Journal of Power and Energy Engineering, 2025, 13(9), 440-454
https://www.scirp.org/journal/jpee
ISSN Online: 2327-5901
ISSN Print: 2327-588X

Time-Series Forecasting of Solar Photovoltaic Power Generation in Malaysia

Irma Wani Jamaludin[1*], Junainah Sardi[1], Muhammad Hakimi Faqihi Abu Bakar[1], Norhafidzah Mohd Saad[2]

[1]Faculty of Electrical Technology and Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Melaka, Malaysia
[2]Faculty of Electrical & Electronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia
Email: *irma@utem.edu.my

How to cite this paper: Jamaludin, I.W., Sardi, J., Bakar, M.H.F.A. and Saad, N.M. (2025) Time-Series Forecasting of Solar Photovoltaic Power Generation in Malaysia. *Journal of Power and Energy Engineering*, **13**, 440-454.
https://doi.org/10.4236/jpee.2025.139027

Received: September 2, 2025
Accepted: September 23, 2025
Published: September 26, 2025

Abstract

The need for accurate forecasting techniques to support the integration of solar power into the national grid is highlighted by Malaysia's growing reliance on renewable energy. Using Global Horizontal Irradiance (GHI) data gathered in Melaka, this study proposes a time-series forecasting method for solar Photovoltaic (PV) power generation. Before analysis, a thorough dataset covering five years was pre-processed to guarantee accuracy, consistency, and dependability. In order to better predict PV output, the Auto-Regressive Integrated Moving Average (ARIMA) model was created and used to account for both short-term and long-term variations in solar irradiance. Comparisons between expected and observed values were used to evaluate the model's performance, and the results showed how well it captured seasonal variations, diurnal cycles, and stochastic fluctuations related to tropical weather. The outcomes validate the ARIMA model's applicability for Malaysian renewable energy forecasting, providing insightful information for improving grid stability, optimizing PV system performance, and guiding sustainable energy planning policy. In a broader sense, the results highlight how important sophisticated time-series models are for tackling the problems caused by weather variability in tropical areas, where precise solar energy forecasting is necessary for long-term energy security and effective resource management.

Keywords

Global Horizon Irradiance, Time-Series Forecasting, Auto-Regressive Integrated Moving Average, Renewable Energy

1. Introduction

Photovoltaic (PV) modules convert sunlight into electricity, and thermal collec-

tors harvest its heat. For solar energy to become a primary source, research and development must address efficiency, storage, and environmental factors including dust, temperature, and humidity [1].

Green energy sources are not yet an affordable alternative to traditional sources due to high production costs, although the gap is shrinking. Solar energy is favourable for electric power generation and could meet the planet's electrical energy needs [2]. Forecasting technologies reduce scheduled-to-actual dispatch fluctuations at load dispatch centers. Most of the power from solar PV panels comes from a specific pattern of sunlight. Irradiance is zero at night and rises at daybreak. Following a daily cycle, it peaks in the afternoon and progressively declines to near-zero around sunset. Weather factors can change irradiance while keeping the pattern.

For short-term (system reserves, dispatch) and long-term (unit commitment, scheduling) energy grid management, solar irradiance forecasting is becoming increasingly crucial. Most forecasting literature covers intraday or day-ahead time horizons. However, the variable solar resource increases the risk of grid instabilities in countries with large PV deployments [3].

Traditional energy sources like fossil fuels have been the main source of energy for many years, but they are being replaced by renewable resources to solve the energy crisis caused by their depletion and the environmental damage caused by greenhouse gas emissions. We need modern energy management systems that integrate energy in real time in distribution networks and power grids. Scientific solutions for integrating solar energy into a real-world grid include an accurate near-real-time forecasting tool for solar energy management and proportionate dispatching to and from a grid system [4]. Solar energy forecasting considers client usage for stability, energy regulation, reverse management, scheduling, and unit commitments.

Clouds and sky movement cause PV system output fluctuations. Energy storage and generation projections are the most common ways to reduce the influence of highly variable renewable energy production on electrical system stability [5]. PV power plant production relies on sun irradiation (Wh/m^2) on PV panels. Thus, global horizontal irradiance (GHI) is the most important input parameter in most PV power prediction systems, and researchers are forecasting GHI to use in PV power forecasts. Different models and input data are used depending on the forecast horizon.

Renewable energy sources (RESs) are resources that are environmentally beneficial and clean. They can generate power with less greenhouse gas or air pollution. RESs are growing rapidly due to technology and regulation, making renewable energy an important investment and competitive area for international cooperation. Meeting rising energy demand while lowering greenhouse gas emissions requires power demand-supply planning. Forecasting RESs for grid management and power generation optimization requires accurate models [6].

Meteorological data for a typical year and solar resource distribution knowledge

are needed to evaluate and optimize solar equipment. Early solar radiation studies included four peninsula monitors, estimations of fluctuations based on top-of-atmosphere radiation, and regression analysis of sunlight hours. Pollution affects solar radiation. Smoke and haze cut UV and global sun exposure 16% - 23% [7].

Solar panels generate electricity from sunlight. Because the sun offers almost un-limited energy, this is renewable. PV systems employ solar radiation to generate power that can be stored in batteries and used during sunny periods to power homes and businesses. PV systems can power loads. Solar energy predictions are made using manual calculations, analytical algorithms, and simplified assumptions based on historical data and meteorology. These techniques might be complicated and expensive because they require a lot of human experience and effort to analyze the data. Manual prediction methods may struggle to understand solar energy production's many link-ages, resulting in inaccurate forecasts [8].

