ARMA MODEL FOR PREDICTION OF SOLAR RADIATION IN KANO-NIGERIA

Ibrahim S. Madugu*[1], B. J. Olufeagba[2], Yinusa A. Adediran[2], A. Abdulkarim[3]
F. Abdulkadir[4], James. U. Inaku[2], Olalekan Ogunbiyi[5], O. Ibrahim[2], A. U. Lawan[6]
and Mubarak A. Afolayan[2]
[1*]Department of Electrical Engineering; Kano University of Sci. & Tech., Nigeria
[2]Department of Electrical and Electronics Engineering, University of Ilorin, Nigeria
[3]Department of Electrical Engineering, Ahmadu Bello University Zaria, Nigeria
[4]Department of Chemistry Education, Kano University of Sci. and Tech., Nigeria
[5]Department of Electrical and Computer Engineering, Kwara State University, Nigeria
[6]Department of Electrical Engineering, Federal Polytechnique Kazaure, Nigeria

Keywords: –
Solar radiation, Time Series Box-Jenkins, Root Mean Square Error, and Sum of Square Error.

Article History: –
Received: January, 2019
Reviewed: February, 2019
Accepted: February, 2019
Published: March, 2019

ABSTRACT

Statistical approach is one of the most prominent ways of forecasting a trend of a time series data such as solar radiation. This paper explore the autoregressive integrated moving average (ARMA) model using Box and Jenkins methodology to determine the most parsimonious model of Kano, Nigeria solar radiation time series. The result obtained showed that ARMA(3,0) has the least value of Root Mean Square Error (RMSE), Sum Of Square (SSE), Mean Absolute Percentage Error (MAPE) and Theil's U-Statistics with 0.0004, 0.0422, 0.0388 and 0.0061 respectively. Therefore ARMA(3,0) is the best model that fits the solar radiation data and hence the most parsimonious.

1. INTRODUCTION

The knowledge of the behaviour of solar radiation is required for the generation of solar energy of a particular location [1]. This is necessary in designing a robust solar autonomous system that will guarantee an almost 100% availability of electricity to the load demand. However, where there is lack of data for the particular region, then a solar radiation data of another region with similar climate could be used from developed empirical equations [2]- [3]. Solar energy microgrid for electricity generation is site specific and requires dependable solar radiation data that can be obtained from meteorological station (such as Nigeria Meteorological Agency, NiMet) or using statistical methods [4]- [5]. The method of using empirical data for solar forecasting requires a very large data usually not less than 5 years and could grapple with other challenges such as: missing data, inaccurate readings due to poor calibration or equipment failure, human errors and so on [1]. However, the statistical method or solar forecasting usually employs the historical data as a base and may present a better and more accurate predictions [6].

Optimum exploitation of solar photovoltaic energy requires appropriate model for the forecasting of the solar radiation and sizing of the solar microgrid components (solar panels, batteries, inverter etc.) [7]. There are various methods of solar forecasting that can be categorized into persistent, satellite data/imagery, numeric weather prediction (NWP), statistical and hybrid methods [8]. Persistent method is the simplest, it correlates past and future and can use short period (1 hour-ahead) [9] - [10]. Statistical method uses mathematical models and can be developed with tools such as autoregressive moving average (ARMA) using a short and medium time period (year, days etc.) [1]. Hybrid method is the combination of more than one method and can easily adjust to any time period [5]. Numerous works were succefully carried out using Box-Jenkins method for solar radiation. In [8], ARMA(2,1) is obtained for the optimum prediction of global solar radiation in Al-Ain, UAE. ARMA Box-Jenkins method is used in [11] and [12] to model global solar radiation data. Stochastic modelling for daily global solar radiation in Marrakesh, Morocco in used in [13].

The fossil fuels such as coal, oil and natural gas was

by far the most common electricity generation plants in the 20th century. This gravely affects the global environment through the greenhouse gas emissions and causing global warming [14]. The United Nations Climate Change Conference (COP21) in 2015 aimed at reducing this menace by encouraging more investments on renewables that are clean and environmentally friendly [15]. Some countries like China and USA even promulgated domestic laws to soothe other nations. The EU (European Union) agreed to increase the renewables energy production to 30% by 2030 and target 100% production by 2050 [15], [16].

