

Optimal Stochastic Forecast Models of Rainfall in South-West Region of Nigeria

Onyeka-Ubaka, J. N.^{1*}, Halid, M, A¹, Ogundeji, R. K.¹

1. Department of Mathematics, University of Lagos, Nigeria.

*Corresponding author: jonyeka-ubaka@unilag.edu.ng

Article Info

Received: 3 March 2021

Revised: 19 July 2021

Accepted: 17 September 2021

Available online: 10 November 2021

Abstract

Rainfall estimates are important components of water resources applications, especially in agriculture, transport. constructing irrigation and drainage systems. This paper aims to stochastically model and forecast the rainfall trend and pattern for a city, each purposively selected in five states of the South-Western Region of Nigeria. The data collected from Nigerian Meteorological Agency (NIMET) website are captured with fractional autoregressive integrated moving average (ARFIMA) and seasonal autoregressive integrated moving average (SARIMA) models. The autocorrelation function (ACF) and partial autocorrelation function (PACF) are used for model identification, the models selected are subjected to diagnostic checks for the models adequacy. Several tests: Augmented Dickey Fuller (ADF), Ljung Box and Jarque Bera tests are used for investigating unit root, serial autocorrelation and normality of residuals, respectively; the mean square error, root mean square error and mean absolute error are employed in validating the optimal stochastic model for each city in all states, in which the model with the lowest error of forecasting of all competing models is suggested as the best. The analyses and findings suggest SARIMA(1,0,1)(1,1,0) [12], SARIMA(3,0,2)(1,0,0) [12], SARIMA(1,0,0)(1,1,0) [12], SARIMA(2,0,2)(2,1,0) [12] and SARIMA(0,0,1)(1,1,0) [12] for (Ibadan) Oyo State, (Ikorodu) Lagos State, (Osogbo) Osun State, (Abeokuta) Ogun State and (Akure) Ondo state, respectively. The seasonal ARIMA (SARIMA) model was proven to be the best optimal stochastic forecast model for forecasting rainfall in the selected cities. The SARIMA model was, therefore, recommended as a veritable technique that will assist decision makers (Government, Farmers, and Policymakers) to establish better strategies “aprior” on the management of rainfall against upcoming weather changes to ensure increase in agricultural yields for the betterment of the citizenry and general economic growth.

Keywords: Optimum Forecast, ARFIMA, SARIMA, Model Diagnostics, Residuals.

MSC2010:37M10; 62M10; 91B25; 91B84

1 Introduction

Rainfall estimates are important components of water resources applications. Accurate rainfall forecasting provides information for any water resources planning systems like transport, agriculture, irrigation schemes, city and environmental plans, reservoir operation strategies, flood and drought

analysis, and for designing and maintaining of hydraulic structures. Nigeria, being an agrarian country, depends highly on climate. Forecast of rainfall will in the long-run ensure stable, sufficient and adequate moisture that can generate growth, efficiency and stability in agricultural yields since all national crop production, agricultural activities as a whole depend heavily on the distribution and quantity of rainfall. There is no doubt that water resources are essential renewable resources which are the basis for development and existence of a society ([1] Usoro and Awakessien, 2018). Nigeria is divided into two major parts; the southern and the northern region. Each of the parts has three sub-regions, which makes it a total of six regions in the country, which are; North-East, North-West, North Central, South-East, South-West and South-South regions. Each of these regions have different features that describe the economic activities, settlements and way of life of the inhabitants. It is indisputable that the amount of rainfall recorded in the southern part of the country is much more than that of the northern part, thereby having a great impact on the type of vegetation in the region. The southern part, for instance, has coastal areas with rainforest and mangrove vegetation whose amount of rainfall are not equal, which is also applicable in the northern part of the country.

Forecasting of rainfall has so many advantages such as helping people plan for when to do certain activities that are influenced by weather, to help people with health related issues to plan the day, to help businesses plan for transportation hazards that can result from the weather as related to driving and flying for example, and most importantly to help farmers and gardeners plan for crop irrigation and protection, among others. After crops have been planted, the minimization or eradication of crop failure during the early stages depend heavily on the amount of rainfall. In addition, wise decisions would be taken with regards to the need for irrigation and the timings, utilization of hydro-electric power, the conservations of water strategies for dams, all depends on reliable rainfall forecast. More so, a reasonable knowledge of the data on rainfall assists in the forecasting of length of growing or rainy season, which is most useful in the selection of crop varieties, crop matching and crop sequences. Several authors have written on SARIMA models adopting several methods such as the non-parametric methods (Artificial Neural Networks) and Parametric Methods (Exponential smoothing, Extrapolation of trend curves, the Holt-Winters forecasting procedure and Box Jenkins procedure). Forecasting rainfall is a very difficult task in the modern world, as a result of facing major environmental problem of global warming which has rendered previously employed methods less effective as a result of changing patterns and variability in rainfall ([2] Box and Jenkins, 1976). [3] Tariq and Abbasabd (2016) used a SARIMA model for Nyala station (Sudan), which was considered appropriate for forecasting monthly rainfalls. [4] Valipour (2015) proved the appropriateness of SARIMA model for long-term runoff forecasting in the United States as compared to the ARIMA models. [5] Umar, et al. (2014) adopted a non-seasonal ARIMA (1, 2, 1) models for rainfall data in North-Eastern Nigeria and observed that the seasonality of the data cannot be captured. Additive and multiplicative methods of decomposing trends, seasonal and cyclical variations in the rainfall data were applied by [6] Abdulrahim, et al. (2013) in Sokoto rainfall data. They observed a high concentration of rainfall in the months of June, July and August, while decrease in the rainfall was observed in March, April and October in the analysis of precipitation indicating the farmers' period of growing crops.

In the South-South region of Nigeria, [7] Etuk, et al. (2013) fitted a long-range SARIMA(5, 1, 0)(0, 1, 1)₁₂ model to rainfall data in Port Harcourt and observed that the fitted model was not equally fair because of the order of the non-seasonal autoregressive part. [8] Ekwe, et al. (2014) analyzed the trend and seasonal behaviour of rainfall data in Nasarawa State using the regression analysis for the period of 20 years (1993-2012) and discovered that the highest rainfall occurred in 1996, and lowest in 2010 with the highest amount of rainfall value recorded in August. [9] Yaya, et al. (2015) fitted trend and seasonal models to a rainfall data from different meteorological stations in the six geographical zones in Nigeria. In the analyses, there existed seasonality pattern in the rainfall across the six zones in the country. In each of the zones, there are increasing amount of rainfall, according to their findings. The trend component with the model $Y = b_0 + b_1X$ was isolated first and after that $U_t = \Phi_{12}U_{t-12} + \varepsilon_t$ was used to fit the data after the trend removal. The behaviours of the residuals from the fitted models were not shown to indicate adequacy of the models. They

conclude that it is not possible for one particular model to be adequate in fitting rainfall data in all the zones. The prediction result can be a guide leading to very important theory evidence for the agricultural managers to make pertinent decisions ([10] Guifen, et al., 2010). [11] Olatayo and Taiwo (2014) introduced three fundamental different approaches for designing a model, the statistical method based on autoregressive integrated moving average (ARIMA), the emerging fuzzy time series (FST) model and the non-parametric method (Theil's regression) and observed that ARIMA (1, 2, 1) was used to derive the weights and the regression coefficients, while the Theil's regression was used to fit a linear model.

The motivation and objective of this paper is to select the optimal stochastic model(s) from the different seasonal autoregressive integrated moving average (SARIMA) models and fractional autoregressive integrated moving average (AFRIMA) models, for forecasting rainfall in the South West Region of Nigeria. The South-Western Region as a whole will greatly benefit from the success of this study by having an insight of the behaviour, pattern or trend of rainfall. Information will also be available to the Government and community to prepare for any form of disasters that can be caused by excessive and/or scarcity of rainfall. The five cities identified and selected for this study are predominantly agricultural dependent, it would be of utmost importance to the farmers, and there would be a significant increase in the yield if there is a timely access to reliable forecast of rainfall data in these cities. The remainder of this paper is organized as follows. Section 2 introduces the theory about the models. Section 3 presents modelling strategy used for the two models, describes the chosen tests and how the forecasting accuracy are evaluated. In Section 4, results and discussion for the data sets of monthly rainfall are presented. Section 5 concludes the paper by summarizing the results.

