GSJ: Volume 12, Issue 9, September 2024, Online: ISSN 2320-9186
www.globalscientificjournal.com

# Modelling Daily Solar Radiation from Photovoltaic System using ARIMA and ARFIMA Models

**Mfon E. Udo**
Economic and Financial Statistics Unit, Department of Statistics,
University of Ibadan, Ibadan, Nigeria
Email address: mfon_udo2018@yahoo.com

**OlaOluwa S. Yaya**
Economic and Financial Statistics Unit, Department of Statistics,
University of Ibadan, Ibadan, Nigeria
Email address: os.yaya@ui.edu.ng

**David Umolo**
Economic and Financial Statistics Unit, Department of Statistics,
University of Ibadan, Ibadan, Nigeria
Federal Polytechnic Orogun, Orogun, Delta State, Nigeria
Email address: umolo.david@fepo.edu.ng



## Abstract

This paper focuses on using two classical time series models: the Autoregressive Integrated Moving Average (ARIMA) model and Autoregressive Fractionally Integrated Moving Average (ARFIMA) model in modeling and forecasting photovoltaic system (PV) via solar radiation. Semi-parametric and parametric methods are used for model parameters estimation while the Akaike information criteria (AIC) is used to judge the goodness of fit between the two models and forecast performance measures considered are mean square error (MSE), root mean square error (RMSE) and mean absolute error (MAE). The data used for this study are daily solar radiation data in Uyo, spans January 1, 2018 to December 31, 2022 (5 years). According to the literature, the ARIMA (p,d,q) and ARFIMA (p,d,q) models are not being used to model solar radiation data applicable to PV system. From the results obtained, ARIMA (1,0,1) and ARFIMA (0,0.2045,1) models have a better goodness of fit for the series and the forecasting ability of the two adequate models show that ARIMA (p,d,q) model is preferred to ARFIMA (p,d,q) model. The ARIMA model forecast performance proves superior to the ARFIMA model for solar radiation. Hence, the ARIMA model may help to improve the forecast accuracy of PV system via solar radiation dataset.

**Keywords**: Photovoltaic (PV), Solar radiation, ARIMA, ARFIMA and Long memory

# I. INTRODUCTION/ BACKGROUND OF THE STUDY

Light is one of the most powerful tools in human life as we know, so it is very important to have sufficient knowledge about it. It comes in different forms like Lasers, bulbs and sun, in which all of them function in different ways to help human life and that of the plants.

Sunlight, which includes visible, ultraviolet, and infrared light, is the source of solar radiation. Everything in nature releases electromagnetic energy. While sunlight travels through Earth's atmosphere, it is partially absorbed, scattered, and reflected by molecules, aerosols, water vapor, and clouds. Both direct and diffuse solar radiation can reach the earth's surface; direct solar radiation is emitted directly into space, while diffuse solar radiation is reflected off from other objects or hindered by clouds. Global solar radiation is the sum of all of the sun's radiation that falls on a horizontal surface, including both direct and diffuse energy. Although it is less intense at these times, global solar radiation is seen in the twilight before sunrise and after sunset. Direct solar radiation is detected from sunrise to sunset. The worldwide solar energy that falls on a horizontal surface is measured using a pyranometer. Its sensor has a horizontal radiation-sensing surface that converts solar radiation energy into heat by absorbing it from the entire sky. This thermal energy can be measured to determine global solar radiation.

Instruments used to measure solar radiation include pyranometers, sensors, sunlight calculators, and various recording devices. It is possible to measure solar insolation on a daily, monthly, and annual average as well as on global isoflux contours, satellite cloud cover data, and solar radiation calculations numerically. According to Kazem et al. (2016) and Mazin et al. (2015), solar energy data tells us how much the sun's potential is at a certain point on Earth over a given length of time. For the purpose of planning and sizing solar energy systems, these statistics are crucial. Every geographic area should have timely, accurate, and relevant data available for design, optimization, and performance evaluation of solar technology. The use of solar energy is facilitated by thorough understanding and in-depth research of the site's potential; for this reason, ground level measurement is a crucial component of solar energy conversion systems.

