Tanzania Journal of Engineering and Technology 2022, **41(2)**:131-140
OPEN ACCESS articles distributed under Creative Commons Attribution Licence [CC BY-ND]
Websites: https://www.ajol.info/index.php/tjet, https://tjet.udsm.ac.tz

ISSN 1821-536X (**print**); ISSN 2619-8789 (**electronic**)
DOI: 10.52339/tjet.v41i2.786

Full Length Research Paper

Modelling Monthly Rainfall of Calabar, Nigeria Using Box-Jenkins (ARIMA) Method

Francis James Ogbozige

Department of Civil Engineering, Federal University Otuoke, PMB 126, Yenagoa, Nigeria
Correspondence Email: engr.ogbozige@gmail.com

ABSTRACT

Rainfall has both positive and negative effects on human activities hence, correct prediction of the period of occurrence is very essential. However, the traditional method of predicting months of heavy rainfall is gradually fading out as irregular rainfall pattern is been experienced in most regions of the world. The rainfall pattern in Calabar city in Southern Nigeria has been reported in past literatures of being irregular. Hence, this research applied the Auto-Regressive Integrated Moving Average (ARIMA) also known as Box-Jenkins method to model the rainfall pattern of Calabar. This was achieved by subjecting 50years monthly rainfall (1971 – 2020) to gretl software version 2021b. The analysed data showed that the rainfall of the study area required just one-time differencing to attain stationarity at 95% confidence, while the order of the Auto-Regressive AR(p) and Moving Average MA(q) models were either 1 or 2 in both cases. Hence ARIMA(1,1,1), ARIMA(1,1,2), ARIMA(2,1,1) and ARIMA(2,1,2) were identified and further analyses revealed that ARIMA(2,1,2) best suited the rainfall of the study area. A diagnostic check was carried out on the selected ARIMA (2,1,2) model and it was observed to be reliable (minimal white noise) thereafter, it was used to forecast the rainfall of the study area for some months.

ARTICLE INFO

Submitted: **April 2, 2022**

Revised: **June 3, 2022**

Accepted: **June 27, 2022**

Published: **June 30, 2022**

Keywords: *Correlogram*, *Differencing*, *Stationarity*, *Time Series*

INTRODUCTION

Rainfall is an important component of the hydrologic cycle as it influences human lives in numerous ways including agriculture, tourism and consequently the economy of a society. Besides, rainfall is a major determining factor for the period of constructing Civil Engineering projects. Hence, most Civil Engineering constructions especially at the execution phase are usually carried out during non-rainy months (i.e. dry season). On the other hand, agricultural activities spike during the rainy period but can as well be boosted during the non-rainy period by means of artificial irrigation systems. The rainfall in a given catchment also dictate the amount of surface runoff, which in turns affect flow rate of streams within the catchment. In other words, the important of understanding the pattern of rainfall in a given catchment cannot be over emphasised. Hence, it has been a common practice in developing countries to predict dry and wet seasons using historical rainfall patterns (Gbangou *et al.*, 2021; Zuonon *et al.*, 2020; Chang'a *et al.*, 2010). However, the global climate change phenomenon has altered the usual periods known for raining thus, such predictions failed in most cases. Past literatures including Adamala (2016), Grimaldi *et al.* (2006) as well as Machiwal and Jha (2006) have reported that hydrologic variables such as rainfall and streamflow are Time Series based. This simply means that previous observed values could be used to predict or forecast future values. However, there are different models used for Time Series analysis including, Auto-Regressive (AR), Moving Average (MA), Auto-Regressive Moving Average (ARMA) and Auto-Regressive Integrated Moving Average (ARIMA). Stationarity of raw data is a vital prerequisite for the application of AR, MA and ARMA models but in reality, stationarity of long-term hydrologic data is not certain hence, ARIMA model is more appropriate since it could render raw data stationary through differencing.

The ARIMA model was developed by George Box and Gwilym Jenkins hence known as Box-Jenkins, in order to take account of non-stationarity of raw data by using a series of differencing operators (Maity, 2018). Thereafter, ARMA is applied to the resulting time series. If p is the order of the AR model written as AR(p), d is the order of differencing operator and q is the order of the MA model written as MA(q) then, ARIMA model is represented as ARIMA(p,d,q). Mathematically, the p^{th} order of an AR(p) model is given in Equation (1) while the q^{th} order of an MA(q) model is expressed in equation (2) as both reported by Maity (2018).