2 Model

(a) Seasonal ARIMA (SARIMA) Models

A time series is said to be homogeneous non-stationarity if $(1 - B)^d y_t$ is stationary for some value of $d \geq 1$. Seasonal autoregressive integrated moving average (SARIMA) models are an adaptation of autoregressive integrated moving average (ARIMA) models to specifically fit seasonal time series. Seasonal differencing in this model reduces the periodic intensity and makes a series static. The SARIMA model is applied to the time series y_t with the following expression ([12] Brockwell and Davis, 1991):

$$\Phi(B^s)\phi(B)\Delta^d\Delta_s^D y_t = \theta_o + \Theta(B^s)\theta(B)\varepsilon_t. \quad (2.1)$$

These models, SARIMA $(p, d, q) \times (P, D, Q)_s$ contain non-seasonal and season components, where s is the seasonal length, for example $s = 12$ for monthly data; B is the lag operator and ε_t is assumed to be a Gaussian white-noise process with mean zero and variance σ^2 ; $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ is the autoregressive process of order p . The process is stationary if the roots of $1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 0$, lie outside a unit circle. The partial autocorrelation function (PACF) vanishes after lag p ; $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p$ is the moving average order q (MA(q)). The MA(q) process is always stationary because $1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p < \infty$. The process is invertible if the roots of $1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p = 0$, lie outside a unit circle; $\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps}$ and $\Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_p B^{ps}$ are the seasonal autoregressive and seasonal moving average models components. The difference operator is Δ^d where d specifies the order of differencing and the seasonal difference operator is Δ_s^D where D is the order of seasonal differencing. The difference operators are applied to transform the observed non-stationary time series y_t to the stationary process y_t^* with the following equation:

$$y_t^* = (1 - B)^d (1 - B^s)^D y_t \quad (2.2)$$

is the backshift operator such that $B^d y_t = y_{t-d}$.

For the state-space model and forecasting of integrated processes, the fact that the observed variable can be replaced by the differenced variable as in Equation (2.2) is used. [13] Box, *et al.* (2008)

propose that the SARIMA $(p, d, q) \times (P, D, Q)_s$ for y_t can be seen as a special form of the equivalent representation of as an ARMA $(p + sP, q + sQ)$ written as

$$\phi(B)^* y_t^* = \theta_0 + \theta(B)^* \varepsilon_t. \quad (2.3)$$

The AR part $[\phi(B)^*]$ in this model is derived by multiplying the autoregressive lag polynomials $\phi(B)$ and $\Phi(B^s)$. Hence, it is

$$\phi(B)^* = \phi(B)\Phi(B^s) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps}). \quad (2.4)$$

For the MA part $[\theta(B)^*]$, multiply the moving average lag polynomials $\theta(B)$ and $\Theta(B^s)$ which give

$$\theta(B)^* = \theta(B)\Theta(B^s) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_p B^p)(1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_p B^{ps}). \quad (2.5)$$

(b) Autoregressive fractionally integrated moving average (ARFIMA)

ARFIMA models are time series models that generalize ARIMA models by allowing non-integer values of the differencing parameter. These models are useful in modelling time series with long memory in which deviations from the long-run mean decay more slowly than an exponential decay. The acronyms "ARFIMA" or "FARIMA" are often used, although it is also conventional to simply extend the "ARIMA (p, d, q) " notation for models, by simply allowing the order of differencing, d , to take fractional values. An ARFIMA model shares the same form of representation as the ARIMA (p, d, q) process, specifically

$$\left(1 - \sum_{i=1}^p \Phi_i B^i\right) (1 - B)^d X_t = \left(1 + \sum_{i=1}^p \Theta_i B^i\right) \varepsilon_t. \quad (2.6)$$

In contrast to the ordinary ARIMA process, the "difference parameter", d , is allowed to take non-integer values.

3 Methodology

The self-projecting time series uses only the time series data of the activity to be used to generate forecasts. This approach is typically useful for short to medium-term forecasting. The underlying goal of the Box-Jenkins forecasting method is to find an appropriate formula so that the residuals are as small as possible and exhibit no pattern. Box and Jenkins bases the model selection on three stages: identification, estimation and diagnostic checking.

The identification part then begins by finding the appropriate order of integration in level, that is, d . This order will be found with two unit root tests, the Augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. [14] Kwiatkowski *et al.* (1992) state that these two tests can be said to complement each other. The Augmented Dickey-Fuller (ADF) test is used to test the null hypothesis of a unit root against the alternative of stationarity and it is based on the following model ([15] Dickey and Fuller, 1979):

$$\Delta y_t = \alpha + \beta t + (\rho - 1)y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \quad (3.1)$$

where α is a constant, β is the coefficient of a simple time trend, ρ is the parameter of interest, Δ is the first difference operator, δ_1 are parameters and p the lag order of the autoregressive process. The choice of including the intercept and/or the time trend should be made beforehand. The lagged differenced variables are included to account for possible serial correlation that would otherwise appear in the error term fit which is assumed to be approximately a white noise process ([16] Banerjee *et al.*, 1993). What is tested is the null hypothesis of a unit root, that is, $\rho = 1$ against the alternative hypothesis of stationarity $|\rho| < 1$. The test statistic that is used is based on the t -statistic

$$DF_T = \frac{\hat{\rho} - 1}{SE\hat{\rho}} \quad (3.2)$$

where the estimated value of the test statistic should be compared to the value of the relevant critical value of the Dickey-Fuller test ([16] Banerjee *et al.*, 1993).

The Kwiatkowski et al. (KPSS) test was developed in 1992 and assumes the following model:

$$y_t = \gamma_t + r_t + \varepsilon_t \tag{3.3}$$

where γ is the coefficient of a simple time trend, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ and r_t is a random walk, that is, $r_t = r_{t-1} + u_t$, where u_t is a white noise process with mean zero and variance σ_u^2 and r_0 is considered to be the intercept. The null hypothesis that σ_u^2 implies testing that the time series is either level ($\gamma = 0$) or trend stationary ($\gamma \neq 0$) against the alternative that it is non-stationary. The test statistic is then derived by first fitting y_t depending on only an intercept, or an intercept and a trend. The resulting residuals e_t are then used to derive a consistent estimate of the variance

$$s^2(l) = T^{-1} \sum_{i=1}^T e_i^2 + 2T^{-1} \sum_{s=1}^l w(s, l) + \sum_{t=s+1}^T e_t e_{t-s}. \tag{3.4}$$

In this equation, $w(s, l) = 1 - \frac{s}{l+1}$ is the Bartlett window and guarantees that the estimated variance is non-negative. The bandwidth l is decided by the Newey-West automatic with the Bartlett kernel ([17] Schwert, 2009). The next step is to derive the partial sum series of the residuals, that is

$$S_t = \sum_{i=1}^t e_i, \quad t = 1, 2, 3, \dots, T. \tag{3.5}$$

The results in Equations (3.4) and (3.5) are then used to derive the Lagrange multiplier based KPSS test statistic

$$\eta_\mu = \frac{\eta_\mu}{S^2(l)} = T^{-1} \frac{\sum s_t^2}{s^2(l)}. \tag{3.6}$$

where the critical values can be found in [14] Kwiatkowski et al. (1992).

The integration order in season, D , is found with the Canova-Hansen (CH) test and the Hylleberg-Engle-Granger-Yoo (HEGY) test. These tests are used to evaluate the seasonal stationarity of each time series and have been shown to complement each other ([18] Hylleberg, 1995). The Canova and Hansen (CH) test was developed in 1992 and is used to test the null hypothesis that the time series process is stationary with deterministic seasonality against the alternative that it has a seasonal unit root. It is closely related to the KPSS test since both are based on the Lagrange Multiplier (LM) statistic. This test assumes the following model:

$$y_t = \mu + x_t' \beta + S_t + e_t \tag{3.7}$$

where y_t is the modelled time series, x_t is a vector of explanatory variables which can be lagged values of y , S_t is a deterministic seasonal component of period $s = 12$ for monthly data and $e_t \sim N(0, \sigma^2)$ is white noise and uncorrelated with x_t and S_t . It should be noted that the time series that is used for this test is assumed stationary in level. If no explanatory variables are included then the error e_t will be the difference between the modelled process and its seasonal component S_t . The requirements of e_t are not strict but it should not appear to have tendencies for serial correlation, heteroskedasticity or seasonal behaviour. Further, the seasonal component can be written like this

$$S_t = d_t' \alpha \tag{3.8}$$

where d_t is a seasonal dummy indicator for the 12 lags and α is a parameter vector representing the seasonal effects. The seasonal component can then be equivalently written on a trigonometric representation, that is

$$S_t = \sum_{j=1}^q f_{jt}' \gamma_j \tag{3.9}$$

where $q = \frac{s}{2} = 6$ for monthly data, for $j < q$, $f_{jt} = [\cos((j/q)\pi t), \sin((j/q)\pi t)]$ and for $j = q$, $f_{jt} = \cos(\pi t)$. Thus the following vectors of $s - 1 \times 1$ objects can be specified as follows:

