GSJ: Volume 10, Issue 8, August 2022, Online: ISSN 2320-9186

www.globalscientificjournal.com

# A COMPARATIVE STUDY OF SARIMA AND SARFIMA MODELS: AN APPLICATION TO SOLAR RADIATION DATA

**UDO M. E. and Shittu O. I.**
**Department of Statistics, University of Ibadan.**
***Corresponding author: oi.shittu@hotmail.com**

## ABSTRACT

The applications of dynamic models AR, MA or ARMA to time series data are very common in time series analysis. In applying these models the stationarity condition of the dataset need to be considered, if the series is non-stationary then ARIMA model is applied for such series. This work therefore focuses on comparing the fitness and predictive ability of $SARIMA(p,d,q)(P,D,Q)_{12}$ and $SARFIMA(p,d,q)(P,D,Q)_{12}$ models to investigate the stationarity, estimate the parameters, select the appropriate model of solar radiation and make forecasts for the two models using monthly solar radiation data. The data is well knows as an environmental data that are usually seasonal and sometime decays slowly to zero on the time plot. Hence, the need to consider $SARIMA(p,d,q)(P,D,Q)_{12}$ and $SARFIMA(p,d,q)(P,D,Q)_{12}$ models. From the literature $SARIMA(p,d,q)(P,D,Q)_{12}$ and $SARFIMA(p,d,q)(P,D,Q)_{12}$ models are not been used to model solar radiation data. The data used in this study are monthly solar radiation data in Uyo, Akwa Ibom State of Nigeria for 32years (1989-2020), Port Harcourt, Ibadan and Sokoto for 5years (2011-2015), respectively. Akaike Information Criteria (AIC) is used to examining the goodness of fit between the two models and root mean square error (RMSE), mean square error (MSE), mean absolute percentage error (MAPE) and mean absolute error (MAE) are used to measure forecasts performance.

Results showed that $SARFIMA(p,d,q)(P,D,Q)_{12}$ model has a better goodness of fit than $SARIMA(p,d,q)(P,D,Q)_{12}$ model in all the cities considered. The forecasts performance measures prove that $SARIMA(p,d,q)(P,D,Q)_{12}$ model has better predictive ability in Uyo, Port Harcourt and Sokoto than $SARFIMA(p,d,q)(P,D,Q)_{12}$ model with exception in Ibadan that $SARFIMA(p,d,q)(P,D,Q)_{12}$ model outperformed $SARIMA(p,d,q)(P,D,Q)_{12}$ model.

The forecasts performance of monthly solar radiation were obtained in Uyo, Port Harcourt and Sokoto using the performance toolswere made using $SARIMA(p,d,q)(P,D,Q)_{12}$ model while $SARFIMA(p,d,q)(P,D,Q)_{12}$ models was better in Ibadan being a dataset that exhibit long memory.

**Key words:** Solar Radiation, Seasonality, Long Memory, SARIMA, SARFIMA, Time Series, Prediction.

## 1.0 Introduction

Autoregressive Integrated Moving Average (ARIMA) model and Autoregressive Fractionally Integrated Moving Average (ARFIMA) model are statistical tools for modeling economic, finance and environmental data. Just as we noticed that Therefore, time is an essential parameter when making plans, so it very important for engineers, geographers, agriculturists and water resource managers to consider time for factors (i.e. dry or rainy season) in their projects.

These data are sometime non stationary at levels, $I(0)$; may require first order or $d^{th}$ differencing (higher order $I(d)$ integration) for $d \geq 1$ to achieve the desired level of stationarity. The autocorrelation function that declines linearly to zero after $I(1)$ or $d^{th}$ differencing is known as first order or $d^{th}$ integrated series while the autocorrelation function (ACF) declines exponentially to d=0 is the stationary $I(0)$ series. In most cases, time series data exhibit neither of these characteristics but have dependencies between the intervals of their observation even after differencing.

ARIMA model proposed by [1] has the capacity to models short range dependent data while long range dependence also known as fractionally integrated series of data, in this case $d$ is not an interger value ($d < 1$) integration can be modeled using ARFIMA model introduced by [2]. A series exhibiting fractionally integrated pattern is characterized with a stable average sequence of long swings. This is captured from ACF declining very slowly over time [3].

The fractional order model technique known as ARFIMA model is the generalization of ARIMA and autoregressive moving average (ARMA) models a conventional integer models. The extension of ARIMA model called Seasonal Autoregressive Integrated Moving Average (SARIMA) model that supports univariate time series data with seasonal component [4]. Long memory time series dataset, especially high frequency trading data, hydrology and network traffic and so on are widely modeled using ARFIMA model. In fact, most of the time series data observations exhibit long memory, these behaviors lead to the development of methodologies that can estimate and predict the autocorrelation function decaying very slowly to zero. One of the best known classes of long memory models is ARFIMA model and its extension, Seasonal Autoregressive Fractionally Integrated Moving Average (SARFIMA) model that can take care of seasonal component was introduced by [5].

[6] test is used to investigate unit root test while [7] test was developed for fractional integrated series because of its sensitivity to long memory. In general, ARFIMA estimators of $d$ are parametric and semi-parametric methods. Parametric method can estimate all the parameters simultaneously while the semi-parametric method estimate in two steps.

Worldwide, time series analyses of solar radiation data are useful in predicting long term average performance of solar energy system [8]. There are several statistical models but the most important thing is to have the appropriate model for solar radiation data. The proposed models for the study of solar radiation data in this work are SARIMA and SARFIMA.

The work is structured as follows: Section 2, present the literature review of the related methodologies and the dataset applied in this paper which comes immediately after this introduction, Section 3 discusses the data and the method, Section 4 renders the results obtained and interpretation, Section 5 gives summary and conclusion.

