = n(\hat{\rho}_n - 1) - \frac{1}{2}\frac{n^2\hat{\sigma}^2}{s_n^2}(\hat{\lambda}_n^2 - \hat{\gamma}_{0,n}) \quad (15)$$

Identification of the model

The sample plot of ACF and PACF will be used to determine the order of the model at both seasonal and non seasonal part. That is (p,d,q) and (P, D,Q)

RESULTS AND DISCUSSION

In this study, the approach and technique of Box Jenkins (1976) was used to compare the performance of SARIMA and SARIMAX model in analyzing and forecasting monthly average volume of rainfall in Lokoja, the kogi state capital using monthly data on Rainfall, Temperature and relative humidity spanning from January 2010 to December, 2022.

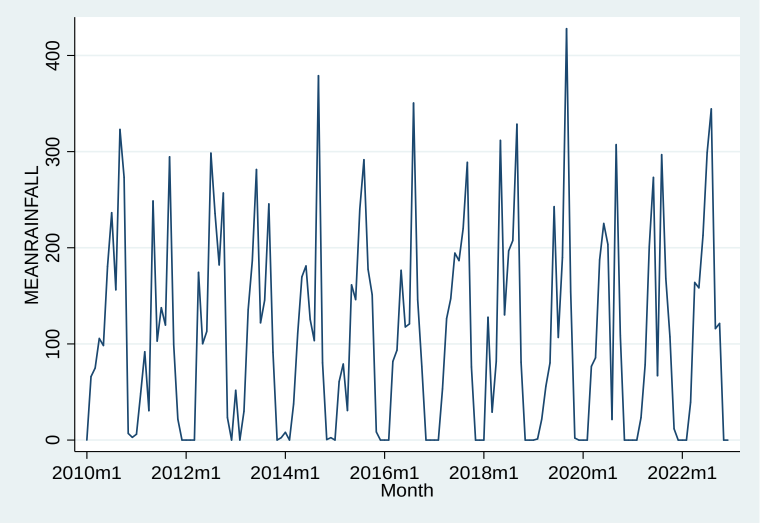

Figure 1: Time Series Plot of Monthly Average Rainfall in Lokoja.