$$f_t = \begin{pmatrix} f_{1t} \\ f_{2t} \\ \vdots \\ f_{qt} \end{pmatrix}$$

$$\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_q \end{pmatrix}$$

will then lead to Equation (3.9) being equivalently written as

$$S_t = f_t' \gamma. \quad (3.10)$$

Putting Equation (3.10) in Equation (3.7) then leads to the model being specified in the following way:

$$y_t = \mu + x_t' \beta + f_t' \gamma + e_t. \quad (3.11)$$

This representation is good because it presents the seasonal components as cyclical where γ_j is the parameter for the seasonal frequency $j\pi/q$ connected to each cyclical process f_t for each seasonal component S_t . It is also important to note that inclusion of lagged variables of y_t in x_t could potentially lead to seasonal unit roots being captured. This could lead to the null hypothesis not being rejected, since the alternative hypothesis is that the time series has a seasonal unit root. Thus since no included lags could lead to the error term e_t being serially correlated the number must be decided with caution. The test statistic is then derived by

$$L = \sum_{i=1}^T \widehat{F}_i' A (A' \Omega^f A)^{-1} A' \widehat{F}_i \quad (3.12)$$

where $\widehat{F}_t = \sum_{i=1}^T f_t \widehat{e}_t$ with \widehat{e}_t being the residuals from the estimation of the model in Equation (3.7),

Ω^f is the long-run covariance matrix of $f_t e_t$ and A is specified to test for the seasonal unit root at one or a number of seasonal lags. If stationarity is rejected at all frequencies then seasonal differencing should be performed to make the time series stationary. The test statistic is then showed to asymptotically follow the Von Misses goodness-of-fit distribution with critical values presented in [19] Canova and Hansen (1992) which the estimated value is compared to.

The HEGY test is an extension on theory from the Dickey-Fuller test to test for seasonal unit roots and was developed by [20] Hylleberg et al. (1990). In their paper they developed the test for quarterly data and an extension to monthly data was later created by [21] Franses (1991) on which this section is based. The first step is to present the seasonal difference operator Δ_s , where it follows that there should be $s = 12$ roots on the unit circle for monthly data. This can be described by the following equation:

$$\begin{aligned} \Delta_z &= (1 - B^{12}) = (1 - B)(1 + B)(1 - iB)(1 + iB) \times \\ &\quad [1 + (\sqrt{3} + i)B/2][1 + (\sqrt{3} - i)B/2] \times [1 - (\sqrt{3} + i)B/2][1 - (\sqrt{3} - i)B/2] \times \\ &\quad [1 + (\sqrt{3} + i)B/2][1 - (\sqrt{3} - i)B/2] \times [1 - (\sqrt{3} + i)B/2][1 + (\sqrt{3} - i)B/2] \end{aligned} \quad (3.13)$$

where B is the lag operator and all polynomials except $(1 - B)$ are connected to seasonal unit roots. The test is then based on the following equation:

$$\begin{aligned} \gamma^*(B)y_{8,t} = & \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-1} + \pi_4 y_{3,t-2} + \pi_5 y_{4,t-1} + \pi_6 y_{4,t-2} + \pi_7 y_{5,t-1} \\ & + \pi_8 y_{5,t-2} + \pi_9 y_{6,t-1} + \pi_{10} y_{6,t-2} + \pi_{11} y_{7,t-1} + \pi_{12} y_{7,t-2} + \mu_t + \varepsilon_t \end{aligned} \quad (3.14)$$

where μ_t is deterministic and specified to include a constant, seasonal dummies and/or a trend, ε_t is a white noise process and $\gamma^*(B)$ is a polynomial of B and the significance of the parameters are what is of interest. Further the y 's are specified as lagged combinations of the observed time series process as follows:

$$\begin{aligned} y_{1,f} &= (1 + B)(1 + B^2)(1 + B^4 + B^8)y_t \\ y_{2,f} &= -(1 - B)(1 + B^2)(1 + B^4 + B^8)y_t \\ y_{3,f} &= -(1 - B^2)(1 + B^4 + B^8)y_t \\ y_{4,f} &= -(1 - B^4)(1 - \sqrt{3}B + B^2)(1 + B^2 + B^4)y_t \\ y_{5,f} &= -(1 - B^4)(1 + \sqrt{3}B + B^2)(1 + B^2 + B^4)y_t \\ y_{6,f} &= -(1 - B^4)(1 - B^2 + B^4)(1 - B + B^2)y_t \\ y_{7,f} &= -(1 - B^4)(1 - B^2 + B^4)(1 + B + B^2)y_t \\ y_{8,f} &= (1 - B^{12})y_t \end{aligned} \quad (3.15)$$

The model in Equation (3.14) is then estimated by ordinary least squares with focus on the estimates of the π s. This is done for the specifications of μ_t being considered relevant. The t -test are then used on $\hat{\pi}_1$ and $\hat{\pi}_2$ to test the one-sided null hypothesis of a unit root for $\hat{\pi}_1$ and a seasonal unit root for $\hat{\pi}_2$ against the alternative hypothesis of no unit root. For $\hat{\pi}_i$ with $i > 2$ a seasonal unit root at a specific frequency is only present when it occurs for connected pairs of parameters. Thus, it follows that the F -test is used to the joint two-sided null hypothesis of a unit root on the connected pairs, $(\hat{\pi}_3, \hat{\pi}_4)$, $(\hat{\pi}_5, \hat{\pi}_6)$, $(\hat{\pi}_7, \hat{\pi}_8)$, $(\hat{\pi}_9, \hat{\pi}_{10})$ and $(\hat{\pi}_{11}, \hat{\pi}_{12})$ against the alternative that there are seasonal unit roots. Critical values for all these tests can be found in [21] Franses (1991). Non-rejection of the null hypothesis of a seasonal unit root for the test of $\hat{\pi}_1$ and all the connected pairs would imply that seasonal differencing of the series should be performed.

The next part is to find the autoregressive orders p and P and the moving average orders q and Q . The correlogram is first used to make guesses for appropriate orders ([2] Box and Jenkins, 1976). The correlogram is used to check the randomness of the data. Box and Jenkins states that the autocorrelation function (ACF) and partial autocorrelation function (PACF) can be used to identify the orders p, q, P, Q for a SARIMA model ([13] Box et al., 2008). The ACF at lag k is given by ([22] Harvey, 1989) as

$$r_k = \frac{c_k}{c_0} \quad (3.16)$$

where c_k is the following autocovariance function:

$$c_k = T^{-1} \sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y}), \quad k = 1, 2, 3, \dots \quad (3.17)$$

and c_0 is the variance, derived by

$$c_0 = T^{-1} \sum_{t=k+1}^T (y_t - \bar{y})^2 \quad (3.18)$$

for the T observations of the process y_t with sample mean \bar{y} . The PACF is the autocorrelation between y_t and its lagged process y_{t+k} while excluding all autocorrelations ranging from y_{t+1} to

y_{t+k-1} . These are estimated with the use of the derived autocorrelations r_j where the following function is first needed

$$r_j = \hat{\phi}_{k1}r_{j-1} + \hat{\phi}_{k2}r_{j-2} + \dots + \hat{\phi}_{kk}r_{j-k}, \quad j = 1, 2, 3, \dots, k \quad (3.19)$$

These target is to extract the parameters $\hat{\phi}_{11}, \hat{\phi}_{22}, \dots, \hat{\phi}_{kk}$ where $\hat{\phi}_{jj}$ is the partial autocorrelation for lag j ([2] Box and Jenkins, 1976). The correlogram includes plots of the sample ACF and the sample PACF both against the time lags k . For complete randomness, the values at all lags should be zero. The most important part of the identification procedure is to look for significant lags meaning the ones lying outside the interval $\pm 2\sqrt{T}$ (23 Hamilton, 1994). Further, the ACF and PACF are used to find appropriate orders for the SARIMA model by using the results in Table 1. It is however assumed that the differencing implied by d and D have been done to the time series, meaning that the SARIMA $(p, d, q) \times (P, D, Q)_s$ has taken its equivalent representation as a SARMA $(p, q) \times (P, Q)_s$ of the differenced series ([2] Box and Jenkins, 1976).

Table 1: ACF and PACF to identify the orders of SARIMA $(p, q) \times (P, Q)_s$, only positive lags are of interest.