Thanks to advances in technology, capturing solar radiation and turning it into electricity is today a straightforward process. Solar energy content can be used directly or transformed via solar technology into a variety of useful energy forms for daily life, including fuels, heat, light, and electricity. Solar energy application is required in a number of industries, including industry, agriculture, engineering, architecture, meteorology, and photovoltaic systems for the generation of electricity. To achieve this, we will concentrate on applying it to the solar system.

Direct solar energy conversion to electricity is accomplished with a Photovoltaic (PV) system. They function whenever the sun is present, but they generate more electricity when the sun is more focused and directly reaches the photovoltaic modules (i.e., the rays of sunlight are perpendicular to the PV modules). PV has the ability to directly power electrical appliances or store energy in a battery. In the absence of utility lines, they are less expensive and require less maintenance than wind turbines, diesel or fuel generators, or batteries by themselves.

Nigeria is a country that lies almost perfectly on the equation, which implies that she gets optimum intense temperature on a daily basis, hence solar radiation. This renewable energy is likely to be a new normal for energy in Nigeria in the coming years. With the spread and awareness of renewable energy in the country, businesses and industries as well as government parastatals and schools will perform better in their daily activities. Recent studies estimated the daily total solar thermal power potential in Nigeria at about 427,000MW, while the present hydro generated electricity the country primarily rely on is about 5000MW.[1] Thus, the needs to prioritise solar energy is necessary.

This present paper considers Uyo, the capital of Akwa Ibom State in Nigeria, as a case study. We attempt to model daily solar radiation on photovoltaic system in Uyo metropolis, Nigeria. We apply both ARFIMA and ARIMA models. The ARFIMA model relies on using a fractional difference operator for stationarity rather than the restrictive unit differencing

[1] https://www.iied.org/sites/default/files/pdfs/migrate/G03512.pdf

assumed in ARIMA models. Thus, ARFIMA model is expected to judge better the data dynamics than the other model.

The rest of the paper is structured as follows: section II showcase the literature review; the data and methods used for the empirical analysis are discussed in section III; the results and the discussion of this paper are presented in section IV and section V gives the conclusion for the study.

## II. LITERATURE REVIEW

Jung et al. (2022) studied Regional photovoltaic power forecasting using Vector Autoregression (VAR) model in South Korea. The work showed that VAR model has a better predictive ability than the Autoregressive Integrated Moving Average (ARIMA) model. Machine Learning techniques were used to examine solar PV systems. Dingari et al. (2020) give analysis of passengers travelling by Air India domestic flights using ARFIMA model and compared with ARIMA model, and the ARFIMA model was proved superior than the ARIMA model.



A combination of Wavelet Transform (WT) and Artificial Intelligence (AI) techniques were used to present one-hour-ahead power output forecasting of a PV system using solar radiation and temperature data. In the study, the application of WT has a high significant impact on ill-behaved PV power time-series data, and AI techniques capture the nonlinear PV instability in a better way (PARAS, 2012). Mohamed (2016) used different methods of fractional parameters estimation to show the efficiency of ARFIMA model with both semi-parametric and parametric approaches.

Titus et al. (2021) studied Guinea fowl production in Kenya using the ARIMA and ARFIMA models for the period of 2010 to 2019. The results obtained indicate that ARIMA and ARFIMA models gave a better fit to the data and were used to forecast evaluated via the Root Means Squared Error (RMSE) in which the ARFIMA model was found to give a better forecast of the Guinea Fowl weights compared to the ARIMA model.

Trevisan and David (2016) investigate Time Series of Soybean and Corn, which are two important Brazilian commodities. The aim of the work was to analyze soybeans and corn time series to compose the spot prince and forecast future prices for the aforementioned commodities. The ARIMA and ARFIMA models were compared, in order to test the better model for prices prediction. The results indicate the ARFIMA model to have higher efficiency than the ARIMA model. Seshadri et al., (2022) gives a favorable result on their applications and have an average accuracy of 86.02% depicted minimal deviation in the study. Thaker and Höller (2022) used the GEKKO optimization tool and assigned an algorithm for each cluster. They also used several other linear regressions time series and machine learning (ML) models to apply on solar energy based on irradiance classification and compare them. The results that the proposed GEKKO optimized model outperforms other machine learning and ensemble models.