Malaysia is strongly committed to creating a low-carbon future by shifting toward a more sustainable economy. The National Energy Transition Roadmap (NETR) [9] outlines a national target of achieving 70% renewable energy capacity in the power mix by 2050, with solar energy expected to contribute the largest share of this expansion. Through the Responsible Transition (RT) Pathway 2050, the country aims to move from heavy reliance on fossil fuels to greener energy sources, including solar power. Total Primary Energy Source (TPES) modelling predicts that Malaysia's energy demand will grow modestly, by 0.2% per year, increasing from 95 million tons of oil equivalent (Mtoe) in 2023 to 102 Mtoe in 2050. The RT Pathway shows significant progress, with coal being phased out and fossil fuel dependency dropping from 96% in 2023 to 77% by 2050.

Solar power, as part of the renewable energy mix, is expected to play a vital role in this transition, contributing to the rise of renewables from just 4% of TPES in 2023 to 23% by 2050. With its abundant sunlight, Malaysia has great potential to expand solar PV generation, which will be essential for achieving these sustainable energy goals.

2. Approaches to Solar Forecasting

The fundamental objective of developing a forecasting model is to generate predictions that approximate, as closely as possible, the actual observed values in a time series. This purpose aligns with the core principles of time-series modeling, which involves analyzing sequential data collected over time to uncover patterns, correlations, and trends that can inform future predictions. A time series is typically defined as a sequence of observations recorded at successive intervals, capturing the temporal evolution of a parameter of interest. In the context of solar forecasting, these observations are often derived from meteorological variables such as temperature, wind speed, humidity, and cloud cover that fluctuate due to naturally occurring random processes [10].

Statistical methods have traditionally been used to establish correlations between historical weather parameters and solar irradiance, particularly on an

hourly or daily basis. For instance, autoregressive integrated moving average (ARIMA) models, exponential smoothing, and regression-based approaches are frequently employed to reconstruct relationships between past and present meteorological conditions and their influence on solar radiation. A notable strength of these statistical models is that they do not require a detailed understanding of the internal physical mechanisms of the solar system under study. Instead, they rely on historical data patterns, making them relatively simple to implement and computationally efficient [11]. However, their performance may be limited when dealing with highly nonlinear or chaotic weather dynamics, which often characterize solar irradiance patterns.

Machine learning (ML) techniques have become strong substitutes or supplements to conventional statistical methods in solar forecasting in recent years. The creation of algorithms that can automatically increase their predictive accuracy with additional data exposure is the focus of machine learning. ML approaches are well-suited to capturing intricate, multifaceted, and nonlinear relationships between input features and target outputs, in contrast to traditional statistical models. Large datasets with meteorological and irradiance measurements are commonly used to train ML models in solar power forecasting. This allows the models to detect complex patterns that are hard to capture with purely statistical methods [12].

Artificial neural networks (ANNs), support vector machines (SVMs), random forests, and, more recently, deep learning architectures like convolutional neural networks (CNNs) and long short-term memory (LSTM) networks are a few frequently used machine learning techniques. By acquiring strong representations of temporal and spatial dependencies, these models have shown great promise in tackling the difficulties associated with solar forecasting. For instance, ANNs and LSTMs are especially useful for short-term forecasting because they can model nonlinear time-dependent relationships. However, the quality and accessibility of training data have a significant impact on the performance of machine learning models. These models can generalize well on large and varied datasets, but their predictive accuracy may be compromised by sparse or noisy data.

Overall, both statistical and machine learning approaches have unique advantages in solar forecasting. Statistical models are valued for their simplicity, interpretability, and efficiency in scenarios with limited data. In contrast, ML techniques excel in capturing complex dynamics and nonlinear relationships but often demand larger datasets and greater computational resources.

3. Methodology

The methodology adopted in this study is designed to ensure the development of accurate and reliable time-series forecasting models for solar irradiance. It is organized into two main phases. First, data preprocessing is conducted to clean, transform, and prepare the raw meteorological dataset, ensuring its quality, consistency, and suitability for analysis. Second, the ARIMA modeling framework is

employed to characterize the temporal dynamics of the data, capturing both short-term dependencies and potential seasonal patterns. This structured approach provides a rigorous foundation for evaluating the forecasting performance of the proposed models.

3.1. Data Preprocessing

The study begins with the collection of historical solar irradiance data from website source Ref. [13], specifically focusing on the all-sky surface direct normal irradiance, Wh/m^2 which is essential for forecasting. Once the data is obtained, it undergoes preprocessing to handle missing values and ensure it is properly structured as a time series dataset. The data clustered to hourly and monthly. The flowchart of the process is shown in Figure 1.