Solar and wind energies are the most prominent due to their potentials. The sun that is the major source of solar energy radiates at a surface temperature of about 5000K and providing 1367W/m^2 of solar energy to the earth at the standard temperature condition (STC). The amount of power the earth receives from the sun is approximately 1.8X10^{11}MW. This is more than the amount the world requires [17].

Most works classify solar PV prediction based on historical solar irradiance data and other meteorological data pattern; the forecast methodology and forecasting horizon. Figure 1 shows the PV power base on historical data classification. There are basically 4 categories in this class that involve: persistence, statistical, machine learning and hybrid methods. In the statistical method (where ARMA belongs), the forecast is performed by statistical analysis of the different input variables using time series of the past data [4].

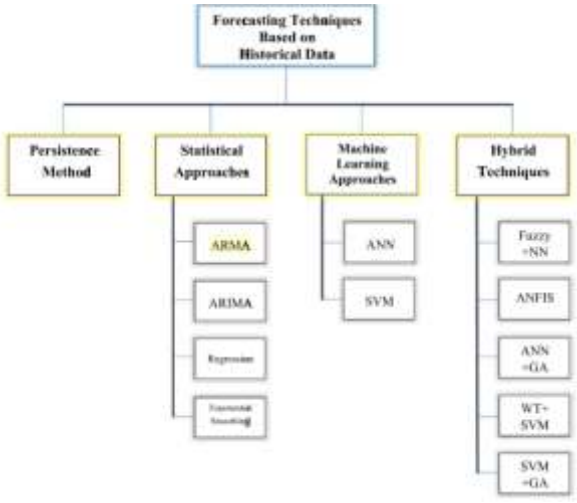


Figure 1. Historical Data Classification of PV Power Forecasting

This work studied and developed a model for the prediction of solar radiation data of Kano (12.05°N; 08.2°E; altitude 472.5 m; 3 air density 1.1705 kg/m3) using time sries ARMA approach. The data used in this study covers 46 years and are obtained from the Nigerian Metrological Agency (NiMet). The selection of ARMA model for the development of forecasting model for the next 20 years (2016 to 2036) is due to its veracity towards a large data.

In section 2 the ARMA model is introduced and a step by step Box and Jenkins methodology are described. Section 3 presents the historical data of the study area solar radiation and other climate variables. In section 4 the results and their discussions were presented.

2. Time Series (ARMA Model)

The ARMA consists of the autoregressive, AR and moving average MA models. The ARMA modeling of any process is given by orders *p* and *q*, which are the combination of the past values of the time series, X_t and past errors. Given by [18]:

$$X_t = \emptyset_1 X_t + \emptyset_2 X_{t-1} + \cdots + \emptyset_p X_{t-p} + e_t - \varphi_1 e_{t-1} - \varphi_2 e_{t-2} \ldots - \varphi_q e_{t-q} \quad (1)$$

where $\emptyset$ and φ are autoregressive and moving average coefficients respectively. While e is the random shock that is assumed to be randomly selected with zero mean and constant variance.

ARMA model has the ability of extracting the statistical behaviour and espouses the Box and Jenkins method [19]. The major draw back of ARMA model (Box-Jenkins) is that the time series must be stationary [20].

A nonstationary time series shows a trend or seasonality, that requires a higher class of model called the autoregressive integrated moving average (ARIMA), deonted as ARIMA(p,d,q). where the d, term indicates the order of differentiation [4].

There are two approaches to identifying the order of d: graphical and test procedure [21]. The graphical method involves the the time plot and the corresponding autocorrelation function, ACF and partial autocorrelation function, PACF plots. The time plot will show a trend and also if the ACF decays very slowly or does not decrease to zero.

In the test procedure a test for the existence of unit root in the time series is carried out. The Augmented Dickey-Fuller test (ADF) uses this procedure to test if the time series is stationary when it has no unit root. This is defined by [22]:

$$\partial X_t = \mu + \gamma t + \phi X_{t-1} + \partial_1 X_{t-1} + \cdots + \partial_p X_{t-p} + e_t \quad (2)$$

where μ is a constant, γ is the trend, p is the autoregressive process order and e_t is a sequence of random variables with zero mean and a unit variance.

The identification of the ARMA model on whether it is purely autoregressive (AR), moving average (MA) or a combination of AR and MA is done using the characteristics of theoretical ACF and PACF. A purely AR process will have its ACF decline exponentially and a PACF that spikes by an order of p while a pure MA process will have its ACF spikes at lags q and an PACF declined exponentially. A mixed process contains a p^{th} order AR and q^{th} order MA components, such process is term ARMA(p, q). However, if the system is de-trended then a 'd' parameter could be added to indicate that the time series was not stationary and the model is called ARIMA(p,d,q) [21].

