Energy Proceedings
Vol 43, 2024
ISSN 2004-2965

Predicting Photovoltaic Power Generation by Machine Learning Using Time Series Analysis

Afroza Nahar [1] *, Rifat Al Mamun Rudro [1], Md. Faruk Abdullah Al Sohan [1], Rubina Islam Reya [1] and Md. Hamid Uddin [1]

1 American International University-Bangladesh, Dhaka -1229, Bangladesh
(*Corresponding Author: afroza@aiub.edu)

ABSTRACT

Negative externalities of fossil fuels together with adjuvant features of solar energy is driving the global espousal of solar energy technologies. This article presents a forecasting model for photovoltaic (PV) power generation using real-time data analysis of two solar plants through machine learning time series model (MLTSM). The work focuses on critical factors such as predictive accuracy, residual distribution, RMSE values, data quality, and model suitability for forecasting. The findings demonstrate that the predictive model achieves an accuracy of 98% for Plant 1 and 91% for Plant 2. Overall, the MLTSM exhibits its effectiveness in enhancing PV power generation forecasting, thereby contributing to the attainment of energy security.

Keywords: solar energy, forecasting, machine learning, time series, energy security

NONMENCLATURE

Abbreviations	
MLTSM	Machine Leaning Time Series Model
PV	Photovoltaic
Symbols	
α, β	Module constants

1. INTRODUCTION

The issue of clean energy has become one of the grave global concerns in the last decade and the United Nations also recognized access to clean and affordable energy as one of the sustainable development goals. Solar energy is accessible all over the globe as a clean and sustainable energy and electricity generation through solar photovoltaic (PV) modules is widely regarded as the straightforward means to produce clean energy [1]. However, the primary challenge with solar power generation consists in dispersed nature characterized by low power density, and the inherent unpredictability of its availability.

Prediction of the availability of solar irradiation is not very easy since it depends on climate conditions that always fluctuate over time. To solve this issue, machine learning (ML) methods have received significant attention from many researchers and developers in the solar power generation field. Atique et al. [2] compared the finding of SVM's superiority, wherein SVM handled non-parametric residuals, captured linear/non-linear relationships, and enhanced accuracy. Limitations encompassed data requirements, overfitting, interpretation challenges, neural network constraints, and the need for accuracy enhancement through research. Kudarihal et al. [3] focused on leveraging PV systems in Hyderabad, a region with moderate solar energy potential, showcasing a daily generation of over 4 kWh AC electricity. The study evaluates an 8-kW rooftop solar plant's operational performance through analyzing the potential of solar energy accessibility. Garibo-Morante et al. [4] introduced an innovative time series modeling approach using univariate and multivariate harmonic decomposition within a state space framework derived from Fourier analysis. This enabled precise signal description and prediction through optimal state estimation, with applications in power demand and wind/solar forecasting. The method offered effective pattern understanding, adaptable modeling, and optimal estimation benefits. Limitations included complexity with non-harmonic signals, data quality dependency, computational demands, and interpretability challenges. Successful applications demonstrated their practical utility within specific contexts, emphasizing informed implementation. Yan et al. [5] studied DC transmission prominence and

This is a paper for 15th International Conference on Applied Energy (ICAE2023), Dec. 3-7, 2023, Doha, Qatar.

introduced a comprehensive model for optimized wind and solar power output within system constraints. The model integrated demand-supply dynamics, various plant types, and DC transmission intricacies. Case study demonstrated practical grid planning insights, acknowledging the role of DC transmission. Limitations included potential oversimplification and data constraints, highlighting the need for cautious real-world application, and understanding of the proposed grid planning approach. Ananthu et al. [6] addressed growing solar energy integration in local grids, proposing a robust solution: data analysis using advanced AI/ML techniques, notably a sophisticated Deep Neural Network (DNN) with LSTM networks. Authors demonstrated accurate PV power generation forecasts using historical data from a 100-kWp solar power plant, surpassing other models like ARIMA, SARIMA (Seasonal Autoregressive Integrated Moving Average), RNN, and fbProphet. The evaluation included metrics like RMSE and CUF. Recognizing limitations in model precision, data reliance, and computational complexity, the study advances solar energy forecasting methodologies through potent AI-driven paradigms in energy systems.

Predicting power generation from solar PV power plants is a necessary step for future development and reliable prediction of the plants power generation capacity is exigent in this regard. For this purpose, it is important to employ new, and intelligent methods to obtain valid and accurate results. Although many researchers already employed machine learning techniques like SVM and ANN an innovative time series modeling approach using univariate and multivariate harmonic decomposition within a state space framework derived from Fourier analysis might play game changer role in this field. But such a model is yet to be developed and tested. In this article, we introduce a comprehensive MLTSM that employs a predictive model performance. This model serves the dual purpose of analyzing diverse functional and performance aspects across a wide spectrum of PV system configurations and optimizing their performance under varying ambient conditions, thereby providing tangible benefits to end-users.