## 2.0 Review of Relevant Literature

ARIMA model to investigate cryptocurrency [9] exchange rate in high volatility environment. [10] showed the efficiency of the different methods used to test and estimate fractional parameter $d$ in the fractionally integrated autoregressive moving average (ARFIMA) model. [11] predict the incidence of hemorrhagic fever with renal syndrome (HFRS) in Weifang between January 1, 2015 and December 31, 2018 using SARFIMA model and compared with SARIMA model.

A studied on exchange rate of UK pound/US dollar was conducted by [12] for a period January 1971 to December 2008 using *ARMA* and *ARFIMA* models. [13], [14], [15] used both ARIMA and ARFIMA models in their research work.

[16] studied RMB exchange rate by building a nonlinear combination model of the autoregressive fractionally integrated moving average (ARFIMA) model, the support vector machine (SVM). model, and back -propagation neural network (BPNN) model to forecast the RMB exchange rate. An investigation was also carried out by [17] on the existence and non-existence of long memory in the Nigerian and US inflation using some standard tests.

The literature reviews related to solar energy can be found in [18], [19], [20], [21], [22], [23].

## 3.0 Data and Methodology

The data used for this study are secondary data collected from meteorological center, Department of Geography and Regional Planning, University of Uyo, Uyo Akwa Ibom for Uyo as a monthly data series and Nigerian Meteorological Agency (NIMET) Oshodi Lagos for Port Harcourt, Ibadan and Sokoto as daily averages solar radiation data [24]. The data for Uyo span 32 year (1989-2020) and that of Port Harcourt, Ibadan and Sokoto spans 5 year (2011-2015). The measurement units for these data are in milliliters for Uyo and Port Harcourt, Ibadan and Sokoto in Watts per square meters (1ml to 13.153 W/m$^2$). It is used in this work as monthly solar radiation for all the cities.

### Methods

SARIMA models are used to incorporate cyclic components in the modeling time series and it's expressed as

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D X_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \qquad (1)$$

Where $d$ and $D$ are, respectively, differencing and seasonal differencing parameters, $s$ is the seasonal period, $B$ is the lag operator, $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$,

$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$, $\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_{2s} B^{2s} - \cdots - \Phi_{Ps} B^{Ps}$ and $\Theta_Q(B^s) = 1 - \Theta_s B^s - \Theta_{2s} B^{2s} - \cdots - \Theta_{Qs} B^{Qs}$ are the polynomials of orders $p, q, P, Q$, respectively and $\varepsilon_t$ is the error term with zero mean and variance $\sigma_\varepsilon^2$.

SARFIMA model can be expressed as $SARFIMA(p,d,q) \times (P,D,Q)_s$, the general form is given in equation (1) above. These two models have the same expression but there are different in application, the differences lies on their differencing parameters. SARIMA model can only works for integer values differencing while SARFIMA model is applicable to fractional integrated processes.

Seasonal fractional integration can also be defined as

$$y_t = (1-B^s)^D \varepsilon_t \tag{2}$$

The seasonal fractional difference operator $(1-B^s)^D$ is a generalization of the binomial expression and can be written as,

$$(1-B^s)^D = 1 - DB^s - \frac{D(1-D)B^{2s}}{2!} - \frac{D(1-D)(2-D)B^{3s}}{3!} - \cdots$$

Theorem 1: Suppose $\phi_p(B)\Phi_P(B^s) = 0$ and $\theta_q(B)\Theta_Q(B^s) = 0$ in (1) have no common zeroes. Then, the following conditions hold:

(i) The process $X_t$ is stationary if $d + D < 0.5, D < 0.5$ and $\phi_p(B)\Phi_P(B^s) \neq 0$, for $|B| \leq 1$.

(ii) The stationary process $X_t$ has a long memory property if $0 < d + D < 0.5, 0 < D < 0.5$ and $\phi_p(B)\Phi_P(B^S) \neq 0$, for $|B| \leq 1$.

(iii) The stationary process $X_t$ has an intermediate memory property if $-0.5 < d + D < 0, -0.5 < D < 0$ and $\phi_p(B)\Phi_P(B^S) \neq 0$, for $|B| \leq 1$.

(iv) The series; $X_t$ is non-stationary if $0.5 \geq d + D < 1$.

## Test Statistics

In this research, augmented Dickey-Fuller (ADF) and Kwiatkowski, Phillips, Schmidt and Shin (KPSS) tests will be used to investigate linearity assumption on solar radiation data for unit root test and fractional integration.

Augmented Dickey-Fuller has three regression equation models that can be used to test for the presence of unit root. There are:

$$\Delta X_t = \beta X_{t-1} + \varepsilon_t \tag{3}$$

$$\Delta X_t = \mu + \beta X_{t-1} + \varepsilon_t \tag{4}$$

$$\Delta X_t = \mu + \beta t + \varepsilon_t \tag{5}$$

Where $\mu, \beta$ and $b$ are the intercept, the unit root parameter and the linear time trend parameter respectively.

The null and alternative hypothesis are tested base on the $t$-statistic

$$t_{ADF} = \frac{\sum_{i=1}^{p} \hat{\phi}_i - 1}{s.e(\sum_{i=1}^{p} \hat{\phi}_i)} = \frac{\hat{\beta}}{s.e(\hat{\beta})}$$

Where $\hat{\beta}$ is the estimate of $\beta$ in the ADF regressions model and $s.e(\hat{\beta})$ is the estimated standard error [25]. The above $t$-statistic is known as Augmented Dickey Fuller (ADF) unit root test statistic. This work tests for two regression models which are in equations (4) and (5), respectively.

[3] popularized KPSS test, by [6]which was originally designed to test the null hypothesis $I(0)$ against its alternative $I(1)$. They conclude that the test can be used to distinguish between short and long memory stationary processes.