$$X(t) = \sum_{i=1}^{p} \phi_i X(t-1) + \varepsilon(t) \quad (1)$$

$$X(t) = \varepsilon(t) \sum_{i=1}^{p} \theta_i \varepsilon(t-1) \quad (2)$$

In equation (1), $X(t)$ is the time series, p is the order of AR model which is the number of lagged values being considered, ϕ_i (for $i \in \{1, 2, 3…., p\}$) are the autoregressive coefficients, and $\varepsilon(t)$ is uncorrelated identically distributed error, also known as white noise. Similarly, in equation (2), θ_i (for $i \in \{1, 2, 3…., q\}$) are the moving average parameters (coefficients) and ε(t-1) is the residual at lag i while $X(t)$ and ε(t) remain same as previously explained. In terms of backshift operator functions, equation (1) and (2) can be written as expressed in equation (3) and (4) respectively.

$$\varepsilon(t) = X(t) - \phi_1 B(X(t)) - \phi_2 B^2(X(t)) - \ldots - \phi_p B^p(X(t)) \quad (3)$$

$$X(t) = \varepsilon(t) - \theta_1 B(\varepsilon(t)) - \theta_2 B^2(\varepsilon(t)) - \ldots - \theta_p B^p(\varepsilon(t)) \quad (4)$$

Hence, ARIMA(p,d,q) model is expressed mathematically as in equation (5).

$$\phi(B)\nabla^d X(t) = \theta(B)\varepsilon(t) \quad (5)$$

where $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ which is the characteristics function of the AR(p), ∇ represents the differencing operation while d is the order of differencing to be decided on the basis of the stationarity of the resulting time series, $X(t)$ is the non-stationary time series, $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$ is the characteristics of function of MA(q) model.

Researchers such as Masum *et al.* (2022), Dwivedi *et al.* (2019), Karimi (2019) and Uba (2015) applied ARIMA to model monthly rainfall pattern in Chattogram (Bangladesh), Junagadh (India), Urmia (Iran) and Maiduguri (Nigeria) respectively, and the results yielded successful predictions. Hence, this research considers Calabar in Southern Nigeria as a case study since literatures including Amadi *et al.* (2021) and Ekpe (2014) have revealed that the rainfall pattern in Calabar fluctuates consistently, yet a predictive rainfall model has not been developed.

METHODS AND MATERIALS

Description of Study Area

Calabar is a city in Southern Nigeria within Latitude 4^0 53' 41.10" to 5^0 7' 37.57" North and Longitude 8^0 14' 14.90" to 8^0 25' 14.03" East, occupying an approximate area of 158km^2. It is the capital of Cross River State and a well-known city for tourism in Nigeria, comprising two Local Government Areas (LGAs) of the state known as Calabar Municipal and Calabar South LGAs. The city shared common boundary with two other LGAs of the state known as Udukpani and Akpabuyo as could be seen in Figure 1.