Identifying the model order may entail trial and error as most data do not exhibit simple pattern. Usually the best method is to look at the plot of ACF and PACF to have a likelihood of the maximum (upper) bound of the parameters, p and q. Then estimates are made on orders less than or equal to the upper values by discriminating among the models. The estimation may involve the use of linear least square estimation, Yule-Walker estimation, maximum likelihood estimation, method of moment etc [23], [24], [22].

Information criteria is used to measure the goodness of fit of the models to select the best fit model or parsimonious model. One of the best way is the use of residual (difference between actual and forecasted data) analysis by plotting the time plot and correlogram (ACF and PACF plots) of the residual. A model with a residual that resembles white noise, having a characteristics of a normal distribution (constant variance and zero mean) fits the data. If the model is adequate the spikes on plot of ACF and PACF of the residuals will fall within the 95% confidence interval bounds. Also, Portmanteau lack-of-fit test could be used defined by [23]:

$$Q = n\sum_{i=1}^{K} e_i \quad (3)$$

where K is taken as the one-quarter of the total number of sample size N of the time series. A correctly selected model order will have Q follows approximately the χ^2 distribution with (K-p-q) degrees of freedom at $N>100$. Moreover, a Ljung-Box-Pierce statistics provides better approximation. Other information criterion used may include the Akaike Information Criterion (equation 4), Schwarz Bayesian Information Criterion (equation 5), Hannan-Quinn information criterion (equation 6) and Durbin-Watson test statistics (equation 7) [21], [25], [26], [27].

$$AIC = \log(\hat{\sigma}^2(k)) + \frac{2}{N}(k) \quad (4)$$

$$SBIC(p,q) = log(\hat{\sigma}^2(k)) + \frac{k}{N} log(N) \quad (5)$$

$$HQIC(p,q) = log\left(\hat{\sigma}^2(k)\right) + \frac{2k}{N} log(log\ (N)) \quad (6)$$

$$V = \frac{\sum_{i=2}^{N}(e_i - e_i - 1)}{\sum_{i=1}^{N} e_i^2} \quad (7)$$

where k is the parameter of parameter estimated and N is the number of observations.

The accuracy of the developed forecast model is tested using the error analysis between the actual data $A(t)$ and the predicted data$F(t)$. There are various tools that check the performance of the ARMA model(s) that may include: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Sum of Square Error (SSE,) Theil's U-statistic etc. These are defined in equations 8 to 12 [28] - [29].

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|A(t) - F(t)| \quad (8)$$

$$MSE = \frac{1}{N}\sqrt{\sum_{t=1}^{N}(A(t) - F(t))^2} \quad (9)$$

$$RMSE = \left(\frac{1}{n}\sqrt{\sum_{t=1}^{n}(A(t) - F(t))^2}\right) \times 100 \quad (10)$$

$$MAPE = \frac{1}{n}\sum_{t-1}^{n}|A(t) - F(t)| \quad (11)$$

$$SSE = \sum_{t-1}^{n}(A(t) - F(t))^2 \quad (12)$$

3. Study Area and Data Analysis

Kano (12.05°N; 08.2°E; altitude 472.5m air density 1.1705 kg/m^3) is the state capital of Kano State in Northern West, Nigeria. It is situated in the Sahelien geographic region, south of the Sahara. Kano is the commercial nerve centre of Northern Nigeria and is the second largest city in Nigeria, after Lagos. The total area of Metropolitan Kano is now 499 square kilometres (193 square miles), with a population of 2,828,861 as of the 2006 Nigerian census [30]. Kano state was created on May 27, 1967 from part of the Northern Region. It has a population of about 9.3 million people (2006 Nigerian census) and is projected to have about 17 million people in 2019 [30].

The Kano metropolis (area covered in this study) is densely populated and heavily industrialized (with more than 500 medium and large scale industries). These industries include textiles and apparels, food and beverages, tanneries, iron smelting, processing, blacksmithing, ceramics, plastics etc. These industries and other business activities acts as a regional hub servicing a market of over 300 million people located in northern Nigeria and the neighbouring countries such as Niger, Chad and Cameroon. These industries and approximately 1.6 million Micro-Small Medium Enterprises require enormous sustainable electricity supply, with a demand of about 300MW [30].