Model	ACF	PACF
AR(p)	Exponentially decreasing or damped sine wave	Spikes to lag p then zero
MA(q)	Spikes to lag q then zero	Exponentially decreasing or damped sine wave
ARMA(p, q)	Exponentially decreasing or damped sine wave after $q - p$ lags	Exponentially decreasing or damped sine wave after $p - q$ lags
SAR(P)	Exponentially decreasing or damped sine wave for all lags times s	Spikes to lag P_s then zero
SMA(Q) _s	Spikes for lag Q_s then zero	Exponentially decreasing or damped sine wave
SARIMA(P,Q) _s	Exponentially decreasing or damped sine wave after lags $(Q - P)_s$	Exponentially decreasing or damped sine wave after lags $(P - Q)_s$

However, this procedure is considered subjective for mixed and seasonal processes. To make the model selection less subjective some frequently used likelihood based information criterions are applied. These are the Akaike information criterion (AIC) ([24] Akaike, 1974), the AIC with correction (AICc) for small samples and the Bayes information criterion (BIC) ([25] Schwarz, 1978). It has been shown that AIC has a tendency to choose a model that is over-parametrized. Further, [26] Burnham and Anderson (2004) suggest that AIC and AICc should be valued over BIC and [12] Brockwell and Davis (1991) propose that AICc is most fit for selecting orders of SARIMA models. Thus, the model selected by AICc will be most valued. However, finding the model that minimizes each criterion would be very time consuming if every model is compared. Fortunately, [27] Hyndman and Khandakar developed an algorithm in 2008 that can be used to speed up this selection process. They suggest an iterative time-saving procedure where the model with the smallest value of some information criterions AIC, AICc or BIC will be found much faster,

since it is now found without comparing every possible model. To derive these information criterions the first thing that is needed is the likelihood function, $L(\hat{\Omega})$ where $\hat{\Omega}$ is the maximum likelihood estimates of the parameters for the SARIMA with $n = p + q + P + Q + 1$ parameters and sample size T . The criterions are then derived by the following equations:

$$\begin{aligned} AIC &= -2 \log[L(\hat{\Omega})] + 2n \\ AIC_c &= \frac{2n(n+1)}{T-n+1} \\ BIC &= -2 \log[L(\hat{\Omega})] + n \log T \end{aligned} \quad (3.20)$$

All non-invertible or non-causal models are rejected. These are found by computing the roots of the lag polynomials $\phi(B)\Phi(B)$ and $\theta(B)\Theta(B)$, if any root is smaller than 1.001 then the model is rejected. If errors arise when fitting the model with the non-linear optimization routine then the model is rejected. It is now assumed that a tentative SARIMA $(p, d, q) \times (P, D, Q)$ s has been identified. Thus, the next step is to estimate it with maximum likelihood estimation approach and then perform diagnostic checks on the residuals. For a good fit these residuals should be distributed as Gaussian white noise, that is, be random, homoscedastic and normal. The diagnostic checking is first performed visually with the standardized residuals and sample autocorrelation function. The Box-Ljung test which is used to test the randomness and serial independence; and the Jarque-Beratest which tests the normality of the residuals are also tested. For this Box-Ljung test, the first step is to extract the residuals $\hat{\varepsilon}_t$ for the fitted model. The T residuals are then used to derive the sample autocorrelations of the residuals with the following equation:

$$\hat{r}_k = \frac{\sum_{t=k+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_{t-k}}{\sum_{t=1}^T \hat{\varepsilon}_t^2}, \quad k = 1, 2, \dots \quad (3.21)$$

This equation is used until a set of autocorrelations $\hat{r}_1, \hat{r}_2, \dots, \hat{r}_m$ have been obtained. These are then used to test the null hypothesis of serially independent residuals versus the alternative hypothesis that they are not serially independent with the following test statistic ([28] Ljung and Box, 1978):

$$\hat{Q}(r) = T(T+2) \sum_{k=1}^m (T-k)^{-1} r_k^2 \quad (3.22)$$

which for an appropriate model was shown to be asymptotically distributed as a $\chi_{1-\alpha}^2(m)$ where m is the number of lagged autocorrelations included and α is the selected significance level. The number of lags should be a function of T , for example, the truncated value of $m = \sqrt{T}$ and that the degrees of freedom should be corrected for SARIMA models to $df = m - p - q - P - Q$. The critical value is included in ([28] Ljung and Box, 1978) and then compared to the value of the test statistic. The null hypothesis of randomness is rejected for large values of the test statistic.

The Jarque-Bera test is used to test the normality of the residuals. The null hypothesis of the test is normality and it is tested against the alternative hypothesis of non-normality. The statistic [29] Jarque and Bera (1980) is written as

$$JB = T \left(\frac{S^2}{6} + \frac{K^2 - 3}{24} \right) \quad (3.23)$$

where T is the number of observations, S is the skewness which is the standardized third moment which if positive decides that it is more likely to observe a bigger number compared to the mean, $S = [E(Y_t - \mu)^3] / [var(Y_t)^{3/2}]$, and the kurtosis which is the standardized fourth moment shows the relationship between the amount of density close to the mean and in the tails. If it is bigger than 3 then we have comparably more density in the tails compared to the normal distribution,

$K = [E(Y_t - \mu)^4]/[var(Y_t)^2]$. The statistic is assumed to be distributed as a $\chi^2_{1-\alpha}(2)$ variable and the null hypothesis of normality is rejected for large values of the statistic ([17] Schwert, 2009). The model should pass all these diagnostic checks to be considered well-fitted and appropriate for forecasting.

In forecasting with the SARIMA $(p, d, q) \times (P, D, Q)_s$ model, the variable $y_t, t = 1, 2, \dots, T$ can be written equivalently as an ARMA $(p + sP, q + sQ)$ for y_t^* from Equation (2.2). The forecast function for the assumed stationary variable y_t^* is then written as

$$(\hat{y}_{t+1/t}^* - \mu) = \phi(B)^*(\hat{y}_t^* - \mu) + \theta(B)^*\hat{\varepsilon}_t \quad (3.24)$$

where $\hat{\varepsilon}_t = y_t - \hat{y}_{t+1/t}^*$ ([23] Hamilton, 1994). The forecast for lead time γ , meaning the time that follows after the last observed information, is then derived by

$$(\hat{y}_{t+s/t}^* - \mu) = \phi(B)^*(\hat{y}_{t+s-1}^* - \mu) + \theta(B)^*\hat{\varepsilon}_{t+s-1} \quad (3.25)$$

That is, the forecast for lead time γ will be derived by the previously observed values of y_t^* , previous forecasts of y^* and the residuals $\hat{\varepsilon}_t$ which have been derived for all time points up to the last observed observation but are equal to zero for the ones where the real values have not yet been observed. ([23] Hamilton, 1994).

4 Results and Discussion

The monthly data of rainfall of the five cities of the South-West Region of Nigeria: Ibadan and Ikorodu respectively in Oyo and Lagos States covered a period of 1973-2018, Osogbo and Abeokuta respectively in Osun and Ogun States covered a period of 1976-2018, while Akure in Ondo State covered a period of 1982-2018. These data, collected from website of the Nigerian Meteorological Agency (NIMET) were plotted as shown in Figures 1-5.

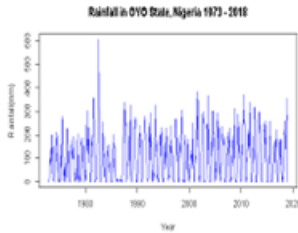


Fig. 1: Time Plots for monthly rainfall in Ibadan

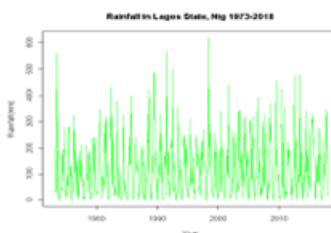


Fig. 2: Time Plots for monthly rainfall in Ikorodu

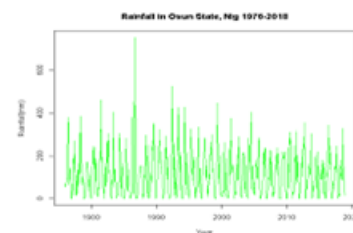


Fig. 3: Time Plots for monthly rainfall in Oshogbo

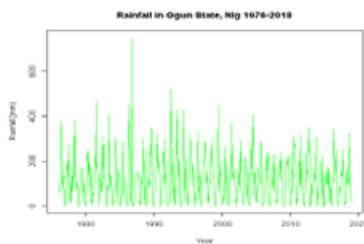


Fig. 4: Time Plots for monthly rainfall in Abeokuta

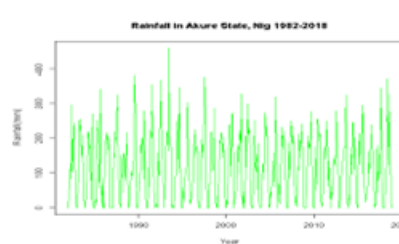


Fig. 5: Time Plots for monthly rainfall in Akure

Figures 1-5 reveal that the time plots are not stationary and may have heavy seasonality due to the fluctuating pattern in the series within the sampled periods. These non-stationarity may be due to irregular variation, seasonal variation and/or other cyclic fluctuations present in the data. These call for transformation of the data as presented on Figures 6-10.