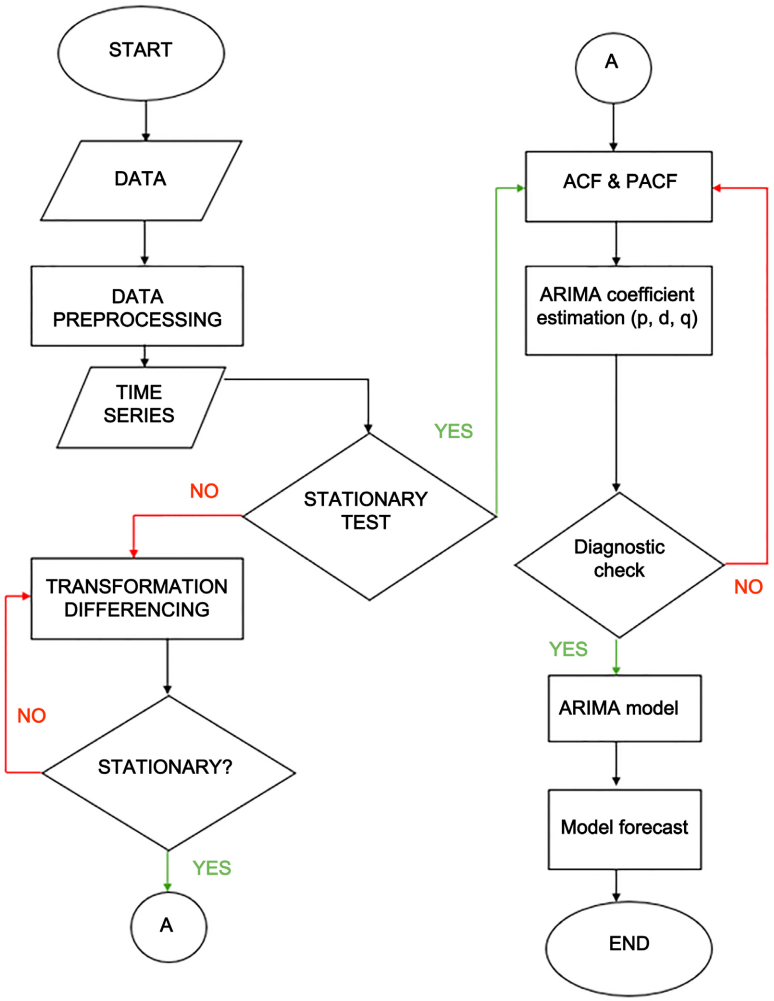

Figure 1. Flowchart of preprocessing data and ARIMA method.

3.2. Auto-Regressive Integrated Moving Average (ARIMA) Method

After preprocessing, a stationarity test is conducted to assess whether the data is stable over time. If the data is non-stationary, transformations and differencing

are applied until stationarity is achieved. This step is crucial to prepare the data for accurate time series modelling.

Following that, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) tests are used to identify the appropriate parameters (*p*, *d*, *q*) for the ARIMA model. For instance, an output of AR (3), I (2), and MA (1) would result in the ARIMA (3, 2, 1) model. A diagnostic check is then performed to validate the model by examining the p-values and selecting the model with the lowest mean squared error (MSE). Once the best model is selected, forecasting is performed, and the predicted values are compared with the actual values to evaluate the model's accuracy and reliability.

The Autoregressive Integrated Moving Average (ARIMA) model is a widely used statistical method for analyzing and forecasting time series data. It follows the Box-Jenkins approach and combines two main components: Autoregressive (AR) and Moving Average (MA). The AR part assumes that the current value of a time series depends on its previous values, while the MA part models the relationship between the current value and past forecast errors. An AR model of order *p*, written as AR (*p*), represents the current value as a linear combination of the past p values and a random error term as in Equation (1). [5]

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t \tag{1}$$

In this model, ε_t represents the forecast error, while $\phi = \left(\phi_1, \phi_2, \phi_3, \cdots, \phi_p\right)$ denotes the set of coefficients for the AR component. The value of p indicates the number of lagged observations used, known as the order of the AR process. For the Moving Average (MA) part, the current value of the time series is influenced by past forecast errors. An MA model of order *q*, written as MA(*q*), uses *q* past error terms to improve forecast accuracy. This approach helps to smooth out random noise and better capture the underlying pattern in the data as in Equation (2).

$$X_t = \varepsilon_t - \sum_{j=1}^{p} \theta_j \varepsilon_{t-j} \tag{2}$$

The ARIMA model combines three key components: Autoregressive (AR), Integration (I), and Moving Average (MA). It is written as ARIMA (*p*, *d*, *q*), where p is the number of lagged observations used in the AR part, d is the number of times the data must be different to become stationary, and q is the number of lagged forecast errors in the MA part. Together, these elements help model and forecast time series data effectively, especially when trends or patterns are present in the data. In this model, t represents the time index, and B is the backshift operator, which means that applying B to a value shifts it one time step back (*i.e.*, $Bx_t = x_{t-1}$). The functions $\varphi(B)$ and $\theta(B)$ are used to represent the Autoregressive (AR) and Moving Average (MA) parts of the model. These functions contain the coefficients that define how past values and past forecast errors are used to make current predictions as in Equations (3)-(5).

$$\left(1-B\right)^d X_t = \frac{\theta\left(B\right)}{\phi\left(B\right)} \varepsilon_t \tag{3}$$

$$\phi(B) = 1 - \phi_1 B^1 - \phi_2 B^2 - \cdots - \phi_P B^P \quad (4)$$

$$\theta(B) = 1 - \theta_1 B^1 - \theta_2 - \cdots - \theta_q B^q \quad (5)$$

The equations presented in Equation (1) until Equation (5) provide a method for manually determining the ARIMA model order. These formulas help identify the optimal autoregressive (AR), integrated (I), and moving average (MA) components needed for accurate modeling. In practice, autocorrelation function (ACF) and partial autocorrelation function (PACF) plots are typically used to guide this selection. However, statistical software offers a more efficient alternative by automatically estimating model parameters and comparing candidate models using information criteria such as the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) [14]. Once the model parameters are established, they are implemented in the forecasting process to yield a robust ARIMA model that optimizes both accuracy and simplicity. The ARIMA model was selected for this study due to the relatively long and consistent time-series dataset (five years of hourly GHI observations). Given that the data became stationary after differencing, ARIMA offered a good balance between interpretability and predictive power. Unlike more complex machine learning methods, ARIMA allows transparent parameter tuning and statistical validation, making it suitable for initial deployment in energy planning where clarity of assumptions and reproducibility are critical.

4. Results

This section presents the results and findings of the project based on the methods described earlier. It focuses on analyzing the forecasting performance of the ARIMA model by comparing the predicted values with actual data.

4.1. Hourly GHI Variation

The hourly Global Horizontal Irradiance (GHI) in Melaka, Malaysia, was analyzed over a 7-day period to capture the short-term diurnal pattern, as shown in **Figure 2**. The results show a consistent rise in irradiance after sunrise, a peak around midday, and a subsequent decline toward sunset, which is characteristic of daily solar radiation profiles in tropical climates. The magnitude of peak irradiance varied slightly across the seven days, with maximum values ranging from approximately 700 Wh/m^2 on cloudier days to over 900 Wh/m^2 on clearer days. These variations highlight the influence of short-term atmospheric conditions such as cloud cover, turbidity, and rainfall.

Although the overall diurnal profile remained stable, the variability in peak values underscores the need for system designers to consider short-term intermittency in PV power generation. Energy storage and hybridization strategies are essential to mitigate these fluctuations, ensuring more reliable energy delivery to the grid.