The 46 years (1st January, 1971 to 31st December, 2016) daily mean solar radiation data (MJ/m^2/day) were collected from the Nigerian Metrological Agency (NiMet). Figure 2 presents the daily average of solar radiation in MJ/m^2/day where due to seasonal effects there are two major highs around March and October.

The maximum solar radiation is recorded in March with an average of 25.15MJ/m^2/day and August has the minimum with an average of 19.57MJ/m^2/day as shown in figure 3, showing the monthly average of 46 years data. The months with the lowest solar radiation coincided with the rainy season as such the influence of cloud cover is imminent.

Figure 4 shows the Kano State monthly maximum monthly mean temperature and rainfall. It can be seen that the months with the highest rainfall are June, July, August and September with 132.5mm, 281mm, 323.9mm and 155.8mm respectively. While the months of Mrch, April and May has the highest temperature. It can also be seen that those months with highest rainfall has the lowest solar energy as shown in figure 3.

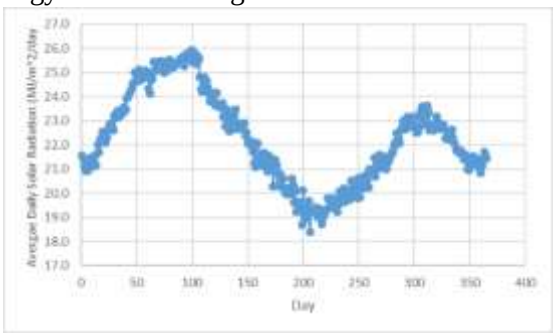


Figure 2. Daily Average Solar Radiation (1971-2016)

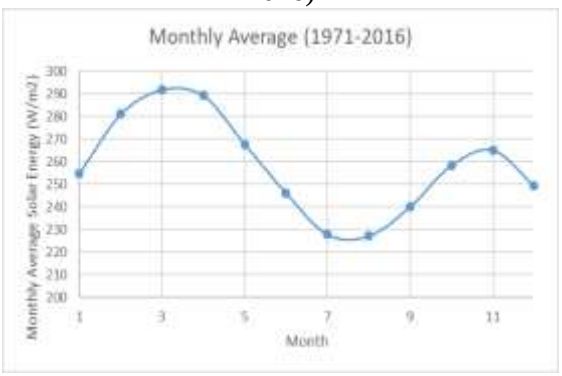


Figure 3. Monthly Average Solar Energy (1971-2016).

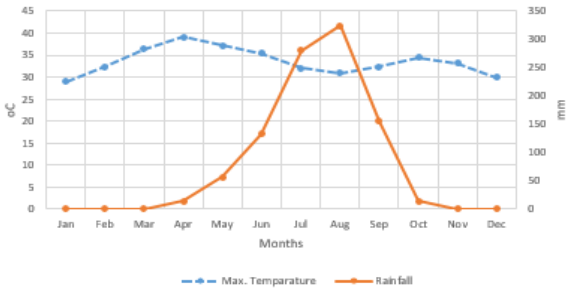


Figure 4. Monthly Maximum Temperature and Rainfall

In figure 5 the yearly minimum, maximum and mean of solar irradiance for the period of 46 years (1971-2016) is presented, where the average, maximum and minimum are 258.4, 400.2 and 129.9W/m^2 respectively.

The histogram of the solar radiation data is shown in figure 6. It can be seen that be seen that the data is normally distributed with a standard deviation of 3.03. The histogram of the quarterly data of 10 years (1971-1980) solar radiation are shown in

figures 7, 8, 9, and 10, for the first, second, third and fourth quarters respectively. It can seen that the first and second quarters show a better distribution while the 3rd and 4th quarters experiences a skewed normal distribution due to wheather turbalance caused by cloud cover in the third quarter and haze in the fourth quarter.