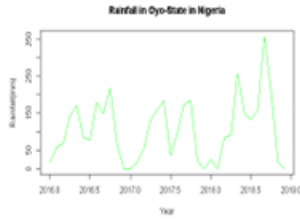


Fig. 6: 1st Differenced for monthly rainfall in Ibadan



Fig. 7: 1st Differenced for monthly rainfall in Ikorodu



Fig. 8: 1st Differenced for monthly rainfall in Oshogbo

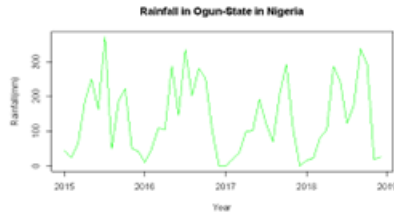


Fig. 9: 1st Differenced of monthly rainfall in Abeokuta



Fig. 10: 1st Differenced for monthly rainfall in Akure

The plots of the first differenced series show a cyclic and seasonal pattern movement from the beginning of 2016 to the beginning of 2019. This implies that the transformed series seem to present some stability in means with approximately constant variances. The model(s) selection and some tests were performed and presented in Tables 2 and 3.

Table 2: Model Identification using Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Test

City	Model	AIC	BIC
Abeokuta	ARIMA(1,0,0)(1,1,0)[12]	7186.598	7203.837
	ARIMA(0,0,1)(1,1,0)[12]	7231.425	7319.623
	ARIMA(1,0,1)(1,1,0)[12]	7074.516	7103.789
	ARIMA(2,0,2)(2,1,0)[12]*	6990.795	7020.965
	ARIMA(3,0,2)(1,0,0)[12]	7689.110	7723.651
Akure	ARIMA(1,0,0)(1,1,0)[12]	5998.132	6138.616
	ARIMA(0,0,1)(1,1,0)[12]*	4885.371	4901.643
	ARIMA(1,0,1)(1,1,0)[12]	5983.113	6321.527
	ARIMA(2,0,2)(2,1,0)[12]	7001.082	7026.942
	ARIMA(3,0,2)(1,0,0)[12]	6942.324	6968.183
Ibadan	ARIMA(1,0,0)(1,1,0)[12]	6298.125	6318.216
	ARIMA(0,0,1)(1,1,0)[12]	6745.561	6784.640
	ARIMA(1,0,1)(1,1,0)[12]*	6231.599	6244.903
	ARIMA(2,0,2)(2,1,0)[12]	7044.606	7066.155
	ARIMA(3,0,2)(1,0,0)[12]	6895.997	6926.166
Ikorodu	ARIMA(1,0,0)(1,1,0)[12]	8567.456	8991.342
	ARIMA(0,0,1)(1,1,0)[12]	8114.536	8347.512
	ARIMA(1,0,1)(1,1,0)[12]	7555.123	7885.451
	ARIMA(2,0,2)(2,1,0)[12]	6881.458	6915.937
	ARIMA(3,0,2)(1,0,0)[12]*	6642.654	6677.134
Osogbo	ARIMA(1,0,0)(1,1,0)[12]*	5936.742	5942.637
	ARIMA(0,0,1)(1,1,0)[12]	6594.011	6778.016
	ARIMA(1,0,1)(1,1,0)[12]	6715.683	6728.624
	ARIMA(2,0,2)(2,1,0)[12]	7147.119	7169.669
	ARIMA(3,0,2)(1,0,0)[12]	7896.233	7962.413

Table 2 displays some set of competing models identified and selected using KPSS test based on Akaike information criterion (AIC) and Bayesian information criterion (BIC). The asterisks (*) are the SARIMA models selected as the best for the five purposively selected cities from the five states of South West Region of Nigeria with minimum values of AIC and BIC. Having selected the models, the stationarity, normality and serial correlation tests were conducted and presented on Table 3.

Table 3: Tests for Stationary, Normality and Serial Correlation on Selected Models

City	Model	ADF	Jarque-Bera	Ljung-Box (\hat{Q}_r)
Abeokuta	ARIMA(2,0,2)(2,1,0)[12]	-7.2291 lag 8 (0.01)	$\chi^2 = 80.65$ $df = 2$ ($< 2.2e^{-16}$)	$\chi^2 = 16.86$ $df = 20$ (0.6619)
Akure	ARIMA(0,0,1)(1,1,0)[12]	-7.6892 lag 7 (0.01)	$\chi^2 = 24.49$ $df = 2$ ($< 4.81e^{-16}$)	$\chi^2 = 32.47$ $df = 20$ (0.5361)
Ibadan	ARIMA(1,0,1)(1,1,0)[12]	-7.5079 lag 8 (0.01)	$\chi^2 = 24.49$ $df = 2$ ($< 24.81e^{-16}$)	$\chi^2 = 20.05$ $df = 20$ (0.4551)
Ikorodu	ARIMA(3,0,2)(1,0,0)[12]	8.9076 lag 8 (0.01)	$\chi^2 = 155.7$ $df = 2$ ($< 2.2e^{-16}$)	$\chi^2 = 96.16$ $df = 20$ ($6.10e^{-12}$)
Osogbo	ARIMA(1,0,0)(1,1,0)[12]	-7.5258 lag 8 (0.01)	$\chi^2 = 300.71$ $df = 2$ ($< 2.2e^{-16}$)	$\chi^2 = 29.50$ $df = 20$ (0.07843)

The bracketed values are the P-values; df = degree of freedom

From Table 3, the Augmented Dickey Fuller shows that there is significant stationarity in the models, since the p-value (0.01) is lesser than alpha level of significance (0.05). The bracketed results of the

Jarque-Bera tests are relatively small indicating that the null hypothesis of normally distributed residuals cannot be rejected for any of the models at 5% significance level. The Box-Ljung is also performed for each model with the number of autocorrelation lags specified as and with degrees of freedom correction for the number of parameters being estimated. The null of randomness cannot be rejected for any of the residuals on the 5% significance level, which is good. Thus, all models seem well-fitted enough and should be suitable for forecasting.

Table 4: Parameter Estimation for Selected SARIMA Models

City	Models	Parameters	Estimate	Std. error	97.5% CI Lower Upper	Other Model Diagnostics
Abokuta	ARIMA(2,0,2)(2,1,0)[12]	$\text{ar1}(\hat{\phi}_1)$ $\text{ar2}(\hat{\phi}_2)$ $\text{ma1}(\hat{\theta}_1)$ $\text{ma2}(\hat{\theta}_2)$ $\text{var}(\hat{\sigma}_1)$ $\text{var}(\hat{\sigma}_2)$ Drift	-0.2449 0.6091 0.2912 -0.6826 -0.6045 -0.2779 0.0559	0.5229 0.5019 0.5116 0.5095 0.0441 0.0440 0.1557	-0.8777 0.9880 -0.1741 2.0110 -3.199 -0.029 -1.2658 -0.784 -0.9908 0.178 -3.641 -1.917 -2.470 0.547	$R^2 = 0.907$ $\text{LogLik} = -2927.12$ $\text{RMSE} = 79.11$; $\text{MAE} = 56.90$; $\text{MPE} = -1402.79$; $\text{MASE} = 0.8290$; $\text{ACP} = 0.033$
Benue	ARIMA(0,0,1)(1,1,0)[12]	$\text{ma1}(\hat{\theta}_1)$ $\text{var}(\hat{\sigma}_1)$ Drift	0.0620 -0.5919 0.0058	0.0917 0.0418 0.1912	0.932 1.655 -0.152 -4.806 -3.695 -3.805	$R^2 = 0.875$ $\text{LogLik} = -2458.66$ $\text{RMSE} = 67.21$; $\text{MAE} =$ 48.34 ; $\text{MPE} = 1609.20$; $\text{MASE} = 0.8172$; $\text{ACP} = 0.036$
Ibadan	ARIMA(1,0,1)(1,1,0)[12]	$\text{ar1}(\hat{\phi}_1)$ $\text{ma1}(\hat{\theta}_1)$ $\text{var}(\hat{\sigma}_1)$ Drift	0.8084 0.2618 0.4144 0.0258	0.1450 0.1616 0.0392 0.2197	2.215 0.7995 -0.995 -0.349 -4.912 -0.576 -3.591 1.665	$R^2 = 0.837$ $\text{LogLik} = -3111.8$ $\text{RMSE} = 75.83$; $\text{MAE} =$ 52.45 ; $\text{MPE} = 1920.87$; $\text{MASE} = 0.8846$; $\text{ACP} = 0.025$
Kogi	ARIMA(3,0,2)(1,0,0)[12]	$\text{ar1}(\hat{\phi}_1)$ $\text{ar2}(\hat{\phi}_2)$ $\text{ar3}(\hat{\phi}_3)$ $\text{ma1}(\hat{\theta}_1)$ $\text{ma2}(\hat{\theta}_2)$ $\text{var}(\hat{\sigma}_1)$	0.1301 0.0404 0.0316 1.0645 0.9640 0.6783	0.0852 0.0482 0.0475 0.0270 0.0271 0.0645	0.459 2.545 -0.840 1.348 -0.611 -1.245 -2.0172 -1.9115 0.118 1.0180 0.519 0.847	$R^2 = 0.935$ $\text{LogLik} = -3317.53$ $\text{RMSE} = 66.54$; $\text{MAE} = 71.24$; $\text{MPE} = 307.06$; $\text{MASE} = 0.5759$; $\text{ACP} = 0.060$
Oyo	ARIMA(1,0,0)(1,1,0)[12]	$\text{ar1}(\hat{\phi}_1)$ $\text{var}(\hat{\sigma}_1)$ Drift	0.0173 -0.4753 0.0173	0.0446 0.0391 0.2254	-0.700 1.047 -3.501 -3.986 -4.851 4.205	$R^2 = 0.821$ $\text{LogLik} = -2964.37$ $\text{RMSE} = 66.45$; $\text{MAE} =$ 55.56 ; $\text{MPE} = 1652.06$; $\text{MASE} = 0.8756$; $\text{ACP} = 0.006$