## III. DATA AND METHOD

In this part, we enlightened the background of the data set used for the study and the methodologies used in examining solar radiation series.

### A. Data for the study

This study used data from daily solar radiation in Uyo metropolis and it spans from January 2018 to December 2022. The data were obtained from meteorological center, university of Uyo branch Akwa Ibom State.

### B. Methods

In this paper, we applied autoregressive integrated moving average (ARIMA) approach proposed by Box and Jenkins (1976) and a long memory model introduced by Grange and Joyeux (1980) known as autoregressive fractional integrated moving average (ARFIMA) model.

- **Autoregressive Integrated Moving Average (ARIMA) Model**

This model has proved to be used in representing both stationary and non-stationary time series. If $d$ is a non-negative integer, then $x_t$ is an $ARIMA\ (p,q)$ process if $y_t = \Phi(B)(1-B)^d X_t$ is a causal $ARMA\ (p,q)$ process. This means that $x_t$ satisfies a difference $\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t$ in equation form is given as

$$\Phi^*(B)X_t \equiv \Phi(B)(1-B)^d X_t = \Theta(\mathrm{B})\varepsilon_t, \quad \varepsilon_t \sim WN(0,\sigma^2) \tag{1}$$

Where $\Phi(B)$ and $\Theta(B)$ are polynomials of degrees $p$ and $q$, respectively, and $\Phi(B) \neq 0$ for $|B| \geq 1$. The polynomial $\Phi(B)$ has a zero of $d$ at $B = 1$. The process $X_t$ is stationary if and only if $d = 0$ in which case it reduces to $ARMA\ (p,q)$ process. $ARIMA$ Models are use in representing data with trend. It should be noted, however, that $ARIMA$ processes can also be appropriate for modeling series with no trend.

- **Autoregressive Fractional Integrated Moving Average (ARFIMA) Model**

The ARFIMA model is use when the dataset exhibits long memory characteristics. The generalized forms of ARFIMA (p,d,q) process is given as

$$\Phi(B)(1-B)^d X_t = \Theta(B)\varepsilon_t, \tag{2}$$

Where, $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ are the polynomial in $B$ and $\phi_i (i = 1,2,\cdots,p)$ and $\theta_j (j = 1,2,\cdots,q)$ are the parameters in the model and $\varepsilon_t$ is the error term.

Let a time series process denoted as $y_t$, then $ARFIMA$ process can be expressed in a simplest form as

$$y_t = (1-B)^d \varepsilon_t \tag{3}$$

Where $p = q = 0$ and $\varepsilon_t \sim N\ (0, \sigma_\varepsilon^2)$. The series is said to be stationary and invertible, when the order of the integration $d$ is distinct between $-0.5 < d < 0.5$,. For $d > 0.5$ indicate that the series is non-stationary in the mean and the variance is unbounded (Pfaff, 2006). At the range of $0 < d < 0.5$ the process has a long memory and for the range $-0.5 < d < 0$, the sum of absolution values of its autocorrelations tends to a constant.

In Maclaurins series expansion and the binomial theorem around $B = 0$ in (3), then, Mclead and Hipel 1978 defined long memory process as:

$$\lim_{x\to\infty}\left(\sum_{k=-N}^{N}|\rho_k|\right) \tag{4}$$

Where $\rho_k$ is the autocorrelation at lag $k$. The autovariance function $(\gamma_k)$ is given as

$$\gamma_k = E(X_t, X_{t-k}) = \frac{\Gamma(1-2d)\Gamma(d+k)}{\Gamma(\mathrm{d})\Gamma(1-\mathrm{d})\Gamma(1-\mathrm{d}+\mathrm{k})}\gamma_u^2 \tag{5}$$

$U_t \simeq (0, \gamma_u^2)$, the autocovariance must have the same sign as $d$, $k \geq 1$ and satisfies

$$\lim_{k\to\infty}[\gamma_k(k)/K^{2d-1}] = \frac{\Gamma(1-2d)}{\Gamma(\mathrm{d})\Gamma(1-\mathrm{d})}\gamma_u^2 \tag{6}$$

The dependence between observations is noted in the hyperbolic decay which is slower than the geometric decay stationary. A stationary and invertible Autoregressive Moving Average $ARMA$ process has auto correlation which is geometrical bounded i.e. for large, $|\rho_k \leq cm^k|$, where $0 < m < 1$ and $c$ is constant so that the limit in (6) is finite and this results in a short memory process.