Figure 5. Yearly Solar Irradiance (1971-2016)

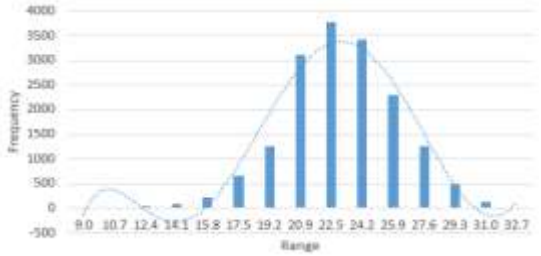


Figure 6. Histogram of the Frequency of the Solar Radiation between 1971 and 2016

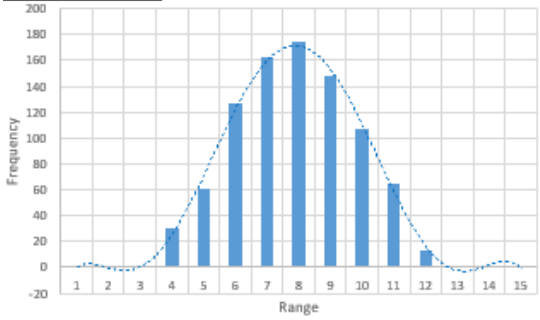


Figure 7. Histogram of the Frequency, 3rd Quarter (1971-1980)

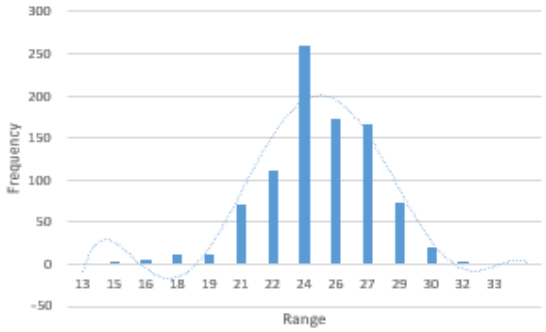


Figure 8. Histogram of the Frequency, 2nd Quarter (1971-1980)

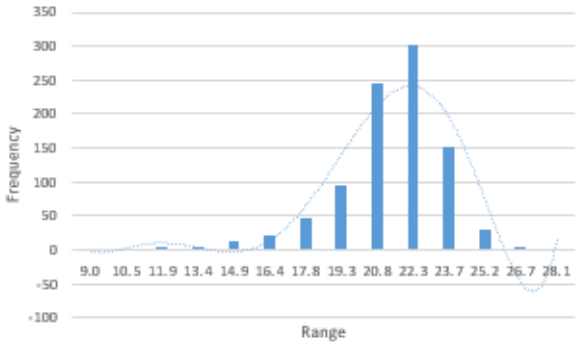


Figure 9. Histogram of the Frequency, 3rd Quarter (1971-1980)

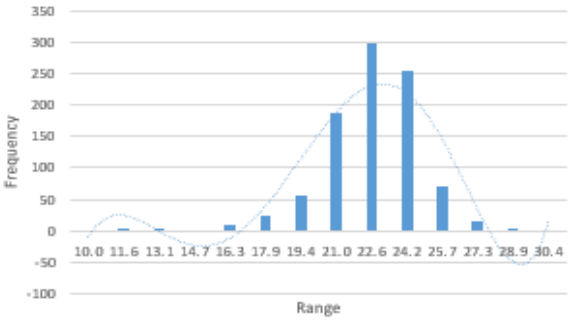


Figure 10. Histogram of the Frequency, 4th Quarter (1971-1980)

4. RESULTS AND DISCUSSION

The trend analysis of the solar radiation data is a nonstationary time series with a slope of -0.0394. However, the first difference of the data showed (Figure 11) a slope of m=0.0004 that is almost zero and makes the time series stationary.

The Autocorrelation Function (ACF) and partial autocorrelation function (PACF) plots shown in figure 12 and 13 respectively show that several lags (k) are within 95% confidence interval (CI) bounds ($\pm 1.96/\sqrt{N}$ or ±0.2889, where N=46), hence a stationary time series. The ACF decays in an oscillation form after the second lag and the PACF plot cuts off just after few lags indicating ARMA(2,0) or higher rank.

Least square parameter estimation method is used to compute several ARMA models; including ARMA(1,0), ARMA(1,1), ARMA(2,0), ARMA(2,1), ARMA(3,0) and ARMA(3,1). In Table I the AIC and mean square error (MSE) of all the models are presented with ARMA(3,0) having the lowest values of AIC and MSE at 15.2006 and

-0.01943. Therefore, ARMA(3,0) is chosen as the most parsimonious model.