The test residuals $\varepsilon_t$ from a regression of $X_t$; on an intercept and time trend and forming the partial sum $\hat{S}_t = \sum_{i=1}^{N} \hat{e}_t$ of the residuals. The long run variance formula $\sigma_N^2(q)$ [26] is computed as $\sigma_N^2 = b_0 + \sum_{j=1}^{q} w_j\,(q) b_j$ with the conditions that $b_j$ is the $j^{th}$ order sample autocavariance of $y_t$ and $w_j(q)$ are the Bartlett window weights given by $w_j(q) = 1 - j/(q+1)$ for $q < N$. Then, the KPSS test is given as,

$$KPSS = N^{-2} \sum \hat{S}_n^2 \,/\hat{\sigma}_N^2(q)$$

## Estimation of Fractional Difference Parameter

Long memory parameter can be estimate in three major ways: namely; non-parametric method, semi-parametric method and parametric method. But we will only consider the semi-parametric and parametric methods.

## Semi-parametric Method

Semi-parametric method of estimating $d$ in the frequency domain proposes by [25] and [26]. This method considers the power sprectrum of the $ARFIMA\ (p,d,q)$ process, $X_t$ given as,

$$f_X(w) = \left|1 - e^{-iw}\right|^{-2d} f_z(w) \tag{6}$$

Where $f_X(w)$ and $f_z(w)$ are the spectral densities of $X_t$ and $X_z$ respectively, can be simplified as;

$$\log[f_X(w)] = -dlog[4sin^2(w/2)] + \log[f_z(w)] \tag{7}$$

$$\log[f_X(w_t)] = \log[f_z(w_t = 0)] - dlog[4sin^2(w_t/2)] + \{\log[f_z(w_t)] - \log[f_z(w_t = 0)]\} \tag{8}$$

In forms of regression equation we have

$$\log[f_X(w_t)] = \alpha + \beta x_t + \varepsilon_t \tag{9}$$

Where $\alpha = \log[f_z(w_t = 0)]$, $x_t = log[4sin^2(w_t/2)]$, $-d = \beta$

$\varepsilon_t = \{\log[f_z(w_t)] - \log[f_z(w_t = 0)]\}$ is the error in the model for $t = 1,2,\cdots,m$ and $\varepsilon_t \sim N(0, \frac{\pi}{6})$. Therefore, we obtain the estimate of $d$ from the regression equation.

## Parametric Method

Parametric method can be estimated using exact maximum likelihood (EML), nonlinear least squares (NLS) and modified profile likelihood (MPL) estimation methods. [27] recommend MPL or EML for small sample size. This work considered EML.

According to the work of Sowell (1992), let $X_t$ be a sample of $N$ observation such that $X_t = [x_1, x_2, \cdots, x_N]'$. We assume that $X_t$ is a stationary normally distributed fractionally integrated time series. Then, $X_t \sim N(X\beta, \Sigma)$

$$\Sigma = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{N-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{N-1} & \cdots & \gamma_1 & \gamma_0 \end{bmatrix}$$

Where, $\Sigma$ is the covariance matrix and $\gamma_k$ is the auto-covariance function.

Let $Z_N = y_t - x_t^{'}\beta$, then $Z_N = [z_1, z_2, \cdots, z]^{'} \sim N(0, \Sigma)$ with the probability density function.

$$f(Z_N, \Sigma) = (2\pi)^{\frac{-N}{2}} |\Sigma|^{\frac{-1}{2}} \exp\left\{-\frac{1}{2} Z_N^{'} \Sigma^{-1} Z_N\right\} \tag{10}$$

Taking the log likelihood function of equation (10), we have

$\ln\mathrm{L}(d, \phi, \theta, \beta, \sigma^2) = f(Z_N, \Sigma) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma| - \frac{1}{2} Z_N^{'} \Sigma^{-1} Z_N$

Let $\Sigma = \sigma^2 R$, then, the log likelihood function become

$$\ln\mathrm{L}(d, \phi, \theta, \beta, \sigma^2) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|R| - \frac{N}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2} Z_N^{'} \mathrm{R}^{-1} Z_N \tag{11}$$

Differentiating with respect to $\sigma^2$, we have

$$\partial\ln\frac{L}{\partial\sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} Z_N^{'} \mathrm{R}^{-1} Z_N$$

$$\hat{\sigma}^2 = N^{-1} Z_N^{'} \mathrm{R}^{-1} Z_N \tag{12}$$

Then, the concentrated likelihood function is

$$l_c(d, \phi, \theta, \beta) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2} - \frac{1}{2}\ln|R| - \frac{N}{2}\ln[N^{-1} Z_N^{'} \mathrm{R}^{-1} Z_N] \tag{13}$$

Differentiating with respect to $\beta$, gives

$$\hat{\beta} = (X^{'} R^{-1} X)^{-1} X^{'} R X_N \tag{14}$$

$$l_c(d, \phi, \theta) = -\frac{N}{2}(1 + \ln(2\pi)) - \frac{1}{2}\ln|R| - \frac{N}{2}\ln[N^{-1} Z_N^{'} \mathrm{R}^{-1} \hat{Z}_N] \tag{15}$$

Where, $\hat{Z}_N = X_t - X\hat{\beta}$, the function used in maximization method is

$$-\frac{1}{2}\ln|R| - \frac{N}{2}\ln\left(\frac{\hat{Z}_N^{'} - R^{-1}\hat{Z}_N}{N}\right)$$

This function maximized with respect to the elements of $R$, which included $d$, $\phi_P(B)$ and $\theta_q(B)$.

Where, $d$ is the fractional differenced parameter, the parameters of the autoregressive polynomial $\phi_P(B)$ and the moving average polynomial $\theta_q(B)$.