## IV. Results and Discussion

This section shows the details of result analysis carried out on solar radiation and a brief explanation of the results. All the analysis in this work was done via R package and Eview.

Table 1: Descriptive Statistics

| Mean | Median | Maximum | Minimum | Std. Dev. | Skewness | Kurtosis | Jarque-Bera |
|---|---|---|---|---|---|---|---|
| 11.286 | 10.800 | 711.000 | 1.000 | 16.788 | 39.772 | 1674.870 | 2.08e+08 |

GSJ: Volume 12, Issue 9, September 2024
ISSN 2320-9186


The table 1 above shows the descriptive statistics of daily solar radiation. The data used consists of 1826 observations from the period 01-01-2018 to 31-12-2022, the interval of 5 years. This series follows normal distribution as indicated by the low value of the jarque-Bera (1987) test.

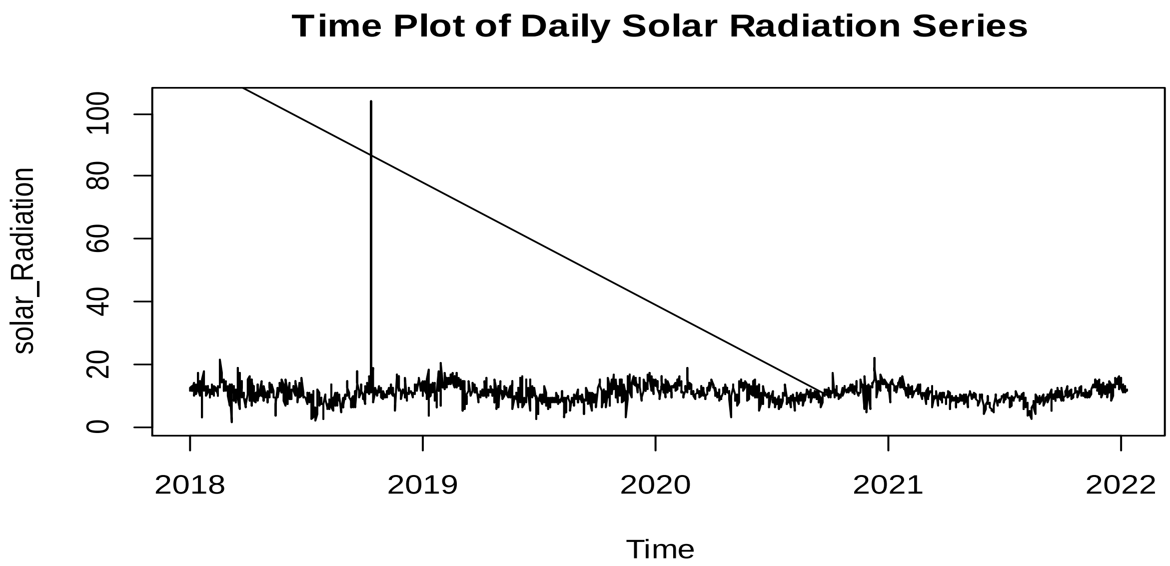


**Figure 1: Solar Radiation plot of the Original Series**

The graphical representation of daily solar radiation dataset in figure 1 revealed the important characteristic of the series. From the plot we can see some level of stationarity in the series but the strong outlier between 2018 and 2019 gives a hint for further investigation of other dynamics properties and this will be done through the help correlogram plot.

