Journal scientific and applied research, vol. 29, 2025
International Journal

ISSN 1314-6289 (Print)
ISSN 2815-4622 (Online)

Original Contribution

EVALUATION OF FORECASTING METHODS FOR PHOTOVOLTAIC SYSTEMS IN TERMS OF ACCURACY AND COMPUTATIONAL REQUIREMENTS

Iliyana St. Ivanova

DEPARTMENT OF COMMUNICATION AND COMPUTER ENGINEERING AND SECURITY TECHNOLOGIES, FACULTY OF TECHNICAL SCIENCES, KONSTANTIN PRESLAVSKY UNIVERSITY OF SHUMEN, SHUMEN 9712,115, UNIVERSITETSKA STR., E-MAIL: i.s.ivanova@shu.bg

ABSTRACT: *Forecasting photovoltaic (PV) production is an essential component for the successful integration of renewable energy into electricity grids. In this paper, four main groups of methods are evaluated: statistical, physical, machine learning-based, and hybrid. An analysis is made in terms of accuracy, computational complexity and applicability using an evaluation scale. The results show that hybrid models appear particularly promising, as they combine the main strengths of the other approaches.*

KEY WORDS: *Photovoltaic system, Forecasting, Machine learning, Statistical models, Hybrid methods.*

1. Introduction

The growing share of renewables in the global energy mix requires reliable methods for forecasting meteorological parameters. Accurate forecasts of electricity generation from PV systems are critical for utilities, especially in countries where legislative initiatives are actively promoting the construction of solar power plants [1,2].

Photovoltaic systems have become a vital component of global electricity generation. However, their variability and limited predictability pose major challenges for grid operators. Effective forecasting supports load management and reduces grid imbalances, leading to greater power system stability [3].

This paper discusses and evaluates four main groups of methods: statistical, physical, machine learning-based, and hybrid. The working principles, advantages and limitations are presented. An evaluation scale is applied to enable comparison in terms of accuracy, computational complexity, interpretability, data volume, forecast period and applicability.

2. Forecasting methods for PV

2.1. Statistical forecasting methods for PV.

Statistical methods are based on time series and use historical data for forecasting. They are effective for short-term forecasts but have limited accuracy during sudden weather changes. These methods are commonly used, for instance:

2.1.1. The ARIMA (AutoRegressive Integrated Moving Average) method is one of the most common statistical methods for time series forecasting and is widely used in PV power generation forecasting. Its basic idea is that the future power output of a PV installation can be expressed as a linear combination of previous values of the power produced (autoregressive component), previous errors (moving average component), and an appropriate differencing that removes trends and seasonality. The model is usually denoted as ARIMA (p, d, q), where the parameter p specifies the number of included previous power values, d the degree of differencing, and q the number of included previous errors.

When forecasting the performance of PV systems, ARIMA is most used for short-term horizons - minutes to a few hours ahead. It is particularly useful when reliable weather forecasts are lacking, and the forecast must be based primarily on historical solar radiation measurements or power output. Among the advantages of ARIMA is its ease of implementation and the fact that it does not require large amounts of training data, making it applicable to newly installed PV systems with a limited measurement base. However, the accuracy of the model drops under highly dynamic atmospheric conditions - for example, when clouds cause unpredictable changes in sunshine. ARIMA ignores external factors such as cloud cover and temperature. Its predictive value is thus confined to situations where historical PV patterns alone can describe future behavior. In cases where there are clear seasonal patterns in the data, the extended SARIMA (Seasonal ARIMA) model is often applied to better reflect the daily and seasonal cycles in solar electricity production.

2.1.2. The SARIMA (Seasonal AutoRegressive Integrated Moving Average) model is an extension of the classical ARIMA model, which allows more accurate modelling of time series with pronounced seasonality. At its core, SARIMA includes the standard AR, I and MA components, but seasonal parameters are added to account for the recurrence of the data over a given period. Thus, the model is denoted as SARIMA (p, d, q) (P, D, Q) s, where the first three parameters describe the non-seasonal part and P, D, Q and s reflect the seasonal autoregressive and moving average components, the degree of seasonal differencing and the length of the seasonal cycle, respectively.

SARIMA is particularly well suited for forecasting the production of PV systems due to the clear daily and seasonal repeatability of solar radiation and generated power. For example, the daily cycle of sunrise and sunset, as well as the annual variation of day length and solar elevation angle, create regular patterns that classical ARIMA has difficulty capturing. By incorporating seasonal components, SARIMA achieves more accurate predictions than traditional statistical models, especially at short- and medium-term horizons. The advantages of SARIMA are related to its ability to detect complex dependencies in the data without requiring external weather information. It is relatively easy to implement and widely applicable in situations where sufficient historical data with distinct seasonality is available. Its drawbacks stem from its linear nature, like ARIMA. The model struggles to capture sudden non-linear changes, such as those caused by rapidly shifting cloud cover. In addition, SARIMA can become overly complex when multiple parameters are included, which can reduce the model's generalizability.

In practice, SARIMA is often used as an intermediate method - more accurate than standard ARIMA due to the seasonal components, but still more limited than modern machine learning techniques. SARIMA has been applied in several studies to predict hourly solar radiation and power generation, with results showing an improvement in accuracy compared to baseline models, especially at intervals of a few hours to a day. In this sense, SARIMA represents a valuable tool that can be used alone or as a component in hybrid models combining statistical and intelligent methods.