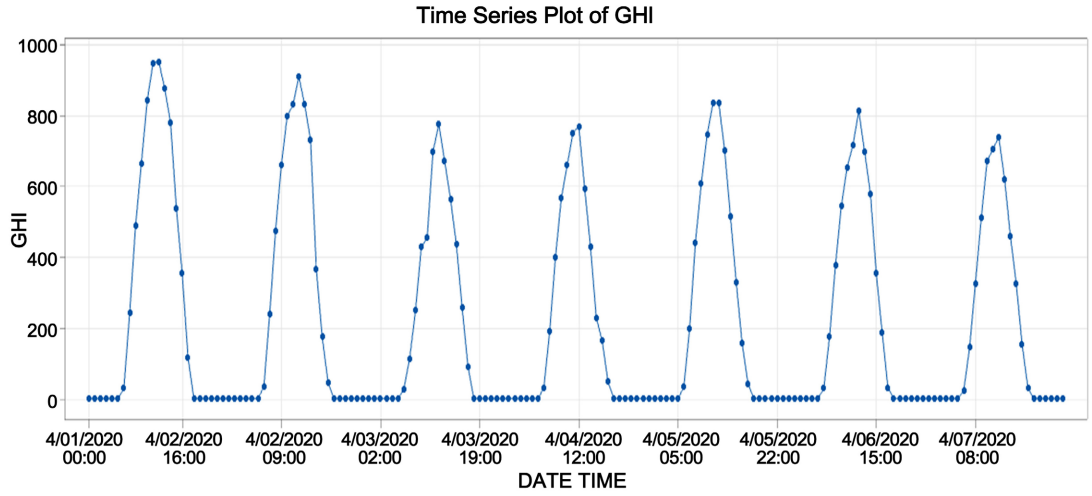

Figure 2. Time series plot of hourly for 7 days.

4.2. Monthly and Seasonal Trend

Long-term monthly GHI data from January 2020 to July 2024 as shown in **Figure 3** were subject to time series decomposition into trend, seasonal, and residual components (**Figure 4**). The trend component remained relatively flat with only minor oscillations, suggesting no significant structural changes in solar irradiance, thereby indicating reliability for long-term planning.

The seasonal component displayed a clear annual cycle, with irradiance peaks in early and mid-year and troughs in mid- to late-year which consistent with Malaysia's monsoon influenced tropical climate. The seasonal amplitude of approximately ±12,000 units quantifies this strong cyclical variation.

The residuals show that there are still random weather-driven outliers even though seasonality rules. These outliers are likely caused by localised weather events like thunderstorms or clouds.

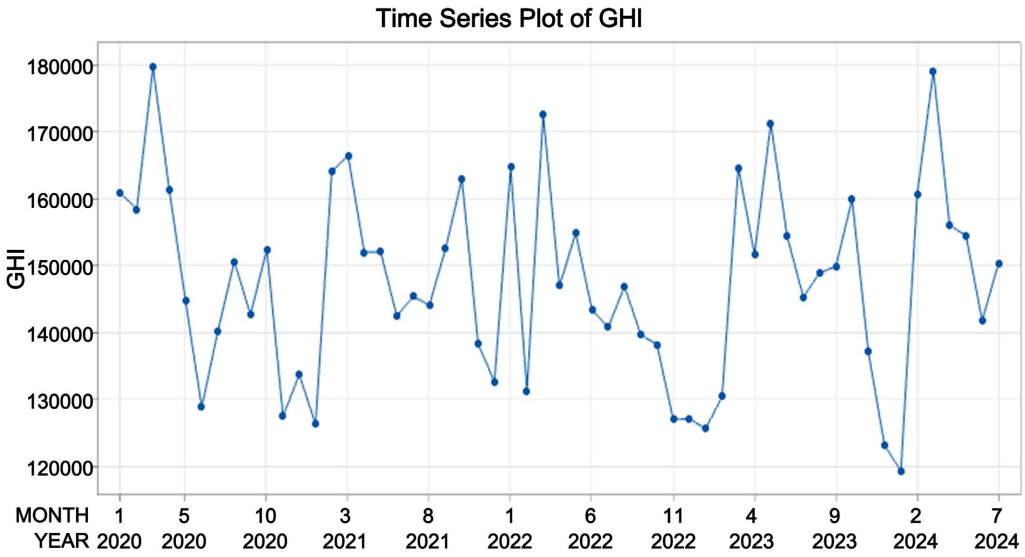

Figure 3. Time series plot of monthly plot of GHI from January 2020 until July 2024.