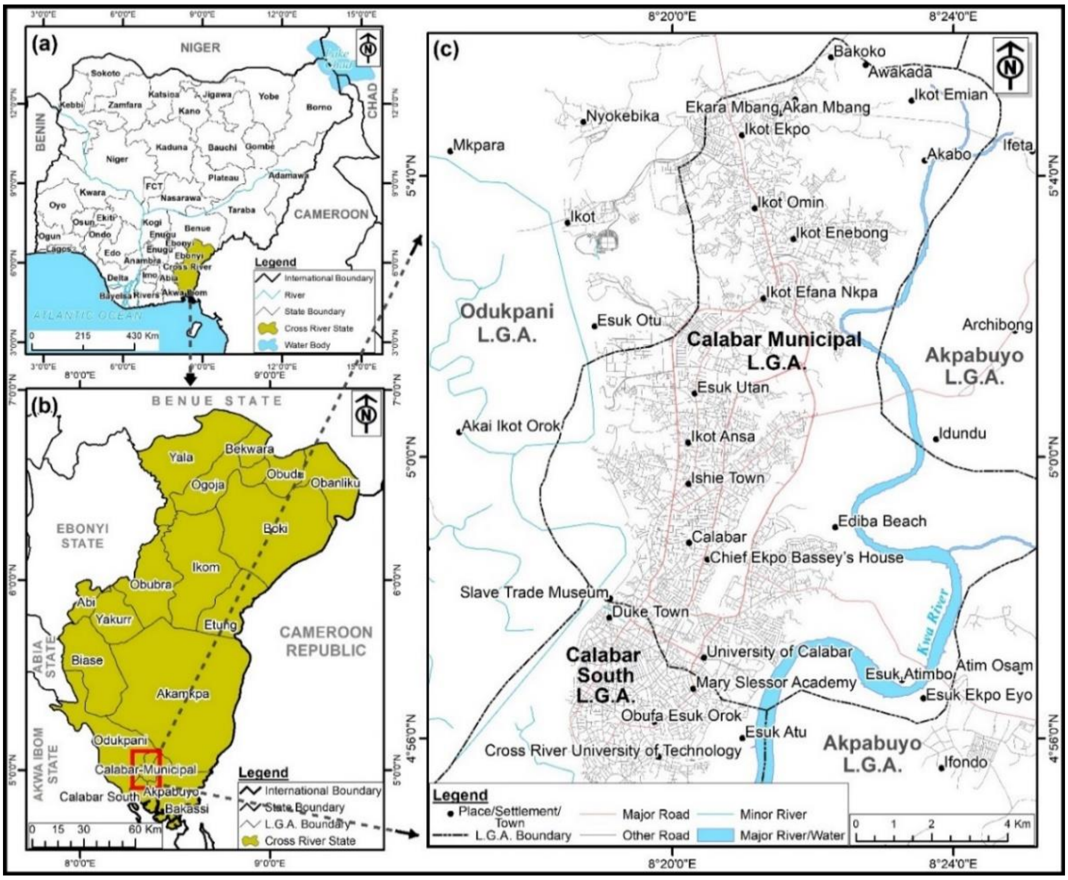


Figure 1: Map of study area

Rainy season in Calabar is usually between the months of April and October with its peak during July or September though, other months of the year experience scanty rainfall occasionally while the dry season is between November and March with its peak during January or February. Notwithstanding, the months with the lowest and highest number of rainy days are January and September respectively, with average monthly rainfall ranging from 50 – 434mm while the average monthly temperature ranged between 24.3 – 27.5^{0}C. The common vegetations are mangrove swamps, rain forest, derived savannah and parkland.

Data Collection and Analysis

The monthly rainfall data of Calabar during January, 1971 till December, 2020 (i.e. 50years) were obtained from the headquarters of Nigerian Meteorological Agency (NIMET), Abuja. In other words, the sample size was 600 (12months into 50). Descriptive statistics with respected to centrality and dispersion of the data were carried out by means of gretl software version 2021b. Stationarity of data, being an important aspect of ARIMA was thereafter checked in the software through Augmented Dickey-Fuller (AGF) test in order to know the required number of differencing (*d*) as reported in Maity (2018).

Correlograms for autocorrelation function (ACF) and partial autocorrelation function (PACF) were plotted so as to understand the range for the order of autoregression AR(p) and moving average MA(q) models, having known the number of differencing (*d*). Thereafter, the various ARIMA models within the identified range of the order (*p*,*d*,*q*) were subjected to Akaike Information Criteria (AIC) test to determine the appropriate values of *p* and *q* in AR(p) and MA(q) respectively since *d* was already known during the AGF test earlier conducted. A diagnostic check on the identified ARIMA model was conducted to know its fitness (Adejumo *et al.*, 2018) and finally, it was used to forecast the monthly rainfall values of the catchment for a period of 24 months.

RESULTS AND ANALYSIS

The basic statistical features or description of the 50years monthly rainfall data used for the research is shown in Table 1 while the rainfall pattern with respect to time is given in Figure 2.

Table 1: Summary of descriptive statistics of rainfall data used

Variables	Statistic
Mean (mm)	245.86
Median (mm)	228.05
Minimum (mm)	0.0000
Maximum (mm)	861.70
Standard deviation	185.46
Skewness	0.5467
Ex. kurtosis	-0.3384
Total observations	600