**4. Results and Interpretation**

The summary of statistics of solar radiation data is given in table 1

GSJ: Volume 10, Issue 8, August 2022
ISSN 2320-9186


**Table 1: Descriptive Statistics**

| Statistics | Uyo | Port Harcourt | Ibadan | Sokoto |
|---|---|---|---|---|
| Mean | 10.6115 | 153.2911 | 142.2442 | 235.0097 |
| Median | 10.7000 | 156.3593 | 151.9150 | 237.9562 |
| Maximum | 17.0000 | 231.9561 | 186.7699 | 308.7152 |
| Minimum | 5.9000 | 101.6585 | 73.27386 | 158.5125 |
| Standard Deviation | 1.8225 | 29.8410 | 29.76073 | 34.08001 |
| Skewness | -0.0031 | -0.0475 | -0.585432 | -0.4401 |
| Kurtosis | 3.1868 | 2.3739 | 2.194653 | 2.8863 |
| Jarque-Bera | 0.5591 | 1.0026 | 5.048767 | 1.9695 |
| Probability | 0.7561 | 0.6058 | 0.080108 | 0.3735 |
| Observations | 384 | 60 | 60 | 60 |

From table 1, the average monthly solar radiation data in Uyo, Port Harcourt is about 10.6%. The series are normally distributed for Uyo and Port Harcourt as indicated by the high p-value and also their Jarque-Bera test values are low.

## Time Plot

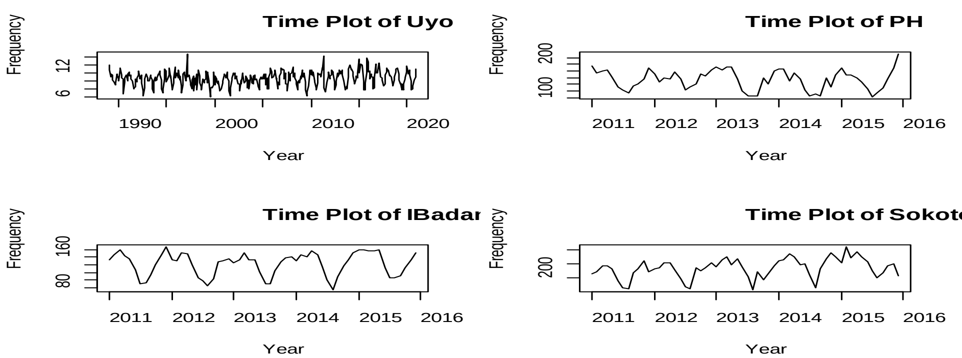


Figure 1: Time Plots of Solar Radiation in four metropolises

The time plots of the original data series (Solar Radiation) for Uyo, Port Harcourt, Ibadan and Sokoto are in figure 1. These plots shows the indication of seasonal variation and non stationary in the series since there is a systematic change in the mean and variance. Also, some spikes in the plotted series in Uyo gives indication of outliers. Hence, we carried out confirmatory test by ploting the ACF and PACF of the series.

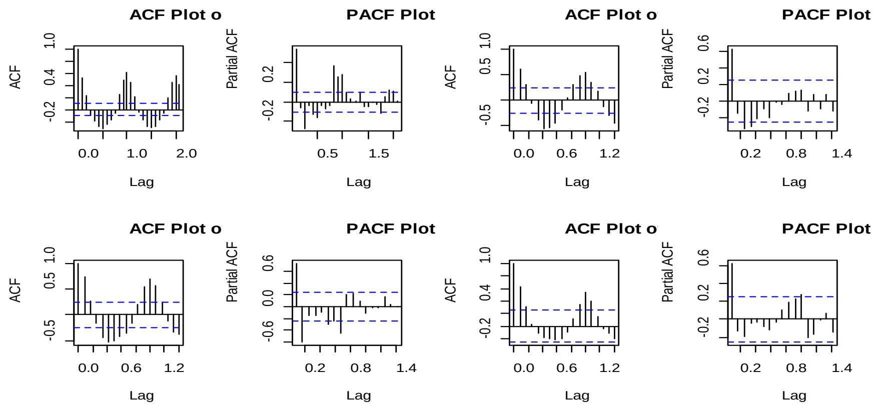


Figure 2: ACF and PACF plots of original series

By visual inspection, ACF plots in figure 2 shows a strong correlation at $s, 2s, 3s, \cdots$, lags where $s = 12$; indicating monthly seasonal variation in the series. The ACF original series are observed to be non-stationary as the spikes decay at a regular pattern and PACF plot are highly significant at lag 1. The ACF plots of solar radiation rata exhibits a slow decay at the seasonal lags which is the behaviour of the seasonal fractionally differenced process. The time series data must undergo transformation to attain stationarity, in order to identify the model for the data.

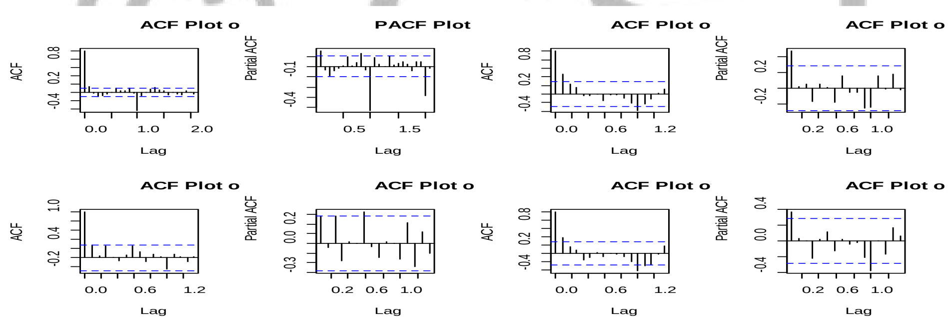


Figure 3: ACF and PACF plot of seasonal differenced series for solar radiation

## Fractional and Unit Root Test

The test of stationarity of solar radiation data are done using Augmented Dickey Fuller (ADF) test and KPSS (Kwiatkowski, Phillips, Schmidt and Shin) test for integer and non-integer difference values, respectively.

The ADF test statistic tests the null hypothesis that the series has a unit root against the alternative of no unit root (stationary). The decision rule is to reject the null hypothesis when the p-value is less than or equal to 0.05.

The KPSS test statistic tests the null hypothesis of stationarity against the alternative that the series has a unit root and to accept the null hypothesis when the test statistic value is less than the critical value.