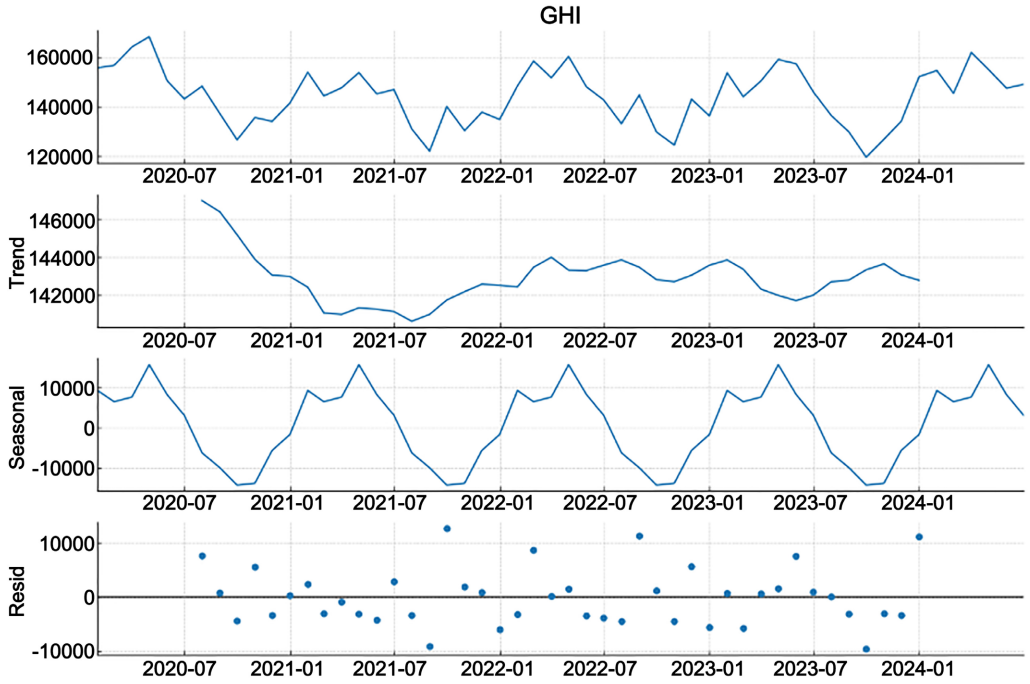

Figure 4. Time series decomposition into trend, seasonal, and residual components.

4.3. Autocorrelation and Partial Autocorrelation Analysis

The autocorrelation (ACF) and partial autocorrelation (PACF) analyses were performed to evaluate the temporal dependence structure of the GHI dataset. Figure 5 shows the ACF and PACF plots of the differenced GHI data. The ACF plot reveals a strong positive autocorrelation at lag 1, followed by values that quickly decline towards zero. Some spikes beyond lag 20 suggest possible seasonal effects. This pattern indicates the presence of a moving average (MA) component in the data.

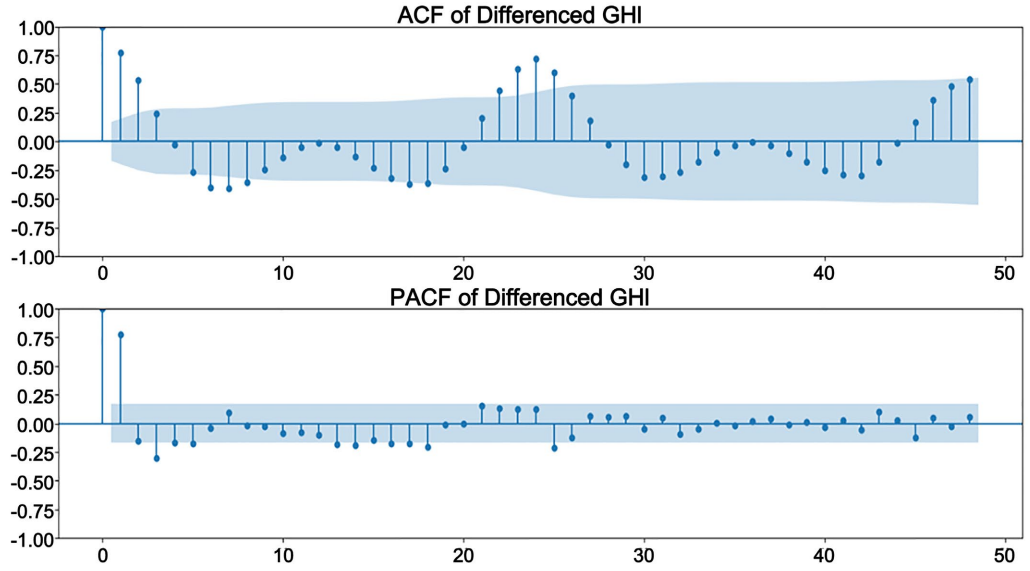

Figure 5. ACF and PACF graph for hourly GHI data.

The PACF plot shows a significant spike at lag 1 and 2, with the rest of the spikes staying within the confidence interval. This suggests an autoregressive (AR) component of order 2, meaning the current value is mostly influenced by the value from one time step before.

Figure 6 shows the ACF and PACF plots generated from historical data between 2020 and 2024. These plots help determine the ARIMA model orders (p, d, q) for the non-seasonal part and (P, D, Q) for the seasonal part. In the ACF plot, the first and second lags exceed the significance limits, suggesting that the moving average order (q) is 2. Similarly, the PACF plot shows significant spikes at the first and second lags, indicating that the autoregressive order (p) is also 2.

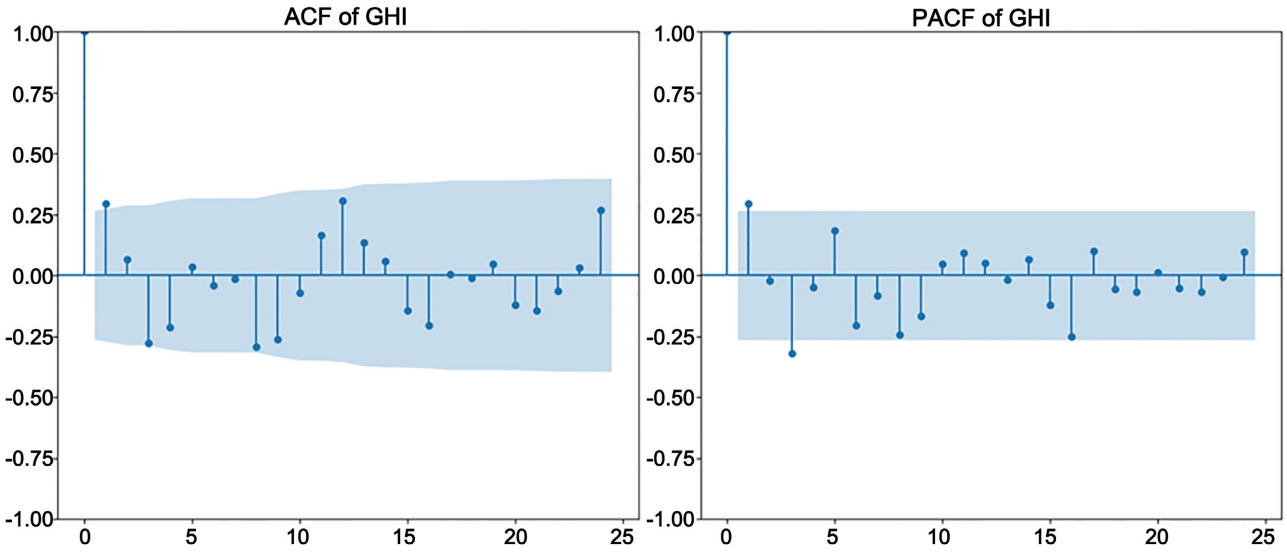

Figure 6. ACF and PACF graph for monthly GHI data.