Table 5: Parameter Estimation for Selected AFRIMA Models

City	Parameters	Estimate	Std. error	Z-value	P> Z	AIC	Log-lik
Abokuta	d	2.997e ⁰¹	8.200e ⁰¹	3.655e ⁰¹	< 2e ⁻¹⁶ **	6361.336	-3175
	ar.ar1	7.955e ⁰¹	7.709e ⁰¹	1.032e ⁰¹	< 2e ⁻¹⁶ **		
	ar.ar2	-1.770e ⁰¹	9.995e ⁰¹	-1.771e ⁰¹	< 2e ⁻¹⁶ **		
	ar.ar3	4.094e ⁰¹	1.959e ⁰¹	2.090e ⁰¹	0.036699*		
	ar.ar4	-1.155e ⁰¹	3.255e ⁰¹	-3.554e ⁰¹	0.000442***		
	ar.ar5	-2.191e ⁰¹	1.350e ⁰¹	-1.647e ⁰¹	< 2e ⁻¹⁶ **		
	ma.ma1	6.730e ⁰¹	7.059e ⁰¹	9.555e ⁰¹	< 2e ⁻¹⁶ **		
Akure	d	0.0000	6.215e ⁰¹	4.145e ⁰¹	< 2e ⁻¹⁶ **	5662.556	-2658
	ar.ar1	0.918	5.755e ⁰¹	1.632e ⁰¹	< 2e ⁻¹⁶ **		
	ar.ar2	-0.208	5.991e ⁰¹	-1.799e ⁰¹	< 2e ⁻¹⁶ **		
	ar.ar3	-0.011	2.829e ⁰¹	2.765e ⁰¹	0.009688*		
	ar.ar4	-0.145	4.255e ⁰¹	-3.164e ⁰¹	< 2e ⁻¹⁶ **		
	ar.ar5	-0.188	2.350e ⁰¹	-2.349e ⁰¹	< 2e ⁻¹⁶ **		
	ma.ma1	0.672	5.205e ⁰¹	5.721e ⁰¹	< 2e ⁻¹⁶ **		
Ibadan	d	3.267e ⁰¹	7.525e ⁰¹	3.927e ⁰¹	< 2e ⁻¹⁶ **		
	ar.ar1	5.199e ⁰¹	6.124e ⁰¹	1.856e ⁰¹	< 2e ⁻¹⁶ **		
	ar.ar2	-2.525e ⁰¹	5.467e ⁰¹	-2.657e ⁰¹	< 2e ⁻¹⁶ **		
	ar.ar3	6.709e ⁰¹	1.255e ⁰¹	5.195e ⁰¹	0.0014992*		
	ar.ar4	-2.164e ⁰¹	5.291e ⁰¹	-5.522e ⁰¹	0.000927***		
	ar.ar5	-5.921e ⁰¹	1.856e ⁰¹	-2.473e ⁰¹	< 2e ⁻¹⁶ **		
	ma.ma1	7.655e ⁰¹	6.152e ⁰¹	9.355e ⁰¹	< 2e ⁻¹⁶ **		
Ikorodu	d	4.555e ⁰¹	2.465e ⁰¹	2.000e ⁰¹	0.039	6759.516	-3361
	ar.ar1	1.055e ⁰¹	1.852e ⁰¹	5.697e ⁰¹	4.990e ⁰¹		
	ar.ar2	-2.854e ⁰¹	3.922e ⁰¹	-7.226e ⁰¹	< 2e ⁻¹⁶ **		
	ar.ar3	4.966e ⁰¹	3.992e ⁰¹	1.244e ⁰¹	0.214		
	ar.ar4	-2.925e ⁰¹	2.417e ⁰¹	-1.209e ⁰¹	< 2e ⁻¹⁶ **		
	ma.ma1	2.703e ⁰¹	6.446e ⁰¹	4.195e ⁰¹	< 2e ⁻¹⁶ **		
	ma.ma2	-2.505e ⁰¹	1.565e ⁰¹	-1.603e ⁰¹	< 2e ⁻¹⁶ **		
	ma.ma3	6.003e ⁰¹	2.402e ⁰¹	5.522e ⁰¹	< 2e ⁻¹⁶ **		
	ma.ma4						
ma.ma5							
Oshodi	d	0.075	6.515e ⁰¹	3.825e ⁰¹	< 2e ⁻¹⁶ **	6119.65	-3052
	ar.ar1	0.546	7.246e ⁰¹	2.802e ⁰¹	< 2e ⁻¹⁶ **		
	ar.ar2	0.205	6.949e ⁰¹	-2.127e ⁰¹	< 2e ⁻¹⁶ **		
	ar.ar3	-0.005	1.876e ⁰¹	3.109e ⁰¹	0.0028541*		
	ar.ar4	-0.036	4.852e ⁰¹	-5.481e ⁰¹	0.000412***		
	ar.ar5	-0.216	1.973e ⁰¹	-1.754e ⁰¹	< 2e ⁻¹⁶ **		
	ma.ma1	-0.867	7.954e ⁰¹	9.305e ⁰¹	< 2e ⁻¹⁶ **		

Asterisks (*) indicates statistically significant

The fitted SARIMA models are found in Table 4. The parameters of the selected SARIMA models seem to be statistically significant considering the low standard error values of their parameter and the estimates lie within the 97.5% confidence intervals. Also, the autocorrelation function (ACF) values for the models are low suggesting statistically good fit for the models. Thus, the mean absolute scale error (MASE), mean absolute error (MAE), root mean square error (RMSE) and Error of Autocorrelation function (ACF) are also checked to ascertain the predictive power of these models. The lower these values are, the more these models can be relied upon in making evidence-based decision and prospective planning. The log-likelihood values are very large confirming the predictive power of the models. All models were proven to be quite well-fitted according to the results of the diagnostic checks. The next step is to extract the residuals and first perform a visual evaluation of each model fit and then execute some statistical tests considering their normality and randomness. Plots of the standardized residuals are found in Figures 11, 12, 13, 14 and 15 for all cities. The residuals seem quite homoscedastic and have a mean very close to zero, thus we move on to forecasting. Table 5 reveals that the log likelihood are high and the p-values of all the parameters of the different models are lesser than the level of significance alpha (0.05). Hence, the parameters of the models are all statistically significant except for Ikorodu where the parameters (d, ar.ar1, ar.ar3) are not statistically significant. These AFRIMA models, though statistical significant, could not capture the empirical features (seasonality and cyclical) in the selected rainfall data. Thus, the paper uses SARIMA models for forecasting the rainfall. The forecast values in the