2. METHODOLOGY

2.1 Workflow of proposed MLSTM

The proposed MLTSM depicts the performance of PV modules and inverters under a given situation and uses power modelling to forecast the performance of the solar power plants. Fig. 1. illustrates the operational procedure of the solar power forecasting system. The model has been validated through real-world dataset which encompasses data (comprising of power generation and sensor readings) collected over 34 days from two solar power plants in India.

2.2 Mathematical Model

In the context of PV systems, vital variables like module temperature (Tm), ambient temperature (Ta), and ground-level sun irradiance (Gir) intricately impact PV performance. These factors, linked with module constants (α, β) and standard conditions (STC) of T0 = 25°C, G0 = 1000 W/m², significantly influence power output (Pac) under standard conditions. The PV power output Pac (considering the inverter) is modeled through Vth (Thevenin voltage) and Ino (Mayer-Norton current) relationships as below:

$$P_{ac} = V_{th} \times I_{t1h} \quad (1)$$

Where V_{th} and I_{no} are defined by:

$$V_{th} = V_o[1 + \beta(T_m - T_0) \quad (2)$$

Scatter plots are used to enhance AC power generation prediction accuracy by visualizing correlations between atmospheric temperature, irradiation, and AC power, ultimately identifying a positive correlation supporting a physics-based model. Then, the data is divided, regressors are chosen, and

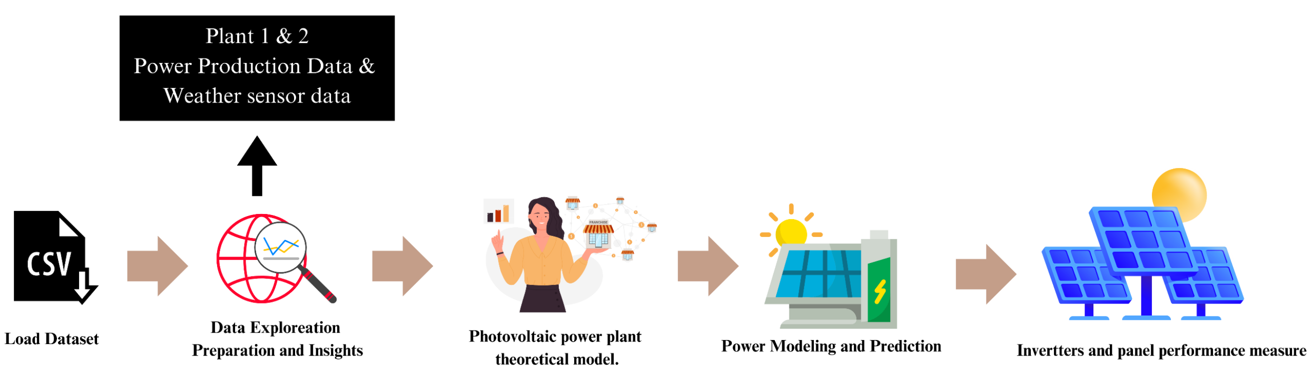

Fig. 1. Methodology Workflow of MLTSM

individual regressors are trained and tested for each inverter to evaluate performance. The aim is to combine these results to enhance predictive capability by accounting for interconnections between these factors. Next, physics-based models are employed to generate new features, enriching the data frames with computed values (eqn. 5) tied to irradiance, ambient temperature, and AC power. These enriched features are utilized for training and evaluating different regressor models such as Linear Regression, Ridge Regression, Decision Tree Regression, Random Forest Regression, and K-Nearest Neighbors Regression. Models are ranked based on negative mean squared error through cross-validation, emphasizing the top performers. Subsequent tasks involve developing models for individual inverters, producing forecasts, and assessing performance using root mean squared errors (RMSE).