**Table 2: Fractional and Unit Root test of Solar Radiation Data**

| City | Test | ADF | | | | KPSS |
|---|---|---|---|---|---|---|
| | | Intercept | | Intercept and Trend | | |
| Uyo | Test Statistic | -2.406388 | | -2.958546 | | 0.135396 |
| | Critical values: 1% | -3.447770 | (0.1406*) | -3.982988 | (0.1455*) | 0.216000 |
| | 5% | -2.869113 | | -3.421983 | | 0.146000 |
| | 10% | -2.570871 | | -3.133816 | | 0.119000 |
| PH | Test Statistic | -5.495229 | | -5.364107 | | 0.068295 |
| | Critical values: 1% | -3.555023 | (< 0.01*) | -4.133834 | (0.0003*) | 0.739000 |
| | 5% | -2.915522 | | -3.493692 | | 0.463000 |
| | 10% | -2.595565 | | -3.175693 | | 0.347000 |
| Ibadan | Test Statistic | -0.824709 | | -1.981252 | | 0.034692 |
| | Critical values: 1% | -3.57446 | (0.8030*) | -4.156734 | (0.5968*) | 0.739000 |
| | 5% | -2.923780 | | -3.504330 | | 0.363000 |
| | 10% | -2.5999 | | -3.181826 | | 0.347000 |
| Sokoto | Test Statistic | -3.576119 | | -4.612533 | | 0.036787 |
| | Critical values: 1% | -3.546099 | (0.0092*) | -4.148465 | (0.0027*) | 0.216000 |
| | 5% | -2.911730 | | -3.500495 | | 0.146000 |
| | 10% | -2.59355 | | -3.179617 | | 0.119000 |

From Table 2, using ADF test for Uyo and Ibadan, we fail to reject the null hypothesis that the monthly solar radiation series has a unit root base on p-value decision. Hence, the series required first order difference to obtain stationarity. According to the KPSS test for monthly solar radiation, the results revealed that the time series is neither *I(1)* nor *I(0)* since it is significant at 10% while in Ibadan the test accept the null hypothesis.

For PH and Sokoto data sets, we reject null hypothesis in ADF test while accepting KPSS test that monthly solar radiation data is stationary at level. Further tests are required to assume the order of differentiation.

## Test and Estimation of order of Integration

Semi-parametric estimator proposed by [26] was applied with (GPH). algorithm The application of this test on the series allows us to test the null hypothesis of a unit root ($d = 1$) against the alternative of fraction integration ($d < 1$). The GPH estimator is based on the regression equation using the periodogram function as an estimate of the spectral density. The value of fractional differencing parameter can be adjusted using the bandwidth parameter. In this work, we used default bandwidth of 0.5 and the estimated fractional parameter results are given in Table 3.

**Table 3: Results of GPH estimate for *d* parameter**

| Estimate | Uyo | PH | Ibadan | Sokoto |
|---|---|---|---|---|
| $\hat{d}$ | 0.6275326 | -0.04795349 | -0.1625215 | 0.4220138 |
| sd.as | 0.1874373 | 0.3863428 | 0.3863428 | 0.3863428 |
| sd.reg | 0.1325419 | 0.6116229 | 0.8034448 | 0.3840207 |

In Table 3, $d$ is GPH estimate, Sd.as is the asymptotic standard deviation and Sd.reg is the standard error deviation. The value of the differenced parameter $d$ is $0.6275 > 0.5$, is indication of non stationarity long memory in the series of Uyo data, PH and Ibadan data have intermediate memory properties because $d$ is $< 0$; and Sokoto fractional differenced parameter is $0.4220$ which is stationary long memory property (that is $0 < d < 0.5$).

Based on that result in Table 2, we applied differencing to Uyo and Ibadan series. The differenced test results are given in Table 4.

**Table 4: Fractional and Unit Root test of Solar Radiation for Differenced Series**

| City | Test | ADF | | | | KPSS |
|---|---|---|---|---|---|---|
| | | Intercept | | Intercept and Trend | | |
| Uyo | Test Statistic | -15.02449 | | -15.00591 | | 0.108747 |
| | Critical values: 1% | -3.447770 | (<0.01*) | -3.982988 | (<0.01*) | 0.216000 |
| | 5% | -2.869113 | | -3.421983 | | 0.146000 |
| | 10% | -2.570871 | | -3.133816 | | 0.119000 |
| Ibadan | Test Statistic | -6.561011 | | -7.340125 | | 0.037755 |
| | Critical values: 1% | -3.57446 | (<0.01*) | -4.161144 | (<0.01*) | 0.739000 |
| | 5% | -2.923780 | | -3.506374 | | 0.363000 |
| | 10% | -2.5999 | | -3.183002 | | 0.347000 |

From table 4, ADF test accept the alternative hypothesis that the series is stationary and the KPSS agree with the null hypothesis. Also, there is a needs to carry out fractional seasonal differencing in all the city.

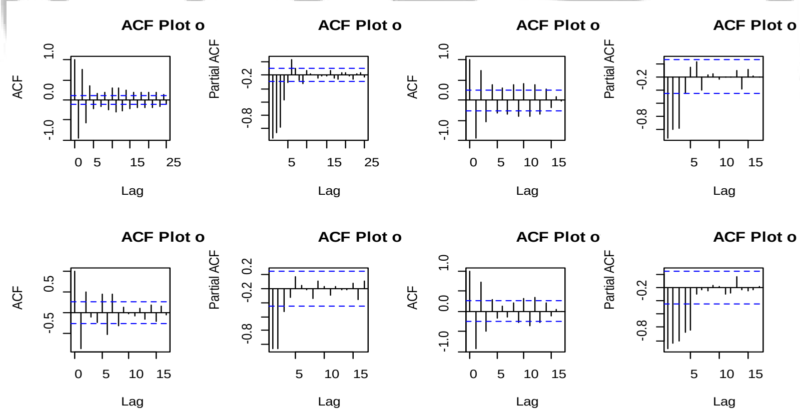


Figure 4: ACF and PACF plots of fractional seasonal differenced

## Construction of the SARIMA model

In this paper, we estimate the parameters of $SARIMA$ model using maximum likelihood estimation procedure. We judge the best model base on AIC value.