selected cities in South West Region of Nigeria are presented in the Appendix

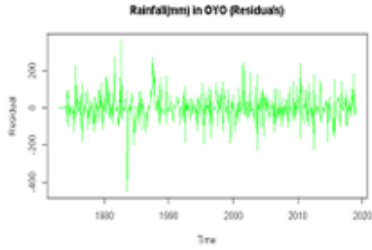


Fig. 11: Residual Plot of Ibadan

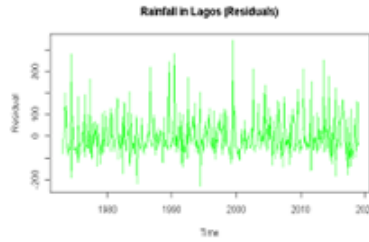


Fig. 12: Residual Plot of Ikorodu

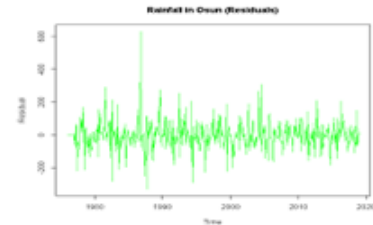


Fig. 13: Residual Plot of Oshogbo

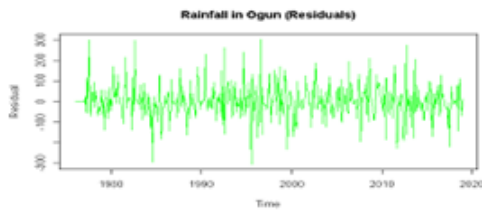


Fig. 14: Residual Plot of Abeokuta

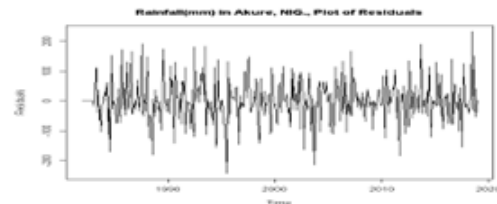


Fig. 15: Residual Plot of Akure (Ondo)

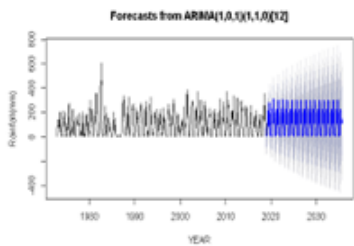


Fig. 16: Forecast Plots of Ibadan

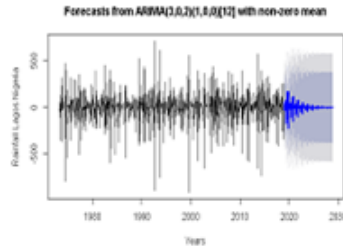


Fig. 17: Forecast Plots of Ikorodu

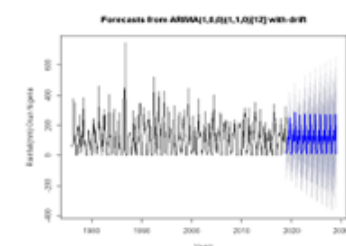


Fig. 18: Forecast Plots of Oshogbo

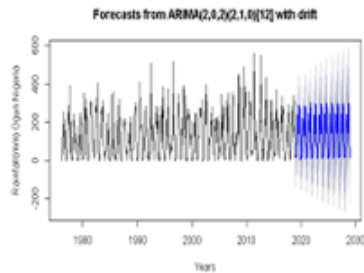


Fig. 19: Forecast Plots of Abeokuta

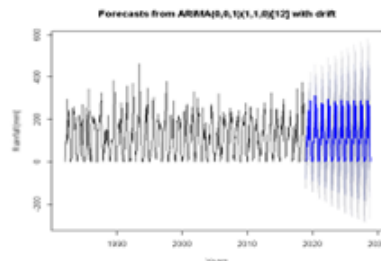


Fig. 20: Forecast Plots of Akure

According to the graphical review of the forecasts, both models seem to produce approximately equal forecasts. That is, the model being more accurate seem to differ from month to month. However, the error measures gave an indication for the SARIMA models producing better forecasts. It can be noted that the values for all the models are within 95% confidence limits. This means that all models performed better compared to the no-change model which just guesses that the next observation will be equal to the present one. That this measure is smaller than one can probably be seen as the minimum requirement for the models being considered to have some kind of predictive ability.

Conclusion

This paper studied the properties, pattern and trend of rainfall in five agricultural dependent cities in South West Region of Nigeria. The time series data were modelled using the techniques of fractional autoregressive integrated moving average (ARFIMA). The adopted models with minimum value of information criterions are considered as the best ([24] Akaike, 1974). In addition, MSE, RMSE, MAE, MPE were also employed to authenticate the accuracy, adequacy and predictive efficiency of the model as a function of reliability on the estimate of the models. The ACF plots of the residuals of the models were examined to see whether the residuals of the model were white noise. Finally, the seasonal ARIMA (SARIMA) has proven to be the best technique for forecasting rainfall in the cities investigated by providing optimal stochastic forecast models. The SARIMA models identified and selected are therefore recommended as very useful techniques that will assist the government, farmers and policymakers in decision making and management of rainfall to increase agricultural yields and set up priorities for equipping themselves against upcoming weather changes.

Acknowledgements

The authors are grateful to the authors referred in this work.

References

- [1] Usoro, A. E., Awakessien, C. E. Time series modelling of rainfall data in different locations in Nigeria. *Journal of Fundamental and Applied Statistics*, 14(1 2), 1-22 (2018).
- [2] Box, G., Jenkins, G. *Time Series Analysis: Forecasting and Control* (Revised Edition). Holden-Day, Oakland, California (1976).
- [3] Tariq, M. M., Abbasabd, A. I. Time series analysis of Nyala rainfall using ARIMA method. *SUST Journal of Engineering and Computer Science*. 17(1), 5–11 (2016).
- [4] Valipour M. Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorological Applications*. 22(3), 592–598 (2015).
- [5] Umar, M. B., Jorg, K., Heiko, B. Spatial-temporal variation and prediction of rainfall in north-eastern, Nigeria. *Climate*(2), 206-222 (2014).
- [6] Abdulrahim, M. A., Fabiyi, I. P., Ismaila, A. Time series analysis of mean monthly rainfall for drought management in Sokoto, Nigeria. *Ethiopian J. Environ. Stud. Managem.*, 6(5), 461 – 470 (2013).
- [7] Etuk, E. H., Moffat, I. U., Chims, B. E. Modelling monthly rainfall data of Port Harcourt, Nigeria by seasonal Box-Jenkins Methods. *International Journal of Sciences*, 2, 60-67 (2013).
- [8] Ekwe, M. C., Johnah , K. J., Johnson, E. I. Mathematical study of monthly and annual rainfall trends in Nasarawa State, Nigeria. *IOSR J. Math.*, 10(1), 56-62 (2014).
- [9] Yaya, O. S., Gil-Alana, L. A., Akomolafe, A. A. Long memory, seasonality and time series trends in the average monthly rainfall in major cities of Nigeria. *CBN Journal of Applied Statistics*, 6(2), 39-58 (2015).
- [10] Guifen, C., Xingmei, X., Guowei, W. The corn output in a time series prediction model. *Institute of Electrical and Electronics Engineers (IEEE)*, 59, 30-501, China (2010).
- [11] Olatayo, T. O., Taiwo, A. I. Statistical Modelling and Prediction of Rainfall Time Series Data. *Global Journal of Computer Science and Technology:G Interdisciplinary*, 14(1), 11-79 (2014).

- [12] Brockwell, P. Davis, R. Time Series: Theory and Methods. Springer, New York, 2nd Edition (1991).
- [13] Box, G., Jenkins, G., Reinsel, G. Time Series Analysis: Forecasting and Control. Wiley, Oxford, 4th Edition (2008).
- [14] Kwiatkowski D., Phillips P.C.B., Schmidt P., Shin Y. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54, 159–178 (1992).
- [15] Dickey, D.A., Fuller, W.A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, Vol. 74 Iss. 366, 427–431 (1979).
- [16] Banerjee, A., Dolado, J., Galbraith, J., Hendry, D. Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data. Oxford University Press Inc., New York (1993).
- [17] Schwert, G. (2009). EViews 7 User's Guide II. Quantitative Micro Software. LLC, Irvine, California.
- [18] Hylleberg, S. Tests for seasonal unit roots: general to specific or specific to general? *Journal of Econometrics*, 69, 5-25 (1995).
- [19] Canova, F., Hansen, B. Are seasonal patterns constant over time? A test for seasonal stability. *Journal of Business and Economic Statistics*, 13(3):237-252 (1992).
- [20] Hylleberg, S., Engle, R., Granger, C., Yoo, B. Seasonal integration and cointegration. *Journal of Econometrics*, 44(1-2), 215-238 (1990).
- [21] Franses, P. Seasonality, non-stationarity and the forecasting of monthly time series. *International Journal of Forecasting*, 7(2), 199-208 (1991).
- [22] Harvey, A. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge, 1st Edition (1989).
- [23] Hamilton, J. Time Series Analysis. Princeton University Press, Princeton, New Jersey (1994).
- [24] Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723 (1974).
- [25] Schwarz, G.E. Estimating the dimension of a model. *Annals of Statistics*, 6(2), 461–464 (1978).
- [26] Burnham, K., Anderson, D. Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33(2), 261-304 (2004).
- [27] Hyndman, R., Khandakar, Y. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, 27(3), 1-22 (2008).
- [28] Ljung, G. M., Box, G.E.P. On a measure of a lack of fit in time series models. *Biometrika*. 65(2), 297- 303 (1978).
- [29] Jarque, C., Bera, A. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3), 255-259 (1980).