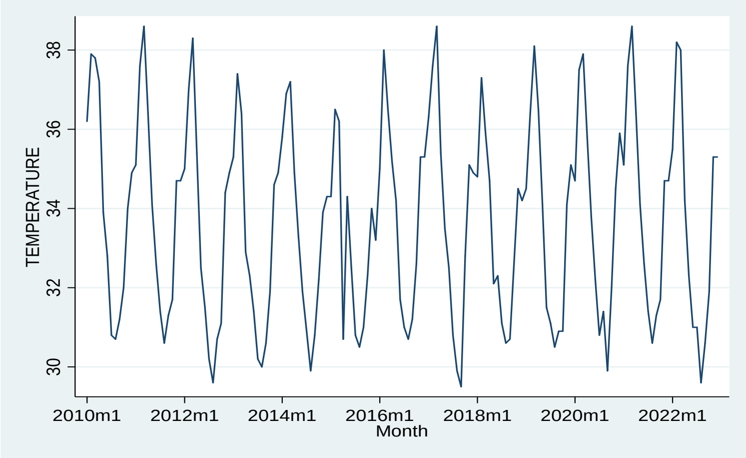

Figure 2: Time series plot of monthly average temperature in Lokoja

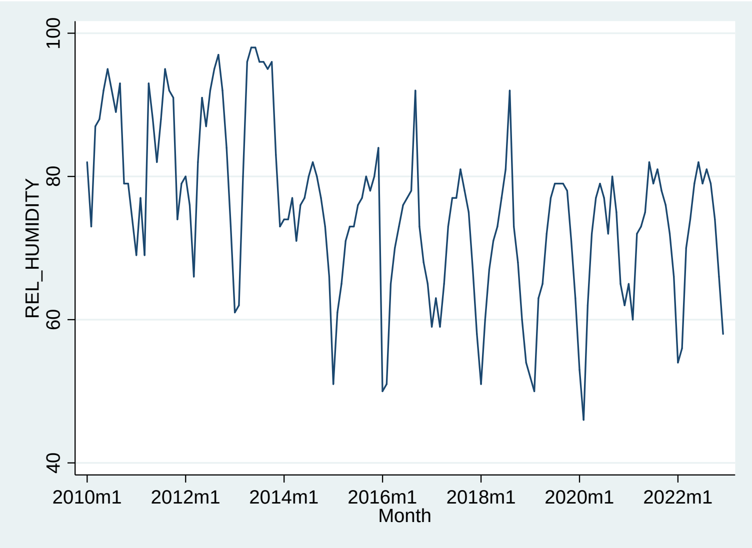

Figure 3: Time series plot of monthly average Relative Humidity in Lokoja

It can be seen in figures 1, the average monthly rainfall does not exhibit trend pattern, but there was evidence of periodic rise and fall. Also figure 2 and 3 display the time plot of original monthly temperature and relative humidity respectively.

Table 1: Dickey Fuller test for unit root of Rainfall series

Test satistic	1% critical value	1% critical value	5 critical value	10% critical value
$Z(t)$	-7.501	-3.492	-2.886	-2.576

Approximate p-value for Z(t) = 0.0000

Table 2: Phillip Perron test for unit root of Rainfall series

Test satistic	1% critical value	1% critical value	5 critical value	10% critical value
$Z(rho)$	-85.812	-19.983	-13.810	-11.073
$Z(t)$	-7.572	-3.492	-2.886	-2.576

Approximate p-value for Z(t) = 0.0000

As revealed by Dickey Fuller unit root test and Phillip Perron test in Table 1 and Table 2 above, the original rainfall series is stationary as the null hypothesis that the series has a unit root is rejected.

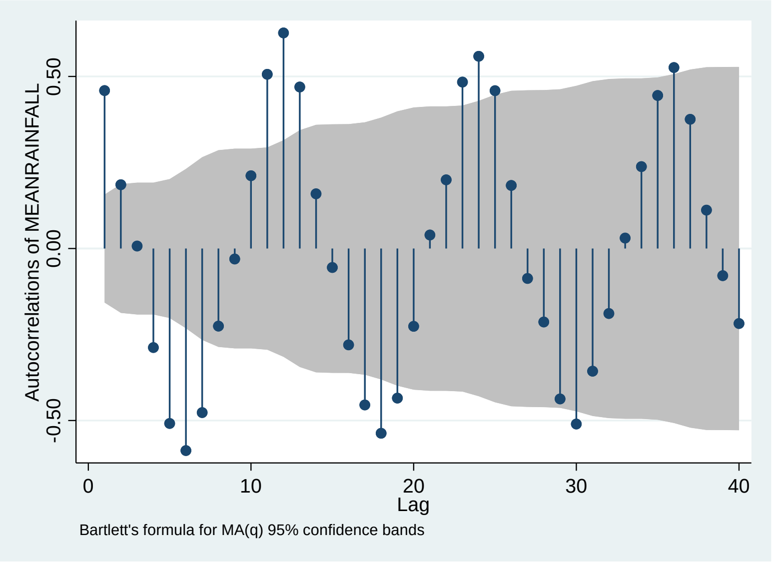

Figure 4: ACF of original rainfall data

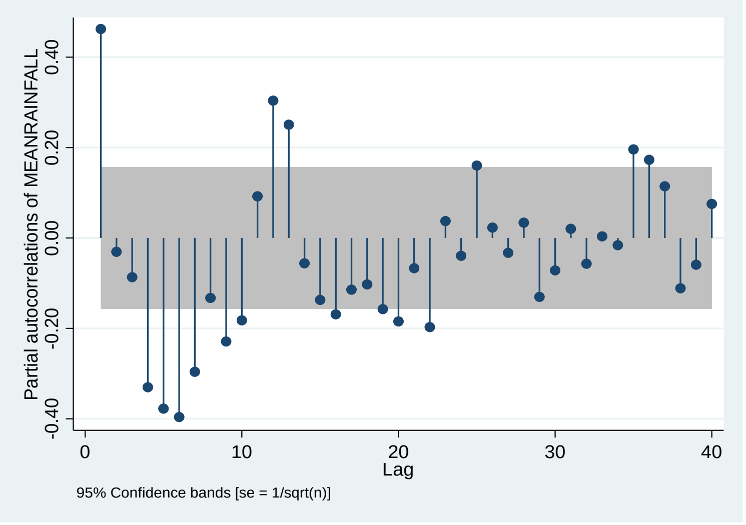

Figure 5: PACF of original rainfall data.