4.3.1. ARIMA Modelling Hourly

Best Model: ARIMA (2, 0, 0) (2, 1, 0)

The optimal forecasting model identified in this study is the ARIMA (2, 0, 0) (2, 1, 0). The non-seasonal component, ARIMA (2, 0, 0), consists of two autoregressive (AR) terms, which indicates that the current value of the time series is predicted using its two most recent observations. Since the data is already stationary in the short term, no non-seasonal differencing is required ($d = 0$). Additionally, there are no moving average (MA) terms ($q = 0$), suggesting that the model does not rely on past forecast errors for predictions. This structure allows the non-seasonal part to effectively capture short-term dependencies and immediate fluctuations in the dataset.

The seasonal component, represented as (2, 1, 0), accounts for recurring daily patterns in the hourly solar irradiance data. Here, the two seasonal autoregressive (SAR) terms use information from observations recorded at the same hour on the previous day and two days earlier (lags of 24 and 48, respectively). Seasonal differencing of order one ($D = 1$) is applied to eliminate systematic daily trends, thereby stabilizing the seasonal cycle. Similar to the non-seasonal part, no seasonal

moving average ($Q = 0$) terms are included, meaning that past seasonal forecast errors are not incorporated into the prediction process. This seasonal specification enables the model to effectively capture and replicate the daily variations of GHI, thereby enhancing the accuracy of forecasts across different time horizons.

Figure 7 shows the time series plot displays historical GHI data alongside forecasted values generated by the SARIMA model. The solid blue line represents actual GHI values, revealing patterns such as trends and seasonality. The red line shows the predicted values, while the shaded area around it indicates the 95% confidence interval. This interval reflects the level of uncertainty in the forecasts. It is narrower for short-term predictions and becomes wider as we move further into the future. This visualization helps to clearly assess how well the model fits past data and how reliable the future forecasts.

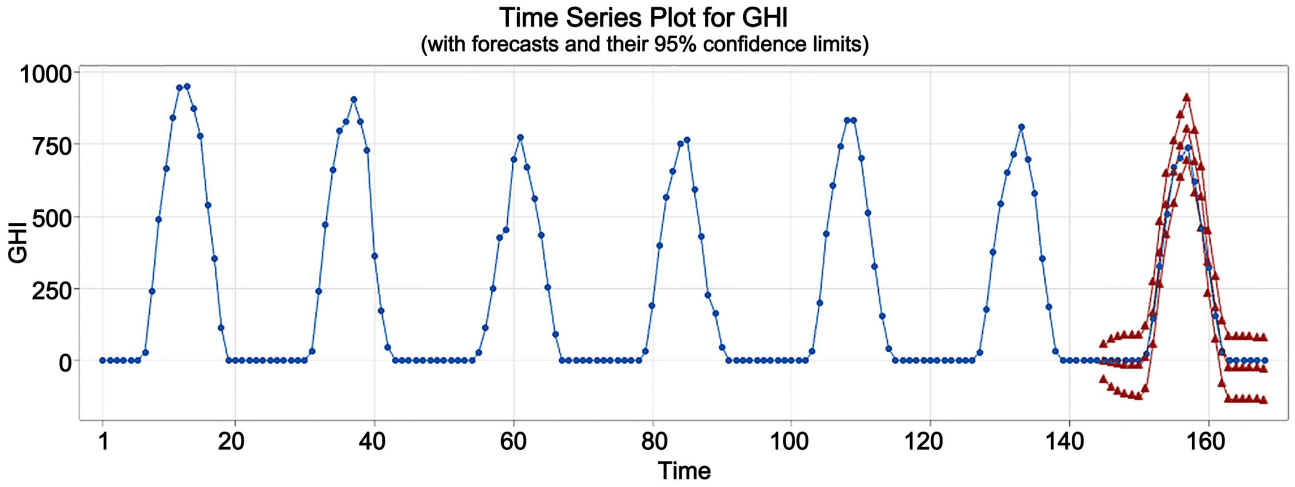

Figure 7. Time series plot with forecast and confidences interval.

4.3.2. ARIMA Modelling Monthly

Best model: ARIMA (2, 0, 2) (0, 0, 0) [12]

For the monthly GHI data, the best-fitting model was identified as ARIMA (2, 0, 2) (0, 0, 0) [12]. Unlike the daily model, which required seasonal adjustments, the monthly series exhibited little evidence of strong recurring seasonal patterns, leading to the selection of a non-seasonal ARIMA structure with added moving average components.

The non-seasonal component, ARIMA (2, 0, 2), captures the short-term dynamics of monthly GHI variations. With two autoregressive (AR) terms ($p = 2$), the model incorporates information from the previous two months to estimate the current month's irradiance. Since the data was already stationary, differencing was unnecessary ($d = 0$). The inclusion of two moving average (MA) terms ($q = 2$) allows the model to account for the influence of past forecast errors over the last two months. This ARMA-type structure provides a flexible balance, enabling the model to capture both persistent trends and noise in the monthly GHI series.