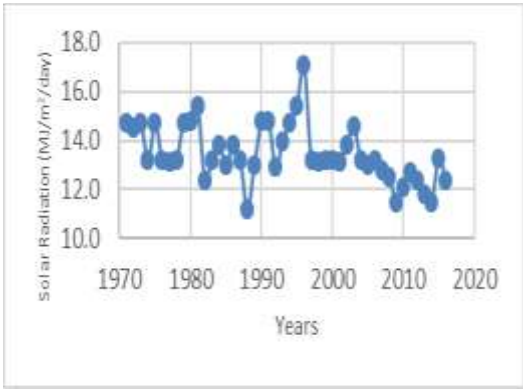


Figure 11. Time Plot for Yearly Minimum Solar Radiation (1971-2016).

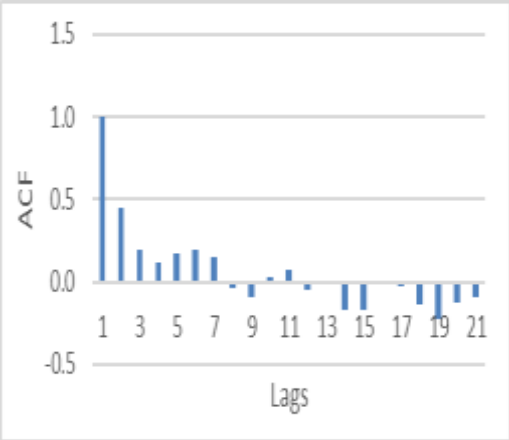


Figure 12. Autocorrelation Function (ACF) of solar radiation (1971-2016) time series.

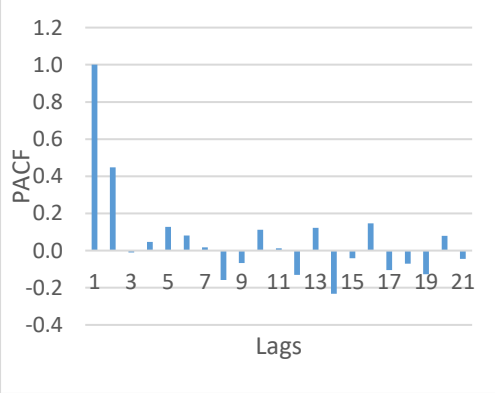


Figure 13. Partial Autocorrelation Function (PACF) of solar radiation (1971-2016) time series.

The ARMA(3,0) model established that the yearly mean solar radiation, y_t depends on its past 3 years values with some residuals 0.4674 (α_1), 0.0466 (α_2) and 0.0820 (α_3) plus a random variable, ε_t of (-0.0843).

The time plot, ACF and PACF of the residual error of the model ARMA(3,0) displayed in figure 14 and figure 15 has almost all of the lags of the correlations within the CI bounds. The time plot of the residual in figure 16 is a behaving like a white noise with approximately zero mean (μ=0.006). This point to the suitability of ARMA(3,0) to the data.

The model ARMA(3,0) is used to forecast the solar radiation data and the result in figure 17 shows a good agreement between the actual data and the forecasted data. The error analysis for all the selected models also show that ARMA(3,0) has the lowest values of MAPE, RMSE and SSE as presented in Table 2.

Table 1: AIC and MSE Values of ARMA Models

	ARMA (1,0)	**ARMA (1,1)**	**ARMA (2,0)**	**ARMA (2,1)**	**ARMA (3,0)**	**ARMA (3,1)**
AIC	18.2776	26.9191	17.1524	26.9191	15.2006	23.4567
MSE	-0.83726	-0. 80962	-0. 81788	-1. 9479	-0. 01943	1. 8346

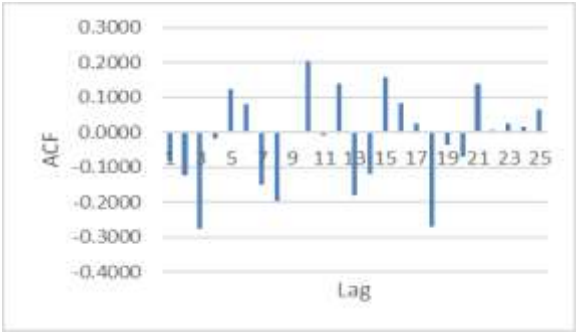


Figure 14. Autocorrelation Function (ACF) Plot of residuals of ARMA (3,0).

The Theil's U-statistics described in equation 18 is another analysis that describe the performance of forecasting model, usually the model with the lowest value is considered the best model. In figure 18 ARMA(3,0) has a value 0.0061. However, it can be seen that ARMA(4,0) has a very slight lower value but is quite insignificant and as such it is not considered.