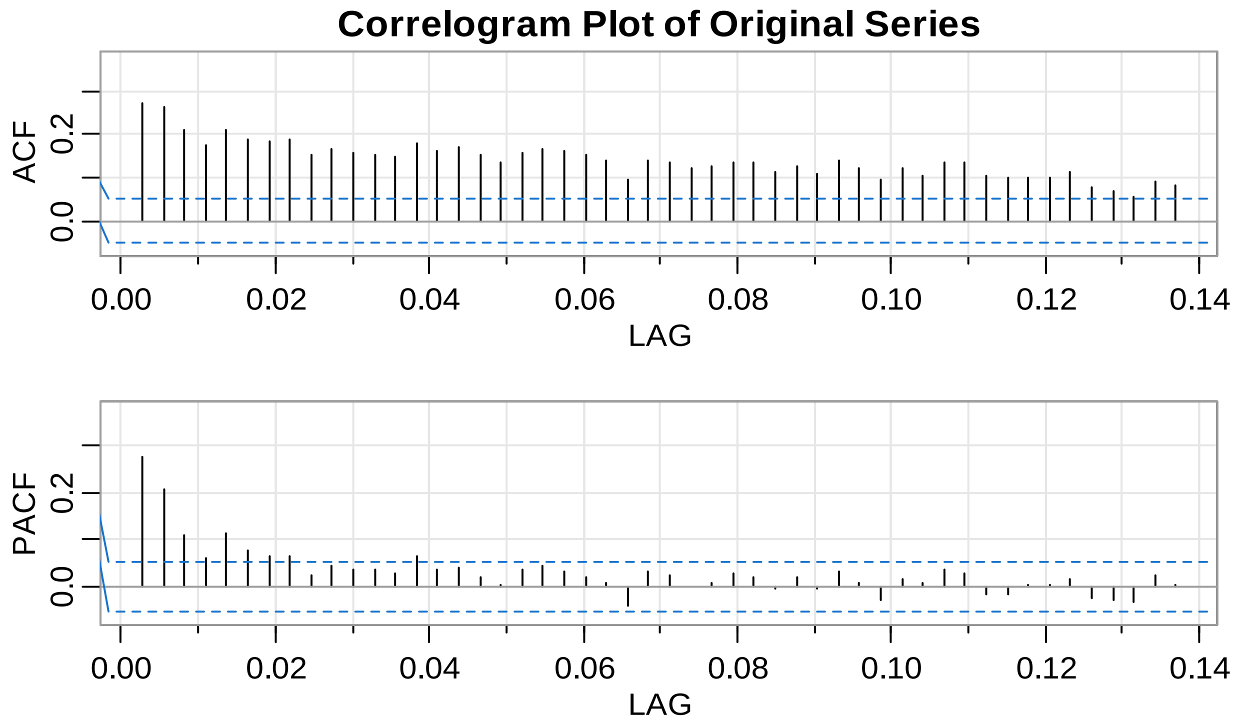


**Figure 2: The Correlogram plot of Daily Solar Radiation**



Autocorrelation function (ACF) and Partial Autocorrelation (PACF) plots are one of the ways that can help in enlightening whether differencing is needed or not in the series. By inspection, the slow decay in the ACF plot above is an indication of long range dependence and the plot also shows non stationary in the dataset meaning that differencing may be needed. Lag one and two in PACF plot are highly significant. Before we go further to difference the series, we need to carry out unit root test.

**Unit Root Test**

This paper used Augmented Dickey Fuller (ADF) test and KPSS test, by Kwiatkowski, Phillips, Schmidt and Shin (1992) designed to test the null hypothesis $I(0)$ against its alternative $I(1)$ test to test the level of stationarity for ARIMA amd ARFIMA models respectively, and the results are given below. The result obtained from table 2 below revealed that daily solar is stationary using ADF test for ARIMA model via probabilit decision rules. In the same table KPSS test for ARFIMA model shows that dataset is stationary in that the test statistic value is less than the critical values at 1%, 5% and 10% respectively; hence, the investigation for the order of differentiation.