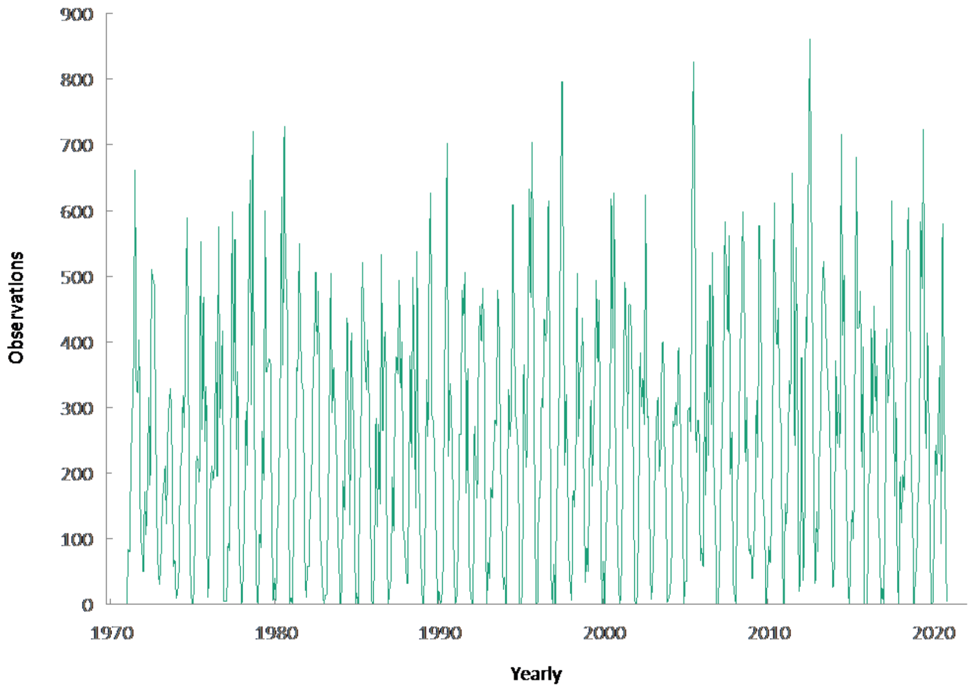


Figure 2: Time series trend of observed rainfall (1971 – 2020)

Since stationarity is a main requirement for Box-Jenkins (ARIMA) model, the data were subjected to stationarity plot using the gretl software version 2021b as could be seen in Figure 3. By mere looking at the information displayed in Figure 3, it seems stationarity already exist in the raw data (i.e. without differencing) as the curve seems to fluctuates about zero (0) at regular interval. However, this could be deceptive since the observed rainfall data are numerous (600) hence, the compression of these data within the horizontal axis might have led to overlapping of the monthly rainfall. Hence, the rainfall data were further subjected to Augmented Dickey-Fuller (AGF) test to understand the number of differencing required to achieve stationarity using three levels of stationarity test namely; test without constant, test with constant, and test with constant and trend as shown in Table 2.

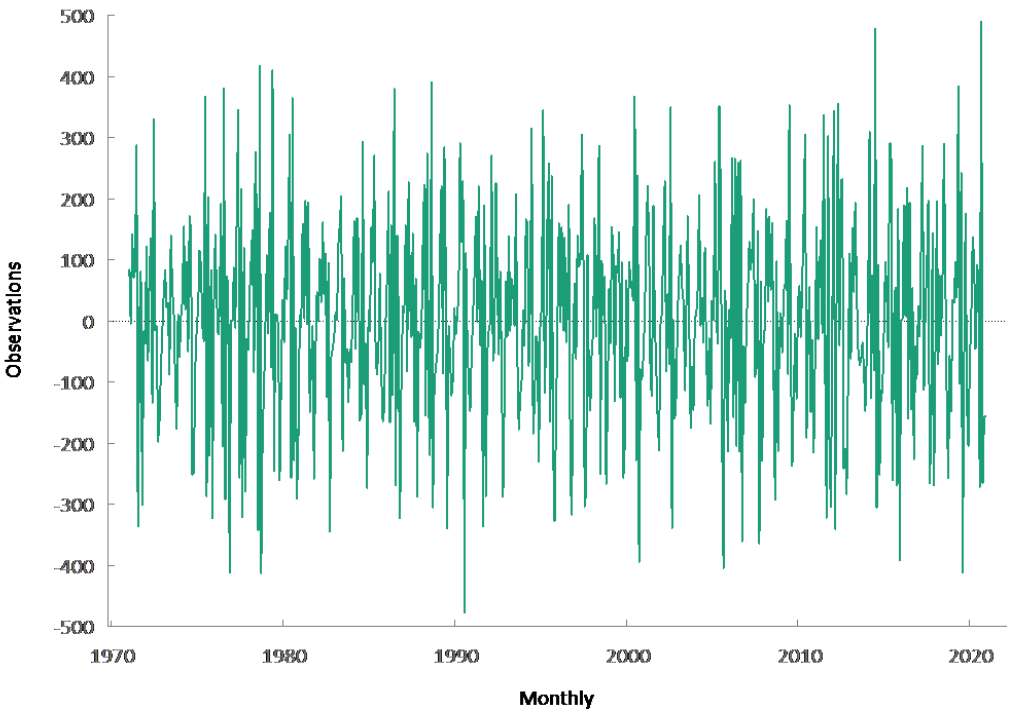


Figure 3: Stationarity plot for rainfall data

Table 2: Augmented Dickey-fuller statistics on rainfall

	Statistic	**p-value**	**comment**
Test without constant	-17.2401	0.0001	Stationary at first difference
Test with constant	-17.2255	0.0001	
Test with constant and trend	-17.2123	0.0001	