Appendix: Forecast values for all Cities in five States of South West of Nigeria

SARIMA(1,0,1)(1,1,0)[12] Ibadan(Oyo State)

	Point Forecast	Lo 95	Hi 95
Jan 2020	19.2632036	-160.043910	198.5703
Feb 2020	4.3095391	-176.085822	184.7049
Mar 2020	76.1800941	-104.492196	256.8524
Apr 2020	101.2123908	-79.530570	281.9554
May 2020	233.3532054	52.592197	414.1142
Jun 2020	157.7195282	-23.046090	338.4851
Jul 2020	109.5532452	-71.213550	290.3200
Aug 2020	142.4292498	-38.337847	323.1963
Sep 2020	312.1727085	131.405535	492.9399
Oct 2020	203.3087541	22.541561	384.0759
Nov 2020	18.8036070	-161.963591	199.5708
Dec 2020	0.1242001	-180.642999	180.8914
Jan 2021	17.4345771	-196.469353	231.3385
Feb 2021	5.5438226	-209.884119	220.9718
Mar 2021	74.1218247	-141.693683	289.9373
Apr 2021	104.0894497	-111.824946	320.0038
May 2021	226.4491421	10.509494	442.3888
Jun 2021	160.1588667	-55.787231	376.1050
Jul 2021	102.7078329	-113.239912	318.6556
Aug 2021	138.0104130	-77.937753	353.9586
Sep 2021	299.0877106	83.139437	515.0360
Oct 2021	201.5227632	-14.425538	417.4711
Nov 2021	19.0099090	-196.938399	234.9582
Dec 2021	0.1312957	-215.817014	216.0796

SARIMA(3,0,2)(1,0,0)[12] Ikorodu(Lagos State)

	Point Forecast	Lo 95	Hi 95
Jan 2020	39.0735298	-408.2169	486.3640
Feb 2020	-16.6535733	-523.8933	490.5862
Mar 2020	33.7146691	-482.6532	550.0825
Apr 2020	-23.0467657	-539.4930	493.3995
May 2020	56.2118652	-460.2395	572.6632
Jun 2020	-116.6145702	-633.0750	399.8459
Jul 2020	46.2711145	-470.1898	562.7320
Aug 2020	-11.4286502	-527.8895	505.0322
Sep 2020	116.7216580	-399.7393	633.1826
Oct 2020	-81.3467555	-597.8077	435.1142
Nov 2020	-155.6052381	-672.0662	360.8557
Dec 2020	120.1301217	-396.3308	636.5910
Jan 2021	26.5035453	-497.6453	550.6523
Feb 2021	-11.2963020	-559.9858	537.3932
Mar 2021	22.8686153	-529.7245	575.4617
Apr 2021	-15.6328220	-568.2596	536.9939
May 2021	38.1285253	-514.5004	590.7575
Jun 2021	-79.1001212	-631.7330	473.5328
Jul 2021	31.3856868	-521.2474	584.0187
Aug 2021	-7.7522223	-560.3853	544.8808
Sep 2021	79.1724843	-473.4606	631.8056
Oct 2021	-55.1778655	-607.8109	497.4552
Nov 2021	-105.5475986	-658.1807	447.0855
Dec 2021	81.4844546	-471.1486	634.1175

SARIMA(1,0,0)(1,1,0)[12] Osogbo(Osun State)

Point Forecast Lo 95 Hi 95

Jan 2020	16.803677	-175.323660	208.9310
Feb 2020	22.052325	-170.081273	214.1859
Mar 2020	137.184100	-54.949499	329.3177
Apr 2020	80.891939	-111.241661	273.0255
May 2020	119.279352	-72.854248	311.4130
Jun 2020	171.325596	-20.808004	363.4592
Jul 2020	107.001073	-85.132527	299.1347
Aug 2020	95.605905	-96.527695	287.7395
Sep 2020	285.767253	93.633653	477.9009
Oct 2020	199.514066	7.380466	391.6477
Nov 2020	40.967212	-151.166388	233.1008
Dec 2020	10.843420	-181.290180	202.9770
Jan 2021	14.221690	-216.423600	244.8670
Feb 2021	26.723344	-203.932543	257.3792
Mar 2021	129.318454	-101.337437	359.9743
Apr 2021	75.582207	-155.073684	306.2381
May 2021	119.579940	-111.075950	350.2358
Jun 2021	168.455134	-62.200757	399.1110
Jul 2021	111.522660	-119.133231	342.1785
Aug 2021	99.406316	-131.249575	330.0622
Sep 2021	269.215667	38.559777	499.8716
Oct 2021	205.827985	-24.827905	436.4839
Nov 2021	39.348201	-191.307689	270.0041
Dec 2021	9.044115	-221.611775	239.7000

SARIMA(2,0,2)(2,1,0)[12] Abeokuta(Ogun State)

	Point Forecast	Lo 95	Hi 95
Jan 2020	22.875661	147.467189	193.2185
Feb 2020	14.556221	-155.919138	185.0316
Mar 2020	78.478807	-92.225857	249.1835
Apr 2020	93.722528	-77.096852	264.5419
May 2020	221.423394	50.463890	392.3829
Jun 2020	205.210375	34.162190	376.2586
Jul 2020	154.774574	-16.364078	325.9132
Aug 2020	129.990384	-41.213262	301.1940
Sep 2020	285.372466	114.108525	456.6364
Oct 2020	282.772426	111.462109	454.0827
Nov 2020	70.326087	-101.025136	241.6773
Dec 2020	9.364165	-162.019698	180.7480
Jan 2021	25.492447	-160.871928	211.8568
Feb 2021	16.071719	-170.400238	202.5437
Mar 2021	85.735122	-100.956184	272.4264
Apr 2021	94.867312	-91.922747	281.6574
May 2021	247.621485	60.700383	434.5426
Jun 2021	210.793479	23.794023	397.7929
Jul 2021	160.598486	-26.484445	347.6814
Aug 2021	145.969660	-41.171434	333.1108
Sep 2021	302.653091	115.456785	489.8494
Oct 2021	282.298689	95.060608	469.5368
Nov 2021	61.061308	-126.214075	248.3367
Dec 2021	12.837710	-174.467175	200.1426

SARIMA(0,0,1)(1,1,0)[12] Akure(Ondo State)

	Point Forecast	Lo 95	Hi 95
Jan 2020	3.1766128	-145.028244	151.3815
Feb 2020	2.7625137	-145.493416	151.0184
Mar 2020	94.8481219	-53.407808	243.1041
Apr 2020	87.4079898	-60.847940	235.6639
May 2020	160.2956319	12.039702	308.5516
Jun 2020	313.5004007	165.244471	461.7563
Jul 2020	114.6611249	-33.594805	262.9171
Aug 2020	272.7976374	124.541708	421.0536
Sep 2020	254.7735112	106.517582	403.0294
Oct 2020	156.6940747	8.438145	304.9500
Nov 2020	22.2013270	-126.054603	170.4573
Dec 2020	0.2473783	-148.008551	148.5033
Jan 2021	5.0223167	-174.169661	184.2143
Feb 2021	4.2975167	-175.003174	183.5982
Mar 2021	109.6099104	-69.690780	288.9106
Apr 2021	78.5150493	-100.785641	257.8157
May 2021	164.3285262	-14.972164	343.6292
Jun 2021	276.1375191	96.836829	455.4382
Jul 2021	109.0034050	-70.297285	288.3041
Aug 2021	239.5427336	60.242043	418.8434
Sep 2021	273.1768038	93.876113	452.4775
Oct 2021	146.1609527	-33.139738	325.4616
Nov 2021	25.3915591	-153.909131	204.6922
Dec 2021	0.3227702	-178.977920	179.6235