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

| **Test** | **ADF** | | | | **KPSS** | |
|---|---|---|---|---|---|---|
| | Intercept | | Intercept and Trend | | Intercept | Intercept and Trend |
| Test Statistic | -41.88140 | | -41.88323 | | 0.118840 | 0.069017 |
| Test Critical values: 1% | -3.433732 | (0.0000*) | -3.963079 | (0.0000*) | 0.739000 | 0.216000 |
| 5% | -2.862920 | | -3.412273 | | 0.463000 | 0.146000 |
| 10% | -2.567552 | | -3.128068 | | 0.347000 | 0.119000 |

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

| Variable | Values |
|---|---|
| $\hat{d}$ | 0.3309008 |
| sd.as | 0.1212829 |
| sd.reg | 0.1044626 |



We used semi-parametric estimation of Geweke and Porter-Hudak (1983) algorithm (GPH) to examine the value of $d$ for ARFIMA model. This test permits researchers to test the null hypothesis of a unit root ($d = 1$) against the alternative of fraction integration ($d < 1$) on the series. The end result is displays on table 3, where the fractional parameter ($d$) fall within the range of stationary long memory process.

**Estimation of Parameters**

The maximum likelihood and semi-parametric methods are used for parameters estimation in this work for ARIMA and ARFIMA models respectively.

**ARIMA Model**

In this part, a classical time series model was employed; Autoregressive Integrated Moving Average (ARIMA). The order of p and q were identified by combination of different values and the best model is picking based on the minimum value of Akaike's information criteria.

**Table 4: ARIMA Models Estimation**

| Models | AIC | Log Likelihood |
|---|---|---|
| ARIMA(1,0,1) | 7736.3 | -3864.15* |
| ARIMA(0,0,1) | 7903.36 | -3948.62 |
| ARIMA(1,0,0) | 7869.31 | -3931.65 |

Table 4 consists of estimated models ARIMA model, where we discovered that ARIMA (1,0,1) model has the minimum value of AIC =7736.3. The parameters estimate is in table 5 below.

**Table 5: Estimated Parameters of $\boldsymbol{ARIMA(1,0,1)}$**

| Parameter | ar1 | ma1 |
|---|---|---|
| Estimate | 0.9705 | -0.8669 |
| s.e. | 0.0182 | 0.1086 |

sigma^2 estimated as 11.16: log likelihood = -3864.15, AIC = 7736.3

The fitted model of $ARIMA(1,0,1)$ is

$$(1+0.9705B)x_{\mathrm{t}} = (1-0.8669B)w_{\mathrm{t}}$$

**ARFIMA Model**

In this section, fractional difference ARMA model was investigated on Uyo daily solar radiation data. The first step applied to estimate long memory parameter was semi-parametric method of Geweke and Porter-Hudak (1983) approach, due to the limitation of GPH method, we also applied a truncated maximum likelihood procedure discussed by Haslett and Raftery (1989) method to complete the process. The adequate model is select from the minimum values of AIC. The models estimation are show in table 6 and table 7 revealed parameters of the adequate model.

**Table 6: ARFIMA Models Estimation**

| Models | AIC | Log Likelihood |
|---|---|---|
| ARFIMA(1,0.2045,1) | 7747.34 | -3869.67 |
| ARFIMA(0,0.2045,1) | 7747.02 | -3870.36 |
| ARFIMA(1,0.2045,0) | 7746.73 | -3870.51 |

**Table 7: Estimated Parameters of $\boldsymbol{ARFIMA(0,0,1)}$**

| Parameter | D | ma1 |
|---|---|---|
| Estimate | 0.2045 | 0.0436 |
| s.e. | 0.0244 | 0.0386 |

sigma[eps] = 3.354903, log likelihood = -3870.51, AIC = 7746.73

The fitted model of $ARFIMA(0,0,1)$ is

$$(1-B)^{0.2045}\,\mathrm{y}_{\mathrm{t}} = 0.0436(1-B)^{0.2045} + \hat{\varepsilon}_{\mathrm{t}}$$

## Forecast Performance

This study used the mean square error (MSE), the root mean square error (RMSE) and the mean absolute error (MAE) (Armstrong and Collopy 1992 and Hyndman and Koehler 2006) to examine the forecasts accuracy of the two adequate models and make comparison between them.

**Table 8: Forecast Accuracy Measures**

| MODEL | MSE | RMSE | MAE |
|---|---|---|---|
| ARIMA(1,0,1) | 11.1559 | 3.3401 | 1.7684 |
| ARFIMA(1,0.2045,0) | 11.2739 | 3.3577 | 1.7771 |

Judging by the results of table 8 above, we can say that ARIMA model slightly outperformed ARFIMA model in forecasting.