The plot of ACF and PACF of figure 4 and 5 above show a sinusoidal wave pattern which indicates a seasonal pattern. The series was however seasonally differenced and the plot of the differenced series confirms stationarity of the differenced series.

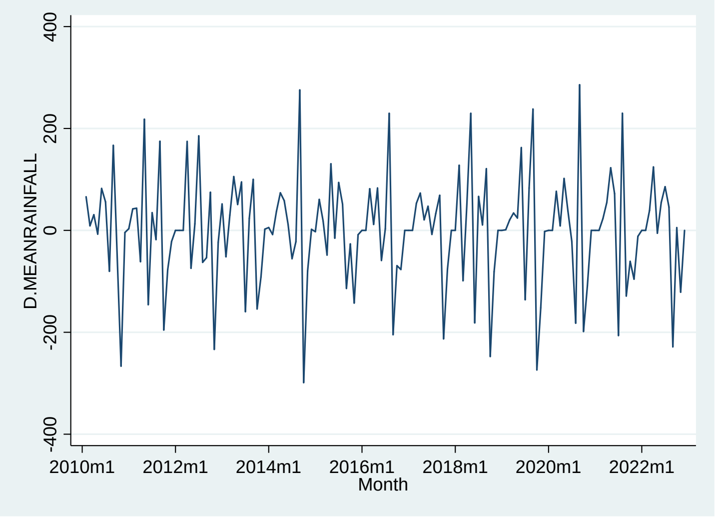

Figure 6: Time plot of the seasonally differenced Rainfall series

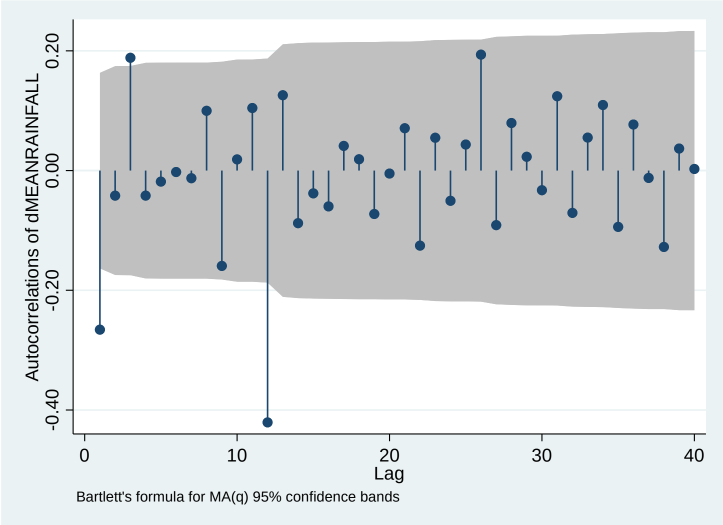

Figure 7: ACF of seasonally differenced rainfall series

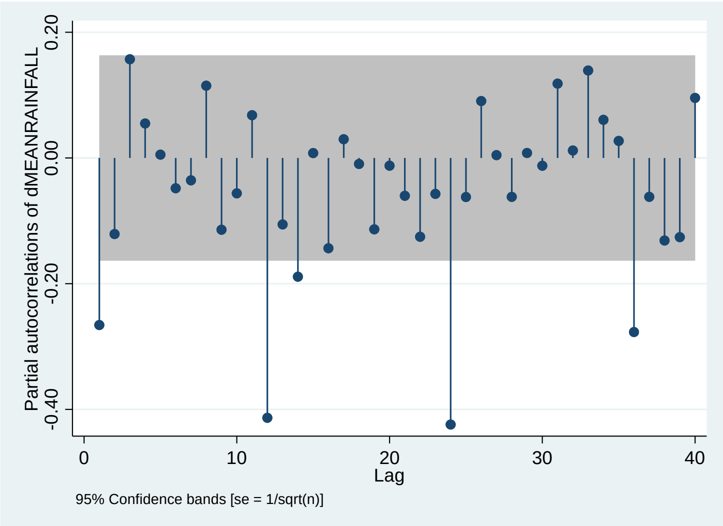

Figure 8: PACF of the differenced rainfall series

Having ascertained that the rainfall series is stationary after taking the first seasonal difference, the ACF of the differenced series can be used to determine the tentative order of non-seasonal and seasonal order of moving average process MA(q) and SMA(Q). PACF is used to determine the non-seasonal and seasonal order of autoregressive process AR(p) and SAR (P). A close inspection of Figure 7 shows a significant spike at lag 1 and lag 3 implying that the order of non seasonal MA components could be up to 3. Also, a significant spike at lag 12 implies SMA(1) is suspected. In Figure 8, a significant spike at lag 1 and significant spikes at lag 12, lag 24, and lag 36 means that AR(1) while seasonal AR process could be up to 3 implying SAR(3) could be enough to appropriately describe the process. Based on the above suspected order of both seasonal and non-seasonal AR and MA, the following combination of autoregressive and moving average process were examined for parameter estimation. SARIMA (1,0,0)(0,1,1), SARIMA (0,0,1)(1,1,1), SARIMA (1,0,0)(1,1,1), SARIMA (1,0,0)(2,1,1), SARIMA(1,0,2)(0,1,1), SARIMA (1,0,2)(3,1,1), SARIMA (1,0,3)(3,1,1). Among the models examined, the models whose parameters are all significant are represented in the table below

Table 3: Model estimation results

Models	AIC	R^2
SARIMA(1,0,0)(0,1,1)$_{12}$	24.229	81.86
SARIMA(1,0,0)(1,1,1)$_{12}$	24.236	81.77
SARIMA(1,0,0)(2,1,1)$_{12}$	24.256	81.72