**Table 5:** The results for the estimated $\boldsymbol{SARIMA}$ Models in the four Cities

| City | Model | AIC | Log Likelihood |
|---|---|---|---|
| Uyo | $SARIMA(1,1,1)(0,1,1)_{12}$ | 1210.96 | -601.48 |
| | $SARIMA(1,1,1)(0,1,2)_{12}$ | 1210.29* | -600.15 |

| | | | |
|---|---|---|---|
| | $SARIMA(0,1,1)(0,1,1)_{12}$ | 1217.3 | -605.65 |
| PH | $SARIMA(1,0,1)(0,1,1)_{12}$ | 412.88 | -202.44 |
| | $SARIMA(1,0,2)(0,1,1)_{12}$ | 414.70 | -202.35 |
| | $SARIMA(1,0,0)(0,1,1)_{12}$ | 411.57* | -202.78 |
| Ibadan | $SARIMA(1,1,1)(1,1,0)_{12}$ | 375.68* | -183.84 |
| | $SARIMA(0,1,1)(1,1,0)_{12}$ | 377.12 | -185.56 |
| | $SARIMA(1,1,1)(2,1,0)_{12}$ | 377.15 | -183.58 |
| Sokoto | $SARIMA(1,0,1)(0,1,1)_{12}$ | 440.43* | -216.21 |
| | $SARIMA(1,0,2)(0,1,1)_{12}$ | 442.25 | -216.12 |
| | $SARIMA(1,0,0)(0,1,1)_{12}$ | 440.62 | -217.31 |

Table 5 above shows the results of the estimated models of SARIMA, the model with the minimum AIC value is indicated by asterisk and it is considered to be the best model for monthly solar radiation series in the respected Cities. The estimated parameters of those models are given in table 6.

Table 6: Estimate of Parameters of of SARIMA Models for the series.

| City | Models | Parameter | ar1 | ma1 | sma1 | sma2 |
|---|---|---|---|---|---|---|
| Uyo | $SARIMA(1,1,1)(0,1,2)_{12}$ | Estimate | 0.1680 | -0.9423 | -0.8652 | -0.0883 |
| | | s.e. | 0.0566 | 0.0227 | 0.0616 | 0.0549 |
| PH | $SARIMA(1,0,0)(0,1,1)_{12}$ | Estimate | 0.5226 | | -0.9996 | |
| | | s.e. | 0.1368 | | 0.3470 | |
| Ibadan | $SARIMA(1,1,1)(1,1,0)_{12}$ | Estimate | 0.3238 | -0.8963 | -0.4693 | |
| | | s.e. | 0.1705 | 0.0865 | 0.1496 | |
| Sokoto | $SARIMA(1,0,1)(0,1,1)_{12}$ | Estimate | 0.8601 | -0.4561 | -0.6060 | |
| | | s.e. | 0.1622 | 0.2638 | 0.2919 | |

## Construction of the SARFIMA model

In SARFIMA model, the order $(p, d, q)$ and the seasonal components $(P, D, Q)$ are specified same as the $SARIMA$ above. Unfortunately, the method of estimation we used to estimate the order of fractional differencing cannot estimate all parameters in the model simultaneously, and we cannot identify the parameter $d$ and $D$ when using GPH estimation method. Therefore, we used exact maximum likelihood estimator to estimate $SARFIMA$ model.

This method estimates the memory parameter and the parameters of the appropriate $SARFIMA$ Model orders simultaneously. In this work, we choose the best model based on Akaike Information Criterion (AIC). The results are shown in Table 7 and the parameters estimate for the adequate models in table 8

Table 7: Estimate of Parameters of of SARFIMA Models for the series.

| City | Model | AIC | Log Likelihood |
|---|---|---|---|
| | $SARFIMA(1,0.1117,0)(0,0.4560,2)_{12}$ | 166.824* | -76.4122 |
| **Uyo** | $SARFIMA(0,0.0758,1)(1,0.3519,1)_{12}$ | 179.519 | -83.7594 |
| | $SARFIMA(0,0.1364,1)(0,0.4580,2)_{12}$ | 167.354 | -76.6772 |
| | $SARFIMA(1,-0.3975,1)(0,0.3988,1)_{12}$ | 355.763 | -169.881 |
| PH | $SARFIMA(1,-0.4685,2)(0,0.3910,1)_{12}$ | 354.458 | -169.229 |
| | $SARFIMA(1,-0.2316,2)(1,0.4033,1)_{12}$ | 352.429* | -17.214 |
| | $SARFIMA(1,0.1137,1)(1,0.4657,0)_{12}$ | 316.598 | -151.299 |
| Ibadan | $SARFIMA(1,-0.0588,1)(1,0.4491,0)_{12}$ | 315.305 | -151.653 |
| | $SARFIMA(1,-0.8637,1)(2,0.4720,0)_{12}$ | 314.263* | -149.132 |

| | | | |
|---|---|---|---|
| Sokoto | $SARFIMA(1,-0.4374,1)(0,0.3868,1)_{12}$ | 382.695 | -184.347 |
| | $SARFIMA(1,-0.5304,2)(0,0.3843,1)_{12}$ | 384.507 | -184.253 |
| | $SARFIMA(1,-0.3688,2)(1,0.3519,1)_{12}$ | 380.743* | -184.372 |