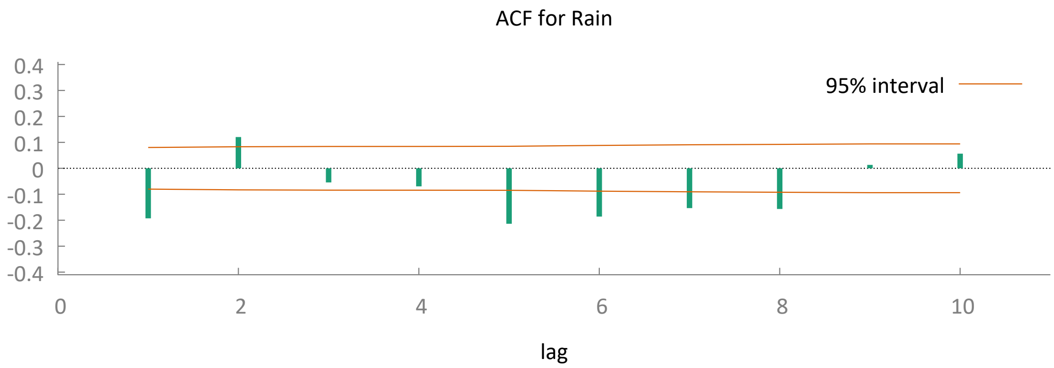
5% critical =3.1456

It is clearly revealed in Table 2 that the number of differencing required to achieve stationarity for the rainfall data is one; i.e. at first differencing (95% confidence). In other words, the value of d in ARIMA(p,d,q) model is 1. Next, was the identification of the order of the Autoregressive (AR) and Moving Average (MA) models. This was achieved by obtaining the correlograms information for the autocorrelation function (ACF) and partial autocorrelation function (PACF) as shown in Table 3 and Figure 4.

Table 3: Bartlett standard error statistics for ACF and PACF of rainfall data

Lag	ACF	PACF	Q-stat. [p-value]
1	-0.1928 ***	-0.1928 ***	22.3664 [0.000]
2	0.1204 ***	0.0865 **	31.1077 [0.000]
3	-0.0543	-0.0172	32.8901 [0.000]
4	-0.0696	-0.0968 **	35.8230 [0.000]
5	-0.2138 ***	-0.2490 ***	63.5172 [0.000]
6	-0.1860 ***	-0.2869 ***	84.5204 [0.000]
7	-0.1533 ***	-0.2707 ***	98.8149 [0.000]
8	-0.1567 ***	-0.3455 ***	113.7743 [0.000]
9	0.0132	-0.3087 ***	113.8809 [0.000]
10	0.0564	-0.3350 ***	115.8266 [0.000]

***, ** and * indicate significance levels at 1%, 5% and 10% respectively.



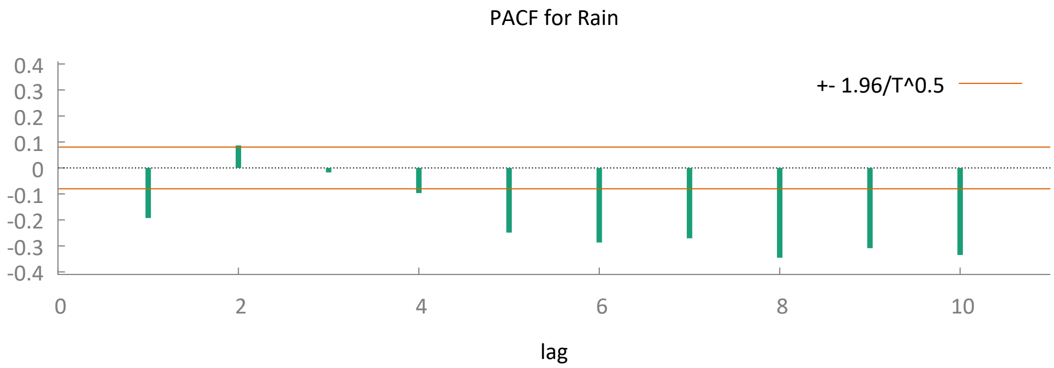


Figure 4: Correlogram plots for ACF and PACF

The information in Table 3 and Figure 4 showed that there is a decay in the third lag (i.e. lag 3) for both ACF and PACF thus, suggesting that the order of the AR(p) and MA(q) models falls on either lag 1 or lag 2 for each case. Since the number of differencing (*d*) to achieve stationarity was identified to be 1, it implies the order of the ARIMA(*p*,*d*,*q*) model is either ARIMA(1,1,1); ARIMA(1,1,2); ARIMA(2,1,1) and ARIMA(2,1,2). However, in order to determine the ARIMA model that best fits the rainfall data, the Akaike Information Criteria (AIC) of the various identified ARIMA(*p*,*d*,*q*) models were determined and their values are given in Table 4.