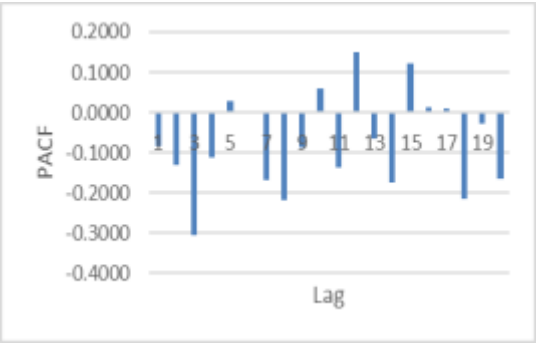


Figure 15. Partial Autocorrelation Function (PACF) Plot of residuals of ARMA (3,0).

Table 2:RMSE, MAPE AND SSE Values of ARMA Models

MODEL	RMSE	MAPE	SSE
ARMA(1,0)	0.0182	1.8201	77.5609
ARMA(1,1)	0.0176	1.7600	77.5762
ARMA(2,0)	0.0178	1.7780	77.5715
ARMA(2,1)	0.0423	4.2347	77.5559
ARMA(3,0)	0.0004	0.0422	0.03885
ARMA(3,1)	0.0399	3.9882	69.9896

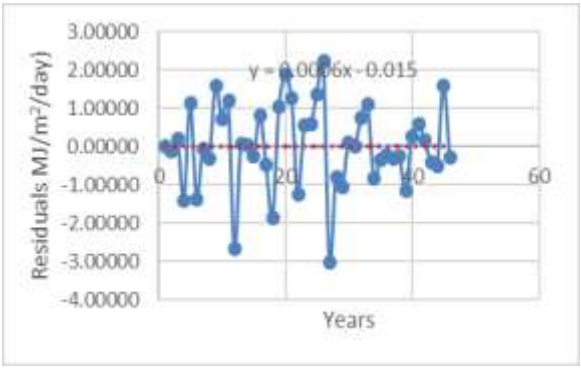


Figure 16. Time Plot of Residuals of ARMA (3,0).

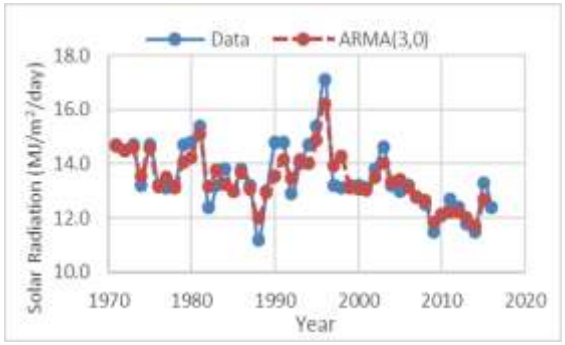


Figure 17. Measured Vs ARMA(3,0) of Mean Yearly Minimum Solar Radiation (1971-2016).

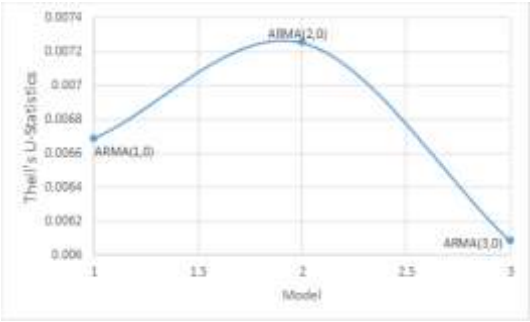


Figure 18. Theil's U-Statistic for ARMA and ARMAP models.

5. CONCLUSION

Result of this work has the maximum mean monthly solar radiation for Kano, Nigeria is 25.15MJ/m^2/day. The data shows a normal distribution and so are the quarterly distributions. The most parsimonious model is ARMA(3,0) having the lowest values of error analysis. Thus having a Theil's-U statistics value of 0.0061 and MAPE value of 0.0422.

REFERENCES

[1] J. Hassan, "ARIMA and regression models for prediction of daily and monthly clearness index," vol. 68, no. 1, 2014.

[2] S. Fritz and T. H. Mcdonald, "Average solar radiation in the United States," *Heat Ventil*, pp. 1-4, 1949.