## V. CONCULSION

Time series analysis is one of the ways used in checking how a variable change over time. More often, it is used to identify timely patterns and trends possess in a dataset. The energy of PV system is intermittent and stochastic in nature, as it depends on solar radiation and others weather characteristics to function. This paper aims to explore the application of the ARIMA and ARFIMA models within the context of photovoltaic (PV) system via solar radiation data to investigate the dynamics property of solar radiation. In today's world, the sufficient knowledge of how PV works has to be in our finger tips because without light numerous things life would not be carried out.

We used Akaike Information Criteria (AIC) as the model assessment tool. The two models explored in this works fit the data well; on examined ARIMA and ARFIMA models fitted on their minimum values of AIC which are 7736.3 and 7746.73. We discovered that

ARIMA (p,d,q) model has the lowest AIC value of 7736.3 making it to be the best fitted model for the series. The forecast performance accuracy of the two models was tested using MSE, RMSE, and MAE. Results obtained in Table 8 proved that the ARIMA (p,d,q) model outperformed the ARFIMA (p,d,q) model slightly.

Inclusion, this paper can be used to develop a good understanding of the overall process of PV and solar radiation. Therefore, the nature of PV system can be modelled and forecast effectively using both the ARIMA and ARFIMA models via solar radiation data.

## REFERENCES

Armstrong, J. and Collopy, F. (1992): Error Measures for generalizing about forecasting methods empirical comparisons. Int. J Forecast. 8(1):69-80.

Dingari, M., Doodipala, M. and Sumalatha, V. (2020): Time Series Analysis for long memory process of Air traffic using ARFIMA. International Journal of Scientific & Technology Research, Volume 8, ISSN 2277-8616.

Geweke, J.F. and Porter-Hadak, S.(1983): The Estimation and Application of Long memory Time Series models, Journal of Time Series Analysis, 4: 221-238.

Haslett, J. and Raftery, A. E. (1989): Space-time modelling with long momery dependence: Assessing Ireland's wind power resource (C/R:89V38p21-50) Applied Statistics, 38, 1-21.

Hyndman, R. and Koehler, A. (2006): Another look at measures of forecast accuracy. Int J Forecast. 22(4):679-88.

Jarque, C. M. and Bera, A. K. (1987): A Test for Normalit of Observations and Regression Residuals. International Statistical Review, 55, 163-172. http://dx.doi.org/10.2307/1403192

Jung, A.-H., Lee, D.-H., Kim, J.-Y., Kim, C.K., Kim, H.-G., Lee, Y.-S.(2022): Regional Photovoltaic Power Forecasting Using Vector Autoregression Model in South Korea. Energies 2022, 15, 7853. https://doi.org/10.3390/en15217853

Mohamed, R. A. (2016): Using ARFIMA Models in Forecasting the total Value of traded Securities on the Arab Republic of Egypt. www.arpapress.com/Volumes/Vo127Issue/IJRRAS_27_1_04.pdf

Paras Mandala , Surya Teja Swarroop Madhira , Ashraf Ul haque , Julian Meng , Ricardo L. Pineda (2012): Forecasting Power Output of Solar Photovoltaic System Using Wavelet Transform and Artificial Intelligence Techniques, Procedia Computer Science 12/ 332 – 337, doi: 10.1016/j.procs.2012.09.080

Samanta, M.; Bharath, K.S.; Jayesh, B.Y. Short-Term Power Forecasting of Solar PV Systems Using Machine Learning Techniques. *Environ. Sci. Comput. Sci.* **2014**, *2014*, 18566286. [**Google Scholar**]

Seshadri, P., TS, B. P., Kumar, A., Keerthana, H., Kavinmathi, G.,Senthilrani, S. (2022): time Series-Based Photovoltaic Power Forecasting to optimize Grid Stability. Electric Power Components and systems 49(1-17):1-10. DOI:10-1080/15325008.2022.2129871.

Thaker, J. and Höller, R. A. (2022): Comparative Study of Time Series Forecasting of solar Energy Based on Irradiance Classification. Energies 15, 2837. https://doi.org/10.3390/en/5082837.