The seasonal component, (0, 0, 0) [12], indicates that although the model was tested with a 12-month seasonal cycle, no seasonal autoregressive (P), seasonal

differencing (D), or seasonal moving average (Q) terms were required. This suggests that the monthly data did not exhibit consistent or strong seasonal patterns that could significantly improve prediction accuracy. In other words, while solar irradiance naturally has annual variations, the data characteristics such as irregular fluctuations caused by weather variability were better explained by the non-seasonal ARMA structure. Adding seasonal parameters did not enhance model performance and risked overfitting.

The strength of the ARIMA (2, 0, 2) (0, 0, 0) [12] model lies in its ability to capture short-term dependencies and error corrections without introducing unnecessary seasonal complexity. Comparative testing against alternative models confirmed that the inclusion of both AR(2) and MA(2) terms provided a superior fit, as reflected in lower error measures such as RMSE and AIC. Models with higher-order terms or added seasonal parameters did not yield significant improvements and often increased model complexity.

Table 1 shows the forecast results from Time Period 49 to 54 using a time series model. For each period, it shows the predicted value (Forecast), the standard error (SE Forecast), and the 95% confidence interval (Lower and Upper limits), along with the actual observed values. The forecasts range between approximately 140,000 to 170,000, with a consistent standard error of 11578.6, indicating the level of uncertainty in the predictions.

Table 1. Forecast data value with confidence interval and actual value.

95% confidence interval from time period 49					
Time period	Forecast	SE forecast	Lower	Upper	Actual
49	140,328	11578.6	117,630	163,027	119,420
50	140,766	11578.6	118,068	163,465	160,604
51	170,209	11578.6	147,510	192,907	179,147
52	153,442	11578.6	130,744	176,141	156,070
53	162,799	11578.6	140,101	185,498	154,605
54	148,207	11578.6	125,508	170,906	141,843

The 95% confidence interval shows the range in which the actual value is expected to fall. For example, in Time Period 49, the forecast was 140,328, with a 95% confidence range from 117,630 to 163,027, and the actual value was 119,420 within the predicted range. Overall, this table helps evaluate how closely the model's forecasts align with real observations, and in most cases, the actual values are reasonably close to or within the forecast intervals, suggesting good model performance.

Figure 8 presents a time series graph comparing historical GHI data with future values forecasted using the SARIMA model. The blue line shows actual GHI values from Time 1 to 54, with clear fluctuations between 120,000 and 190,000, indicating trends and seasonal patterns. The red line represents forecasted values,

while the red line around it illustrates the 95% confidence interval, showing where future values are likely to fall. This interval is narrow immediately after the historical data but widens over time, reflecting increasing uncertainty. The graph helps visualize how well the model captures past patterns and projects future outcomes.

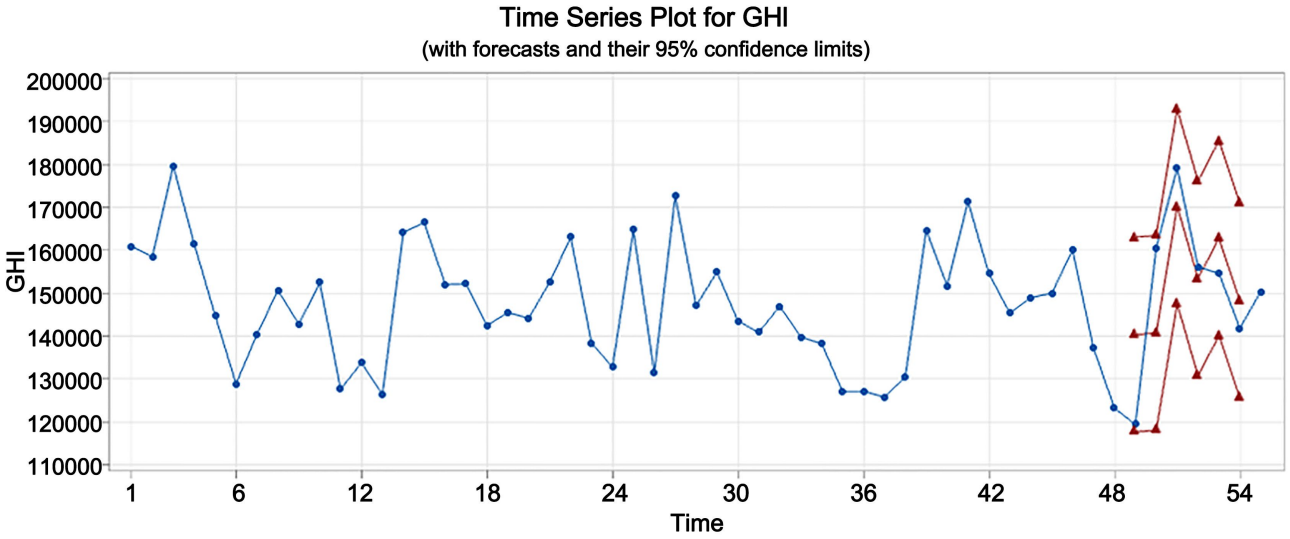

Figure 8. Forecast graph for monthly against the actual value with the confidence interval.

The forecasting analysis of GHI further supports these findings. Historical data (Time 1 - 48) fluctuated between 130,000 and 170,000, showing strong seasonal cycles without a significant long-term trend. Forecasted values (Time 49 - 54) follow a similar cyclical behavior, ranging from approximately 125,000 to 190,000. The 95% confidence intervals reflect moderate uncertainty due to short-term variability, but the overall stability of the seasonal pattern is well captured. This suggests that solar irradiance in Malaysia can be reliably projected for energy planning purposes, provided that seasonal adjustments and variability within the confidence bounds are considered. The results reinforce the potential of Malaysia's solar resource for supporting photovoltaic energy systems, while emphasizing the importance of incorporating both predictable seasonal cycles and short-term variability into system design and forecasting models.

The multi-scale analysis confirms that GHI in Melaka, Malaysia is characterized by predictable diurnal cycles, strong seasonal variation, and stable long-term behavior. The ACF and PACF results reinforced the presence of temporal dependence and annual seasonality, supporting the application of seasonal time series models for forecasting. From a practical perspective, these findings demonstrate the suitability of Melaka for solar photovoltaic deployment, while providing important insights for energy system design, storage optimization, and renewable energy policy in Malaysia.