Table 4: Akaike Information Criteria (AIC) values for identified ARIMA models

Order of ARIMA	AIC
ARIMA (1,1,1)	7808.242
ARIMA (1,1,2)	7794.398

ARIMA (2,1,1)	7797.702
ARIMA (2,1,2)	7690.860

Based on the data in Table 4, ARIMA (2,1,2) has the least AIC value (7690.860) hence, it was selected as the best model for the monthly rainfall of the catchment. A diagnostic check was carried out on the best fitted model being ARIMA (2,1,2) using Q-Q plot and normality residual testing as shown in Figure 5 and Figure 6 respectively. It is conspicuous that Figure 5 has minimal *white noise* as most of the data points are fitted in the line of best fit just as the data points in Figure 6 are mostly within the normal distribution curve, signifying that the ARIMA (2,1,2) model is reliable.

Having seen that the ARIMA (2,1,2) model is reliable, it was used to forecast the monthly rainfall of the catchment (Calabar) up to the year 2022 as shown in Figure 7 while the forecasted or predicted values are clearly presented in Table 5

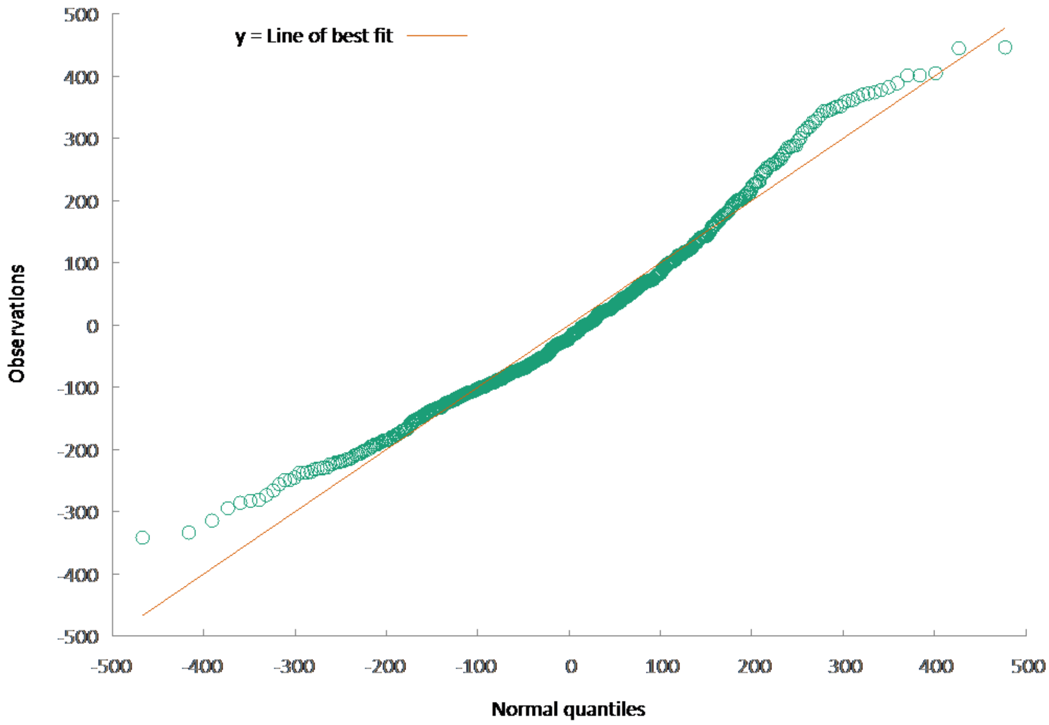


Figure 5: Q-Q plot of ARIMA (2,1,2) model

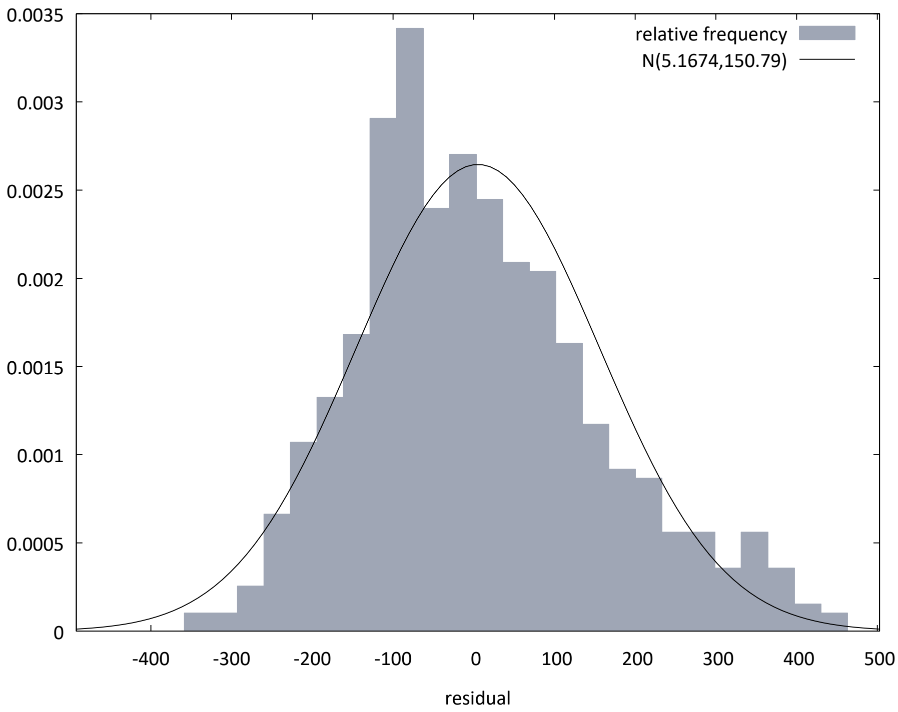


Figure 6: Normality residual testing of ARIMA (2,1,2) model

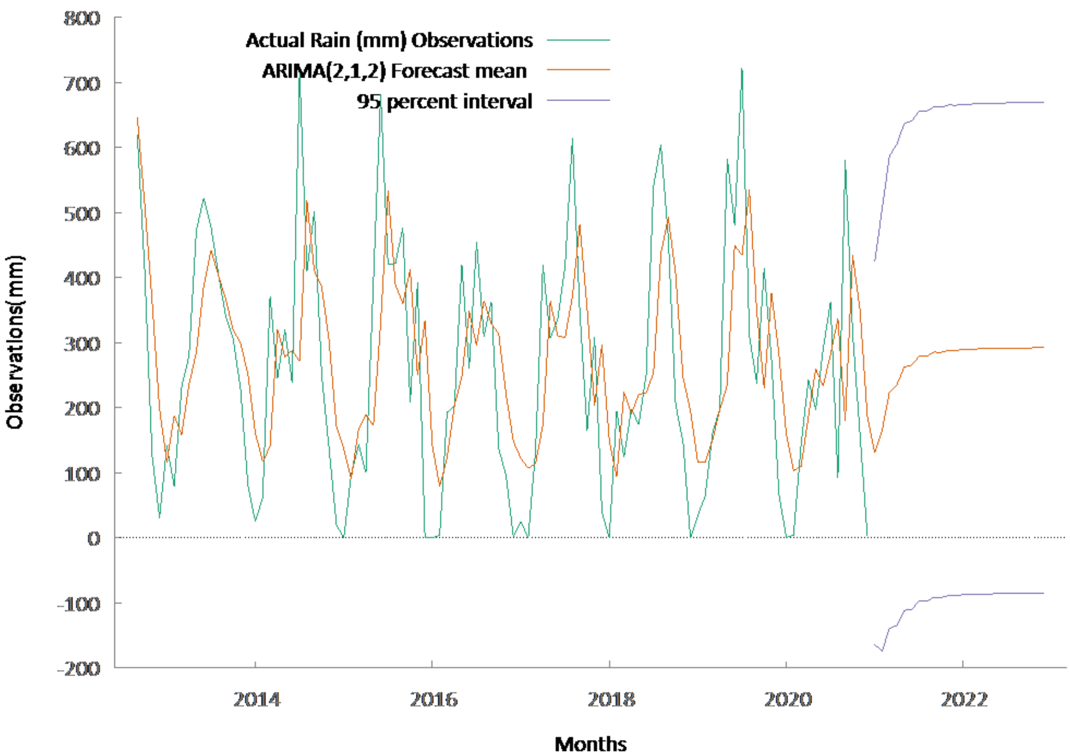


Figure 7: Monthly forecast of Calabar rain using ARIMA (2,1,2) model

Table 5: Forecasted values of rainfall

Date	Prediction (mm)	Lower (mm)	Upper (mm)
January, 2021	43.9	0.0	54.9
February, 2021	0.0	0.0	7.2
March, 2021	98.0	0.0	122.5
April, 2021	199.5	0.0	249.4
May, 2021	231.9	0.0	289.9
June, 2021	343.4	0.0	429.3
July, 2021	300.6	0.0	375.8
August, 2021	360.9	0.0	451.1
September, 2021	388.7	0.0	485.9
October, 2021	80.8	0.0	101.0
November, 2021	169.3	0.0	211.6
December, 2021	65.8	0.0	82.3
January, 2022	17.3	0.0	21.6
February, 2022	44.4	0.0	55.5
March, 2022	141.7	0.0	177.1
April, 2022	245.5	0.0	304.4
May, 2022	318.9	0.0	398.6
June, 2022	326.7	0.0	408.4
July, 2022	466.4	0.0	583.0
August, 2022	382.0	0.0	477.5
September, 2022	397.1	0.0	496.4
October, 2022	219.5	0.0	274.4
November, 2022	192.1	0.0	240.1
December, 2022	29.6	0.0	37.0