[3] A. J. Drummond and E. Vowinckel, "The distribution of solar radiation throughout Southern Africa," *Journal of Meteorology*, vol. 14, no. 4, pp. 343-353, 1957.

[4] K. D. Utpal, S. Kok, S. Mehdi, M. Saad, Y. Moh, V. Willem, H. Bend and S. Alex, "Forecasting of photovoltaic power generation and model optimization: A review," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 912-928, 2018.

[5] A. Abdulkarim, S. Abdelkader and D. J. Morrow, "Statistical analyses of wind and solar energy resources for the development of hybrid microgrid," in *2nd International Congress on Energy Efficiency and Energy Related Materials (ENEFM2014) Springer Proceedings in Energy.*, Switzerland, 2015.

[6] M. T. A. N. Y. E. H. S. J. Kudo, "Forecasting electric power generation in a photovoltaic

power system for an energy network," *Electr Eng Jpn*, pp. 16-23, 2009.

[7] C. A. A. M. I. S. I. O. a. A. Y. A. Mekontso, "Review of Optimization Techniques for Sizing Renewable Energy Systems," *Computer Engineering and Applications*, vol. 8, no. 1, pp. 13-30, 2019.

[8] H. N. A. H. Hassan A, "Time-series regression model for prediction of mean daily global solar radiation in Al-Ain, UAE.," *ISRN Renew Energy*, 2012.

[9] S. Ferrari, M. Lazzaroni, V. Piuri, L. Cristaldi and M. Faifer, "Statistical models approach for solar radiation prediction," *IEEE*, 2013.

[10] X. Yang, J. Ren and H. Yue, "Photovoltaic power forecasting with a rough set combination method.," in *11th international conference on control (CONTROL)*, UKACC, 2016.

[11] H. o. W. W. M. S. Z. Sulaiman MY, "Analysis of residuals in daily solar radiation time series," *Renew Energy* , vol. 11, no. 1, pp. 97-105, 1997.

[12] R. A. G. T. S. K. Zaharim A, "Time series analysis of solar radiation data in the tropics," *Eur J Sci Res*, vol. 25, no. 4, pp. 672-8, 2009.

[13] A. M. W. A. Zeroual A, "Stochastic modeling of daily global solar radiation measured in Marrakesh, Morocco," *Renew Energy*, vol. 6, no. 7, pp. 787-93, 1995.

[14] A. L. C. M. J. Zervos, "A 100% renewable energy vision for the European union-re-thinking 2050," EU, 2010.

[15] A. Zervos, "Renewables 2016 Global Status Report," 2016.

[16] REN21, "Renewables 2015 Global Status Report," REN21 Secretariat, Paris, France, 2015.

[17] C. A. Gueymard, "The sun's total and spectral irradiance for solar energy applications and and solar radiation models," *Solar Energy*, vol. 76, pp. 423-453, 2004.

[18] R. Huang, T. Huang and R. Gadh, "Solar generation prediction using the ARMA model in a laboratory-level microgrid," in *IEEE Third International Conference on Smart Grid Communications*, 2012.

[19] J. Boland, Time series modelling of solar radiationModeling solar radiation at the earth's surface, Springer, 2008.

[20] R. Shumway and D. Stoffer, Time Series Analysis, London: Springer, 2011.

[21] F. Pelgrin, "Lecture 5: Box-Jenkins methodology," Nice, 2011.

[22] J. Wu and C. Chan, "The prediction of monthly average solar radiation with TDNN and ARIMA," in *2012 11th International Conference on Machine Learning and Applications*, 2012.

[23] G. E. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, Time series analysis: forecasting and control, John Wiley & Sons, 2015.

[24] G. Box, G. Jenkins and G. Reinsel, "Time series analysis: forecasting and control," New York, USA, 1994.

[25] M. G. Kendall, A. Stuart and J. K. Ord, The Advanced Theory of Statistics, London: Griffin, 1983.

[26] G. Ljung and G. Box, "On a measure of lack of fit in time series models," *Biometrika*, pp. 297-303, 1978.

[27] W. Enders, Applied econometric time series, New York, NY, USA: John Wiley & Sons, 2010.

[28] O. GALIP, H. ARIF and G. ASIR, "Statistical Analysis of Solar Radiation Data," *Energy Sources*, pp. 1088-1097, 2010.

[29] R. McGee and M. Yaffee, Introduction to time series analysis and forecasting, Brooklyn, New York: Academic Press, INC., 1999.