The best fitted model based on the models with the highest R^2 and the one with the least of AIC is SARIMA(1,0,0)(0,1,1)$_{12}$ Having selected the best fitted SARIMA model, we incorporate the exogenous variables of Temperature and Relative humidity to form SARIMAX model and estimate its parameters. The performance of the SARIMAX model in terms of the criteria used earlier.

Table 4: Estimation of SARIMAX model

Models	AIC	R^2
SARIMA(1,0,0)(0,1,1)$_{12}$	24.229	81.86
SARIMAX(1,0,0)(0,1,1)$_{12}$	24.216	81.92

Table 5: Estimation of parameters of SARIMA(1,0,0)(0,1,1)$_{12}$

Dependent Variable: D(MEANRAINFALL,12)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 06/11/23 Time: 09:43
Sample: 2011M01 2022M12
Included observations: 144
Convergence achieved after 19 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.901914	0.035968	-25.07532	0.0000
MA(12)	-0.093492	0.089047	-5.039918	0.0395
SIGMASQ	1.85E+09	2.55E+08	7.234548	0.0000
R-squared	0.818601	Mean dependent var		453.1653
Adjusted R-squared	0.816028	S.D. dependent var		101276.8
S.E. of regression	43439.60	Akaike info criterion		24.22875
Sum squared resid	2.66E+11	Schwarz criterion		24.29063
Log likelihood	-1741.470	Hannan-Quinn criter.		24.25390
Durbin-Watson stat	3.497855			

Table 6: Estimation of parameters of SARIMAX(1,0,0)(0,1,1)$_{12}$

Dependent Variable: D(MEANRAINFALL,12)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 06/11/23 Time: 10:01
Sample: 2011M01 2022M12
Included observations: 144
Convergence achieved after 10 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
TEMPERATURE	-3.799859	248.3049	-0.015303	0.0478
RELATIVE_HUMIDITY	2.143548	118.1027	0.018150	0.0355
AR(1)	-0.901913	0.035992	-25.05888	0.0000
MA(12)	-0.093514	0.089177	-1.048630	0.0296
SIGMASQ	1.85E+09	2.58E+08	7.158702	0.0000
R-squared	0.819202	Mean dependent var		453.1653
Adjusted R-squared	0.813382	S.D. dependent var		101276.8
S.E. of regression	43750.90	Akaike info criterion		24.21642
Sum squared resid	2.66E+11	Schwarz criterion		24.35965
Log likelihood	-1741.470	Hannan-Quinn criter.		24.29843
Durbin-Watson stat	3.497834			