3. *RESULT ANALYSIS*

Critical factors for faithful prediction of PV power generation include predictive accuracy of the model, residual distribution, RMSE values, data quality, and model relevance for prediction. A comparative portrayal of the actual and MLTSM predicted power generation (termed as AC_Power) for two specific inverters within Plant 1 and Plant 2 have been presented in Fig. 2 and Fig. 3.

$$I_{no} = I_o \,[1 + \alpha\,(T_m - T_0)] \quad (3)$$

The module temperature is directly proportional to the Irradiation and to the ambient temperature and can be represented by the following empirical formula:

$$Tm = 30 - 0.0175\,(G_{ir} - 300) + 1.14\,(T_a - 25) \quad (4)$$

Replacing equations (2), (3) and (4) into equation (1) gives a third-degree polynomial equation that can be expressed as follows:

$$P_{ac} = K_1G^3_{ir} + K_2G^2_{ir} + K_3G^2_{ir}T_a + K_4G_{ir}T^2_a + K_5G_{ir}T_a + K_6G_{ir} \quad (4)$$

Where K_1, K_2, ..., K_6 are constants.

3.1 Framing the prediction model

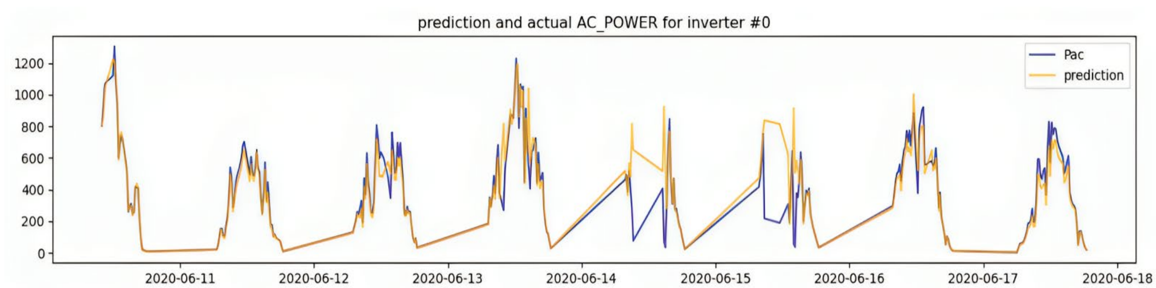

Fig. 3. Prediction and actual AC_POWER Plant 2

The residual distribution patterns suggest that Plant 1 has significant kurtosis and left-skewed residuals, while Plant 2 has a wider distribution variance and inverters with different accuracy. RMSE values are similar, with Plant 1 having lower values and Plant 2 greater. The data quality study suggests Plant 1's data could be better than Plant 2's. Plant 1's inverter-specific accuracy is consistently good, but Plant 2's is mixed, with several inverters below the 80% R-squared threshold. Plant 1 has a 98% R-squared score for aggregate accuracy, whereas Plant 2 has 91%. Plant 1 is highly trustworthy, and Plant 2 is good for forecasting insights. Plant 1's great precision limits progress, while Plant 2's model allows for it.

4. CONCLUSIONS

In this article, we introduce a comprehensive MLTSM model that can faithfully forecast power generation patterns of different PV systems across a broad spectrum of configurations and ambient conditions. During the model development process, features are carefully chosen based on their temporal variability and their interaction with irradiance, ambient temperature, and AC power data is considered. The results indicate that the model's accuracy in predicting photovoltaic (PV) power generation exceeds 90%, demonstrating a highly satisfactory performance.

ACKNOWLEDGEMENT

Authors would like to express their gratitude to American International University–Bangladesh (AIUB) for providing technical support to carry out this research.

DECLARATION OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. All authors read and approved the final manuscript.

REFERENCE

[1] Nahar, A., Hasanuzzaman, M., & Rahim, N. A. Numerical and experimental investigation on the performance of a photovoltaic thermal collector with parallel plate flow channel under different operating conditions in Malaysia. Solar Energy 2017;144:517-528.

[2] Atique, S., Noureen, S., Roy, V., Bayne, S., & Macfie, J. (2020, April 1). *Time series forecasting of total daily solar energy generation: A comparative analysis between ARIMA and machine learning techniques*. IEEE Conference Publication | IEEE Xplore

[3] Kudarihal, C. S., Kumar, S., & Gupta, M. (2023, February 22). *Econometrics and Time Series analysis of a Grid-Connected rooftop solar system and Prosumers experience in a smart Grids scenario: a case study*. IEEE Conference Publication.

[4] Garibo-Morante, A. A., & Tellez, F. O. Univariate and Multivariate Time Series Modeling using a Harmonic Decomposition Methodology. IEEE Latin America Transactions 2021;20:372-378.

[5] Yan, Z., Guo, L., Li, C., Wang, S., & Sang, Z. (2022, May 27). *Research on Wind-Solar-Thermal-Storage Configuration schemes of DC Near-field based on Time Sequential Simulation*. IEEE Conference Publication

[6] Ananthu, D. P., & K, N. (2021, April 8). *A study of 100kWp PV Plant Output Power Forecasting: A case study*. IEEE Conference Publication.