Despite its promising performance, this study has several limitations. The univariate ARIMA model relies solely on historical GHI values and does not incor-

porate exogenous meteorological variables such as temperature, humidity, or cloud cover, which are known to influence solar irradiance. Future research could adopt multivariate models, including ARIMAX or hybrid machine learning approaches, to better capture these external drivers and potentially improve forecast accuracy.

5. Conclusions

This study examined the potential of time-series forecasting for solar PV power generation in Malaysia using GHI data collected in Melaka. The analysis of hourly, daily, and monthly irradiance trends revealed consistent diurnal and seasonal patterns, with peak values exceeding 900 Wh/m^2 under clear-sky conditions and lower peaks of around 700 Wh/m^2 during cloudy days. These variations underscore the influence of tropical weather on solar energy availability and highlight the importance of robust forecasting approaches. The Auto-Regressive Integrated Moving Average (ARIMA) model, supported by ACF and PACF analysis, effectively captured both short-term fluctuations and long-term seasonal trends in GHI, producing reliable forecasts when compared to observed data.

The results demonstrate that ARIMA provides a practical and statistically sound framework for solar energy forecasting in tropical climates, enabling improved PV system optimization and grid integration. In the Malaysian context, accurate irradiance prediction is vital for supporting renewable energy targets, enhancing grid reliability, and informing sustainable energy policy. For instance, short-term solar forecasts could assist grid operators in determining appropriate spinning reserve capacity, while reliable monthly projections could guide energy storage dispatch strategies to mitigate intermittency. Future research could extend this work by incorporating hybrid models, such as ARIMA combined with machine learning approaches, to further improve accuracy under highly variable weather conditions. Overall, the study affirms the critical role of advanced time-series models in advancing solar energy planning and long-term energy security.

Acknowledgements

The authors would like to express their most heartfelt appreciation to the Centre for Research and Innovation Management (CRIM), Universiti Teknikal Malaysia Melaka (UTeM) for their technical and administrative support throughout this study.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Yousif, J.H., Kazem, H.A., Alattar, N.N. and Elhassan, I.I. (2019) A Comparison Study Based on Artificial Neural Network for Assessing PV/T Solar Energy Production. *Case Studies in Thermal Engineering*, **13**, Article ID: 100407.

https://doi.org/10.1016/j.csite.2019.100407

[2] Prema, V., Bhaskar, M.S., Almakhles, D., Gowtham, N. and Rao, K.U. (2022) Critical Review of Data, Models and Performance Metrics for Wind and Solar Power Forecast. *IEEE Access*, **10**, 667-688. https://doi.org/10.1109/access.2021.3137419

[3] Nobre, A.M., Severiano, C.A., Karthik, S., Kubis, M., Zhao, L., Martins, F.R., *et al.* (2016) PV Power Conversion and Short-Term Forecasting in a Tropical, Densely-Built Environment in Singapore. *Renewable Energy*, **94**, 496-509. https://doi.org/10.1016/j.renene.2016.03.075

[4] Huynh, A.N., Deo, R.C., An-Vo, D., Ali, M., Raj, N. and Abdulla, S. (2020) Near Real-Time Global Solar Radiation Forecasting at Multiple Time-Step Horizons Using the Long Short-Term Memory Network. *Energies*, **13**, Article 3517. https://doi.org/10.3390/en13143517

[5] Thaker, J. and Höller, R. (2022) A Comparative Study of Time Series Forecasting of Solar Energy Based on Irradiance Classification. *Energies*, **15**, Article 2837. https://doi.org/10.3390/en15082837

[6] Al-Dahidi, S., Madhiarasan, M., Al-Ghussain, L., Abubaker, A.M., Ahmad, A.D., Alrbai, M., *et al.* (2024) Forecasting Solar Photovoltaic Power Production: A Comprehensive Review and Innovative Data-Driven Modeling Framework. *Energies*, **17**, Article 4145. https://doi.org/10.3390/en17164145

[7] Engel-Cox, J.A., Nair, N.L. and Ford, J.L. (2012) Evaluation of Solar and Meteorological Data Relevant to Solar Energy Technology Performance in Malaysia. *Journal of Sustainable Energy & Environment*, **3**, 115-124.

[8] Ledmaoui, Y., El Maghraoui, A., El Aroussi, M., Saadane, R., Chebak, A. and Chehri, A. (2023) Forecasting Solar Energy Production: A Comparative Study of Machine Learning Algorithms. *Energy Reports*, **10**, 1004-1012. https://doi.org/10.1016/j.egyr.2023.07.042

[9] (2023) National Energy Transition Roadmap (NETR). https://ekonomi.gov.my/sites/default/files/2023-08/National%20Energy%20Transition%20Roadmap.pdf

[10] Albeladi, K., Zafar, B. and Mueen, A. (2023) Time Series Forecasting Using LSTM and Arima. *International Journal of Advanced Computer Science and Applications*, **14**, 313-320. https://doi.org/10.14569/ijacsa.2023.0140133

[11] Sobri, S., Koohi-Kamali, S. and Rahim, N.A. (2018) Solar Photovoltaic Generation Forecasting Methods: A Review. *Energy Conversion and Management*, **156**, 459-497. https://doi.org/10.1016/j.enconman.2017.11.019

[12] AlShafeey, M. and Csáki, C. (2021) Evaluating Neural Network and Linear Regression Photovoltaic Power Forecasting Models Based on Different Input Methods. *Energy Reports*, **7**, 7601-7614. https://doi.org/10.1016/j.egyr.2021.10.125

[13] https://power.larc.nasa.gov/data-access-viewer/

[14] Box, G.E.P., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2016) Time Series Analysis: Forecasting and Control. 5th Edition, Wiley.