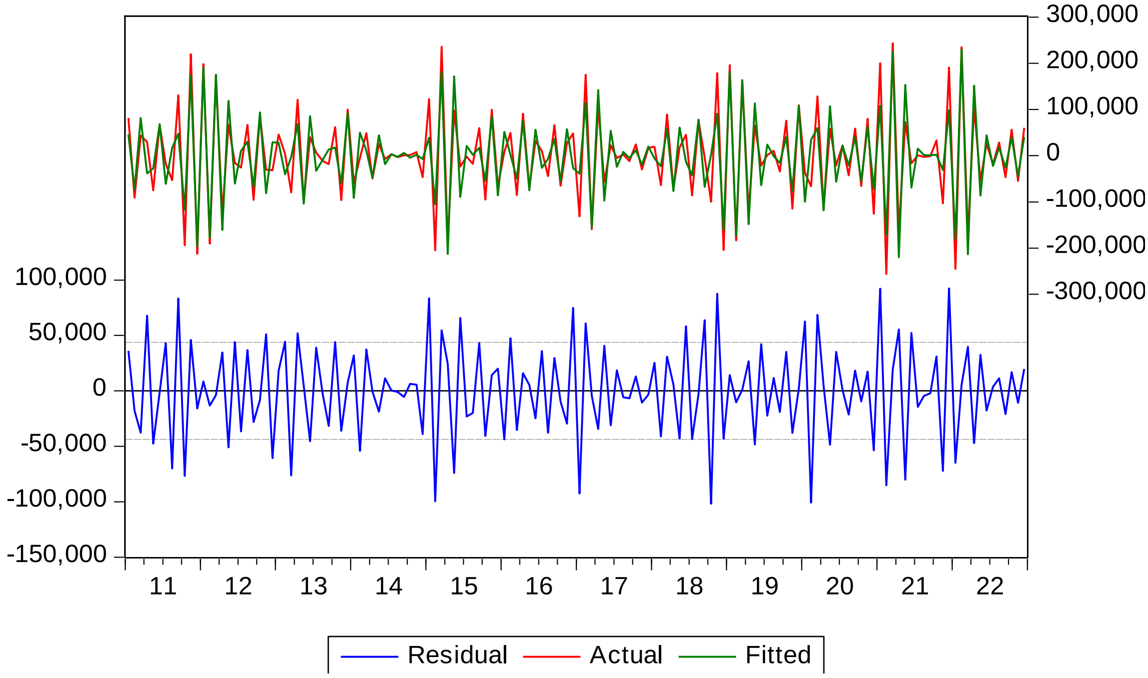

Figure 9: Graph of residual, actual and fitted model

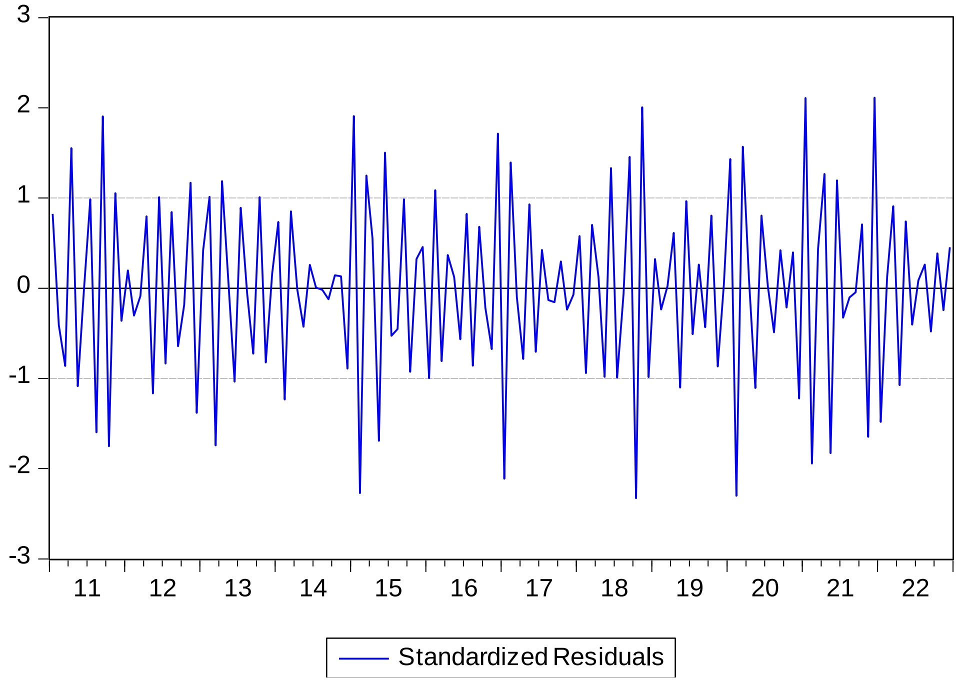

Figure 10: graph of the standardized residuals of the fitted SARIMAX model

SARIMAX model equation for forecasting monthly mean average rainfall

The selected model: SARIMAX(1,0,0)(0,1,1)$_{12}$ is expressed as follows:

$$\theta(B)(1-B)^D X_t = \vartheta(B)\psi(B)^S \varepsilon_t + \sum_{i=1}^{2} \beta_i y_{it} \quad (16)$$

$$X_t = \theta X_{t-1} + X_{t-1} + \theta X_{t-12} + \varepsilon_{t-1} - \psi\varepsilon_{t-12} + \beta_1 y_{1t} + \beta_2 y_{2t} \quad (17)$$

$$\hat{X}_t = -0.90193 X_{t-1} + X_{t-1} - 0.90193 X_{t-12} + \varepsilon_{t-1} + 0.0093514\psi\varepsilon_{t-12} - 3.799859 y_{1t} + 2.143548\, y_{2t} \quad (18)$$

CONCLUSION

The original monthly data on rainfall were confirmed to be stationary by Dickey Fuller unit root test and Phillip Perron unit root test. The plot of ACF and PACF reveal a seasonal pattern in the series which was eliminated by conducting one seasonal differencing. The plot of ACF and PACF were used to select the order of the suspected tentative models. Using the Box and Jenkin methodology, the R^2 and AIC of models with significant parameters were estimated. The model with the least of AIC and highest of R^2 **(** SARIMA(1,0,0)(0,1,1)$_{12}$)was selected as the best fitted SARIMA model. The Exogenous factors (average monthly temperature and relative humidity) were included and the resulting SARIMAX model was estimated. The AIC of the estimated SARIMAX model was smaller and hence adjudged to have performed better than SARIMA model. The goodness of fit of the selected model was confirmed by the standardized residual plot which shows that the residuals are white noise meaning there were uncorrelated. Hence, SARIMAX model with the model equation given in (18) can be used for forecasting monthly average rainfall in Lokoja.

REFERENCES

Box, G.E.P & Jenkins, G.M. (1976). Time Series Analysis: Forecasting and Control.Holden-Day. San Francisco. 4:435-478

Emmanuel, S.U and Bakari, H. R. (2015). An Application of time series Analysis in modeling monthly Rainfall data for Maiduguri, North eastern Nigeria. *Journal of Mathematical theory and modeling*. 5(11):24-33.

Jibril Y. K., kamaluddin A., & Muhammed M. M. (2019). Forecasting Performance of Arima and Sarima Models on Monthly Average Temperature of Zaria, Nigeria. *Journal Of Sciecnce Technology And Education* **7**(3).

Okorie I. E and Akpanta A. C. (2015). Threshold Excess Analysis of Ikeja Monthly Rainfall in Nigeria: *International Journal of Statistics and Applications.* 1(5): 15 – 20.

Oruonye E. D. (2014). An Assessment of the Trends of Climatic Variables in Taraba State Nigeria. *Global Journal of Science Frontier Research*, *14*(H4), 1–14.

Mahsin M.D, Yesmin A & Monira B. (2012). Modeling Rainfall in Dhaka Division of Bangladesh Using Time Series Analysis. *Journal of Mathematical Modelling and Application* 1: 67-73.

Peng C., Aichen N., Duanyang L., Wei J. & Bin M. (2018). Time Series Forecasting of Temperatures using SARIMA: An Example from Nanjing. IOP Conference Series: *Materials Science and Engineering* 394 052024

Samuel O. A. & Adam M. B. (2020) Modelling and Forecasting Seasonal Behavior of Rainfall in Abuja, Nigeria; A SARIMA Approach. *American Journal of Mathematics and Statistics*, 10(1): 10-19.

Seyed A, Shamsnia M, Naeem S, Ali L (2011). Modelling weather parameter using stochastic methods (ARIMA Model) (Case Study: Abadeh station, Iran). International conference on Environment and Industrial innovation IPCBEE.

Wiredu S. , Nasiru S. & Asamoah Y. G. (2013). Proposed Seasonal Autoregressive Integrated moving Average Model for Forecasting Rainfall Pattern in the Navrongo Municipality, Ghana. *Journal of Environment and Earth Science.* **3 (**12).