CONCLUSION AND RECOMMENDATIONS

Based on the analysed results obtained from this research, it could be concluded that the monthly rainfall pattern of Calabar city follows ARIMA (2,1,2) model. The research have also shown that the maximum monthly rainfall in Calabar metropolis occur in between the months of July to September while the months with the least rainfall fluctuate between December, January and February. In addition, the maximum and least monthly rainfall in Calabar does not show a steady increase nor decrease from year to year as the values fluctuates irregularly thus, affirming the reports of Amadi *et al*. (2021) and Ekpe (2014).

Construction companies operating within the study area are hereby advised to make use of the model to predict periods of downpour for a given year in order to avoid wastage of construction materials especially during the execution phase of projects.

REFERENCE

Adamala, S. (2016). Time Series Analysis: A Hydrological Prospective. *American Journal of Scientific Research and Essays*, 1(1): 31 – 40.

Adejumo, A.O., Oloyede, T.O., Adejumo, O.A., Oguntunde, P.E., Odetunmibi, O.A., Ikoba, N.A. and Job. O. (2018). Modelling of Enugu State Monthly Rainfall Using Box and Jenkins

Methodology. *Annals (Computer Science Series),* 1691): 156 – 163.

Amadi, S.O., Agbor, M.E. and Udo, S.O. (2021). Vulnerability of Calabar Rainfall to Climatic Variability Events: A Critical Factor in Integrated Water Resources Management in the Tropical Coastal Location in Southeastern Nigeria. *International Journal of Sustaible Development and Planning,* 16(1): 115 – 122.

Chang'a, L.B., Yanda, P.Z. and Ngana, J. (2010). Indigenous Knowledge in Seasonal Rainfall Prediction in Tanzania: A Case of the South-Western Highland of Tanzania. *Journal of Geography and Regional Planning,* 3(4): 66-72.

Dwivedi, D.K., Kelaiya, J.H. and Sharma, G.R. (2019). Forecasting Monthly Rainfall using Autoregressive Integrated Moving Average Model (ARIMA) and Artificial Neural Network (ANN) Model: A Case Study of Junagadh, Gujarat, India. *Journal of Applied and Natural Sciences,* 11(1): 35 – 41.

Ekpe, I.A. (2014). Rainfall Anomalies and Trend in Calabar: An Imperative for Human Adaptation to Climate Change. A Thesis Submitted to the Graduate School, University of Calabar, Nigeria.

Gbangou, T., Slobbe, E.V., Ludwig, F., Kranjac-Berisavljevic, G. and Paparrizos, S. (2021). Harnessing Local Forecasting Knowledge on Weather and Climate in Ghana: Documentation, Skills, and Integration with Scientific Forecasting Knowledge. *Weather, Climate, and Society*, 13: 23 – 37.

Grimaldi, S., Koutsoyiannis, D., Piccolo, D. and Napolitano, F. (2006). Time Series Analysis in Hydrology. *Physics and Chemistry of Earth, Parts A/B/C,* 31(18): 1097 – 1098.

Karimi, B. (2019). Monthly Rainfall Prediction using ARIMA and Gene Expression Programming: A Case Study in Urmia, Iran. *Online Journal of Engineering Sciences and Technology,* 2(3): 8 – 14.

Machiwal, D. and Jha, M.K. (2006). Time Series Analysis of Hydrologic Data for Water Resources Planning and Management: A Review. *Journal of Hydrology and Hydromechanics,* 54(3): 237 – 257.

Maity, R. (2018). Statistical Methods in Hydrology and Hydroclimatology. 1st Ed. Springer Nature, Singapore.

Masum, M.H., Islam, R., Hossen, M.A. and Akhiel, A.A. (2022). Time Series Prediction of Rainfall and Temperature Trend using ARIMA Model. *Journal of Scientific Research,* 14(1), 215 – 227.

Uba, E.S. (2015). An Application of Time Series Analysis in Modeling Monthly Rainfall Data for Maiduguri, North Eastern Nigeria. *Mathematical Theory and Modeling*, 5(11): 24 – 33.

Zounon, H.N., Baco, N.M., Dagoudo, B.A. and Sahagui, S.K. (2020). Traditional Ecological Knowledge of Predicting Rain for Climate Adapting in North Benin. *International Journal for Innovation Education and Research,* 8(1): 141 – 155.