Table 8: Estimated Parameters of SARFIMA models for the series

| City | Model | Parameter | Estimate | Std. Error | z-value | Pr(>\|z\|) |
|---|---|---|---|---|---|---|
| Uyo | $\boldsymbol{SARFIMA(1,0.1117,0)(0,0.4560,2)_{12}}$ | phi(1) | 0.1346702 | 0.0891628 | 1.51039 | 0.130945 |
| | | theta.12(1) | 0.2984788 | 0.0686513 | 4.34775 | 1.3754e-05 *** |
| | | theta.12(2) | 0.1247637 | 0.0571489 | 2.18313 | 0.029026 * |
| | | d.f | 0.1116659 | 0.0674833 | 1.65472 | 0.097982 . |
| | | d.f.12 | 0.4559956 | 0.0222194 | 20.52245 | < 2.22e-16 *** |
| | | zbar | 10.6114583 | | | |
| | | AIC | 166.824 | | | |
| | | Log Likelihood | -76.4122 | | | |
| | | $\sigma^2$ | 1.41906 | | | |
| PH | $\boldsymbol{SARFIMA(1,-0.2316,0)(0,0.4033,1)_{12}}$ | phi(1) | 0.757209 | 0.136492 | 5.54765 | 2.8953e-08 *** |
| | | theta.12(1) | 0.000000 | NA | NA | NA |
| | | d.f | -0.231608 | 0.192406 | -1.20375 | 0.22869 |
| | | d.f.12 | 0.403324 | 0.039950 | 10.09572 | < 2.22e-16 *** |
| | | zbar | 153.291083 | | | |
| | | AIC | 352.429 | | | |
| | | Log Likelihood | -170.214 | | | |
| | | $\sigma^2$ | 254.446 | | | |
| Ibadan | $\boldsymbol{SARFIMA(1,-0.8637,1)(2,0.4720,0)_{12}}$ | phi(1) | 0.9277675 | 0.0890505 | 10.41844 | < 2.22e-16 *** |
| | | theta(1) | -0.6459958 | 0.1750143 | -3.69110 | 0.00022328 *** |
| | | phi.12(1) | -0.2861464 | 0.1897894 | -1.50771 | 0.13163002 |
| | | phi.12(2) | -0.0284297 | 0.2186778 | -0.13001 | 0.89656080 |
| | | d.f | -0.8636865 | 0.2164948 | -3.98941 | 6.6238e-05 *** |
| | | d.f.12 | 0.4719875 | 0.0214476 | 22.00653 | < 2.22e-16 *** |
| | | zbar | 142.2441710 | | | |
| | | AIC | 314.263 | | | |
| | | Log Likelihood | -149.132 | | | |
| | | $\sigma^2$ | 108.351 | | | |
| Sokoto | $\boldsymbol{SARFIMA(1,-0.3688,2)(1,0.3880,1)_{12}}$ | phi(1) | 0.8795555 | 0.1219016 | 7.21529 | 5.3819e-13 *** |
| | | theta.12(1) | 0.0000000 | NA | | |
| | | d.f | -0.3687582 | 0.2147790 | -1.71692 | 0.085994 . |
| | | d.f.12 | 0.3879681 | 0.0458476 | 8.46213 | < 2.22e-16 *** |
| | | zbar | 235.0096517 | | | |
| | | AIC | 380.743 | | | |
| | | Log Likelihood | -184.372 | | | |
| | | $\sigma^2$ | 418.582 | | | |

## Comparison between $SARIMA$ and $SARFIMA$ models performance

According to table 9 below, SARFIMA model failed to produce better forecast estimates compared to SARIMA indicated by the high values of MSE, RMSE, MAE and MAPE for the cities of Uyo, PH and Sokoto. In Ibadan SARFIMA model has a better forecast estimate than SARIMA model. The smaller the estimated error value, the better the forecasting performance of the model.

**Table 9:** Forecast performance measures for **SARIMA** model and **SARFIMA** model.

| City | Model | MSE | RMSE | MAE | MAPE |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| Uyo | $SARIMA(1,1,1)(0,1,2)_{12}$ | 1.3287 | 1.1527 | 0.8510 | 8.15634 |
| | $SARFIMA(1,0.1117,0)(0,0.4560,2)_{12}$ | 1.4006 | 1.1853 | 0.8940 | 8.6939 |
| PH | $SARIMA(1,0,1)(0,1,1)_{12}$ | 145.4653 | 12.0609 | 8.9849 | 5.9344 |
| | $SARFIMA(1,-0.2316,0)(0,0.4033,1)_{12}$ | 237.0938 | 15.3979 | 12.6610 | 8.3903 |
| Ibadan | $SARIMA(1,1,1)(1,1,0)_{12}$ | 104.5883 | 10.2266 | 7.7164 | 5.7821 |
| | $SARFIMA(1,-0.8637,1)(2,0.4720,0)_{12}$ | 97.51615 | 9.8750 | 7.6418 | 5.6207 |
| Sokoto | $SARIMA(1,0,1)(0,1,1)_{12}$ | 339.4733 | 18.4248 | 13.5063 | 5.6704 |
| | $SARFIMA(1,-0.3688,0)(0,0.3880,1)_{12}$ | 390.6763 | 19.7655 | 16.1129 | 10.8752 |

## Conclusion

This work studies and analyze the nature of monthly solar radiation in four metropolises in Nigeria using SARIMA and SARFIMA processes.

Examining the two models from their AIC values, which are 1210.29, 411.57, 375.68 and 440.43 in SARIMA models while SARFIMA models have 165.492, 352.429, 314.263 and 380.743 for Uyo, PH, Ibadan and Sokoto, respectively. We observed that SARFIMA models have the least values of AIC for solar radiation compare to SARIMA models in all the cities. In terms of forecast, performance measures: MSE, RMSE, MAE and MAPE results show that SARIMA models have the minimum error values in Uyo, PH and Sokoto while SARFIMA model is better for the analysis solar radiation in Ibadan.

Judging from the AIC values of the two appropriate models, we can conclude that SARFIMA was best model for analyzing solar radiation in the four metropolises except in Ibadan. Even though the forecasts SARFIMA models was adjudge to be better that SARIMA in three of the four series, the forecast performance is poor except in Ibadan where long memory was exhibited.

It should also be noticed that good and best fit models may not yield good forecast values for the future.

## References

[1] Box, G. E. P. and Jenkins, G.M. (1976): Time series analysis forecasting and control,2nd ed. Holden-Day. San Francisco.
[2] Granger, C. and Joyeux (1980), "An Introduction to long memory time series models and Fraction differencing". J. Time Series Analysis. 1:15-19
[3] Mills, T. C. and Markellos, R. N. (2008): The Econometric Modelling of Financial Time Series. Third Edition, p 85-89.
[4] Brownlee, J (2018): A Gentle Introduction to SARIMA for Time Series Forecasting in python. https://machinelearningmastery.com.
[5] Porter-Hudak, S. (1990): An application of the seasonal fractionally differenced model to the monetary aggregates. J Am Stat Assoc. 85(410):338-44.
[6] Dickey, D. A. and Fuller, W. A. (1979): Distribution of the Estimators for Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association, 74: 427531

[7] Kwiatkowski, D. P, Schmidt, P. and Shin, Y. (1992): Testing the Null Hypothesis of Stationary against the Alternative of a unit Root, Journal of Econometrics, 54, 159-178.

[8] Jhohura, F. T. and Rayhan, M. I. (2011): An Assesement of Renewable Energy in Bangladesh through ARIMA, Holt's, ARCH-GARCH models. Institute of Statistical Research and Training (ISRT), Dhaka University, Dhaka-1000, Bangladesh.

[9] Nashirah, A. B. and Sofian, R. (2017): Autoregressive Integrated Moving Average (ARIMA) model for Forecasting Cryptocurrency Exchange Rate in High Volatility Environment: A New Insight of Bitcoin Transaction. International Journal of Advanced Engineering Research and Science (IJAERS) https//dx.doi.org/10.22161/ijaers.4.11.20.

[10] Olatayo, T. O. and Adedotun, A.F. (2014): On the Test and Estimation of Fractional Parameter in ARFIMA Model: Bootstrap Approach. Applied Mathematical Sciences, Vol.8, no.96, 4783-4792. HIKARI Ltd, www.m-hikari.com.

[11] Qi, C., Zhang, D., Zhu, Y., Liu, L., Li, C., Wang, Z., and Li, X. (2020): SARFIMA model prediction for infectious diseases: application to hemorrhagic fever with renal syndrome and comparing with SARIMA.

[12] Shittu, O.I. and Yaya, O. S. (2009): An Application to US Dollar/ UK Pound Foreign exchange rate. https://www.ressearchgate.net/publication/200145442.

[13] Adesina, O. S., Onanaye, S. A., Okewole, D. and Egere, A. (2020): Forecasting of New Cases of COVID-19 in Nigeria using Autoregressive Fractionally Integrated Moving Average Models. Doi:10.9734/ARJOM/2020/v16:930226

[14] Titus, C. M., Wanjoya, A. and Mageto, T. (2021): Time Series Modeling of Guinea Fowls Production in Kenya using the ARIMA and ARFIMA models. International Journal of Data Science and Analysis Vol.7, pp. 1-7. Doi: 10.11648/j.ijdsa20210701.11.

[15] Trevisan, L. R. and David, S. A. (2016): Some Comments on Fractionally Integration processes Involving Two Agricultural Commodities. European Scientific Journal ISSN: 1857-7881,(print)e- ISSN: 1857-7431

[16] Xie, C., Mao, Z. and Wang, G. (2015): Forecasting RMB Exchange Rate Based on a Nonlinear Combination Model of ARFIMA, SVM, and BPNN

[17] Yaya, O. S. and Shittu, O. I. (2010): Long Memory and Estimation of memory parameters: Nigerian and US Inflation Rates http://ssrn.com /abstract=1658900.

[18] Amadi, S. O., Dike, T. and Nwokolo, S. C. (2020): Global Solar Radiation Characteristics at Calabar and Port Harcourt Cities in Nigeria. ISSN: 2376-2144 futureenergysp.com/index.php/tre. Doi: 10.1773/tre.2020.6.2.00114.

[19] Dada, BM and Okogbu, EC (2017): Estimating Daily Solar Radiation from Monthly Values Over Selected Nigeria Station for Solar Energy Utilization. J Fundam Renewable Energy App l7: 240. Dio: 10.4172/2090451.1000240.

.

[20] Ibeh, G. F., Agbo, G. A., Ekpe, J. E. and Isikwue, B. C. (2012): Estimation of Global Solar Radiation at Calabar using Two Models. Journal of Natural Sciences Research. ISSN 2224-3186 (paper) ISSN 2225-0921(online) vol.2, No.5, www.iiste.org.

[21] Ibrahim, L. K. Asare, B. K. Yakubu, M. and Dauda, U. (2011): Modelling of Sokoto Daily Average Temperature: A Fractional Integration Approach. Nigerian Journal of Basic and Applied Science (2011), 19(1):21-30, ISSN 0794-5698.

[22] Irfan Uçkan (2018): Analysis of Solar Radiation Data in Van, Turkey. Celal Bayar University Journal of Science Volume 14, Issue 4, p 421-427.

[23] Peng, C., Aichen, N., Duanyang, L., Wei, J. and Bin, M. (2018): Time Series Forecasting of Temperatures using SARIMA: An Example from Nanjing. IOP conf. ser: Mater, Sci. Eng.394 052024. http://dx.doi.org/10.12988

[24] Adejumo, A. O., Suleiman, E. A., Okagbue, H. I.(2017): Exploration of solar radiation data from three geo-political zones in Nigeria. www.elsevier.com/locate/dib

[25] Geweke, J. and Potter-Hudak, S. (1983): The Estimation and Application of Long memory Time series models, Journal of Time Series Analysis, 4: 221-238.

[26] Robinson, P.M. (1995): Log-Periodogram Regression for Time Series with long Range Dependence. Annals of Statistics 23: 1048-1072.

[27] Yaya, O. S. and Shittu, (2009): On the Efficiency of Estimation Techniques for differencing Parameter of Fractional Integrated Time Series in Proceedings of 33rd Annual Conference of Nigerian Statistical Society, Zamfara State, Nigeria.