2.1.3. The VAR (Vector Autoregression) model extends the classical autoregressive approach to capture interdependencies among multiple variables simultaneously, instead of a single time series. In a VAR, each variable is expressed as a linear function of its own past values and those of the other variables in the system. This results in a set of equations that captures the dynamic interrelationships in multivariate data.

In forecasting PV generation, VAR is commonly used because it allows simultaneous modelling of multiple factors affecting the generated power. In addition to the time series of solar radiation, ambient temperature, wind speed, cloud cover, or even power output of neighboring PV installations can be included in the model. Thus, VAR is a more comprehensive model compared to ARIMA and SARIMA, which work mainly with single data series. The VAR model captures complex linear interdependencies among multiple variables and yields results that remain interpretable. This model is particularly useful when the data shows strong interdependence, as is often observed with meteorological variables. In addition, VAR models are also used to conduct scenario analyses using impulse response functions, which show how a change in one variable affects the others in the system.

The VAR model is characterized by high sensitivity to many parameters, especially when multiple variables are included in the analysis. This leads to the risk of loss of generalizability and the need for larger amounts of data for robust parameter estimation. Although effective for short- and medium-term forecasts, VAR has limited capacity to capture nonlinearities and abrupt atmospheric changes, which reduces its accuracy under dynamic weather conditions [4,5].

2.2 Physical methods

Physical methods for forecasting PV generation model the processes that convert solar radiation into electricity. Unlike statistical approaches, which rely mainly on historical data, these methods combine meteorological forecasts with the technical characteristics of PV installations. Input data includes solar radiation, cloud cover, temperature, wind speed and atmospheric transparency, while solar-geometric models consider the position of the sun, tilt and orientation of the panels. The results obtained are converted into expected power generation using simulation tools such as PVGIS, System Advisor Model (SAM), PVWatts, etc.

Leading this category are Numerical Weather Prediction (NWP) based methods that combine predicted weather conditions (temperature, cloud cover, radiation, wind speed) with physical models of the PV modules. This estimates the influence of factors such as panel temperature, angle of incidence of radiation, shading and other system characteristics. Such methods are particularly suitable for medium-term (a few hours to days) and long-term (weeks) forecasts, where historical data alone do not provide reliable predictive value.

Physical models provide a more accurate representation of the dynamics of PV systems and their interaction with the environment, reducing the risk of system biases. Their main limitation, however, is the high computational complexity. Numerical weather models are highly computationally intensive, and their accuracy depends on the quality of the input data and on the ability of the physical models to describe the real characteristics of PV systems. Despite these limitations, physical methods can offer more reliable predictions in the medium and long term compared to statistical methods. The integration of satellite data further improves the accuracy of short-term forecasts. Physical methods are also applicable to newly built PV installations for which long-term measurements are not available [6,7,8].

2.3. Machine learning methods

Machine learning methods provide an approach to predicting the output of photovoltaic (PV) systems using algorithms that capture complex and non-linear relationships between input factors and output power. In contrast to statistical and physical models based on predefined dependencies, machine learning algorithms learn directly from data and extract hidden relationships. Input data includes historical values of power output, solar radiation, temperature, cloud cover and other meteorological parameters.

Methods in this category include:

- Artificial Neural Networks (ANN) - classical neural networks that capture non-linear dependencies between historical data and output. They are particularly suitable for short- and medium-term forecasting.
- Long Short-Term Memory (LSTM) - recurrent networks designed to model long-term dependencies and seasonal fluctuations in time series. They achieve high accuracy in forecasting PV generation.
- Convolutional Neural Networks (CNN) - neural networks oriented to extract local and spatial dependencies. They find applications in weather maps and spatial data analysis.
- Random Forest - an ensemble method based on multiple decision trees that provides greater stability and robustness of predictions.
- Gradient Boosting - an ensemble approach where models are built sequentially and focus on correcting residual errors. This method is suitable for handling complex and dynamic data [9,10,11,12,13].

2.4. Hybrid Methods

Hybrid methods for predicting production from photovoltaic (PV) systems combine the advantages of different approaches - statistical, physical and machine learning algorithms. Their main goal is to overcome the limitations of individual methods. Hybrid methods for predicting production from PV systems can be implemented in several ways.

a) Combining physical and machine-learning models, where weather forecasts or physical simulations are used as input to neural networks or other algorithms.
b) Combining statistical and machine learning methods, for example when a baseline forecast obtained by ARIMA is improved using LSTM or Gradient Boosting.
c) Ensemble hybrids that combine the predictions of statistical, physical and intelligent models through weighting techniques or meta-models.
d) Integration of data from different sources, such as satellite imagery, historical measurements, and forecast atmospheric data, resulting in higher accuracy [14,15,16,17,18].

3. Comparative analysis of forecasting methods using an evaluation scale

The feasibility of different methods for forecasting PV production is assessed using a five-point scale, enabling comparison across key criteria. The scale is defined as follows:

1- Very low: minimum accuracy/minimum computational requirements.

2- Low: limited accuracy/ computational complexity required.

3- Medium: satisfactory accuracy/moderate computational demands.

4- High: good accuracy/high computational complexity.

5- Very high: excellent accuracy/extremely high computational complexity.

Table 1 presents a comparative analysis of the four main categories of methods - statistical, physical, machine learning and hybrid methods.

Table 1 Evaluation of forecasting methods

Method	Statistical methods	Physical methods	Machine learning methods	Hybrid Methods
Accuracy	3	4	4-5	5
Computational complexity	2	5	3-4	5
Interpretability	4-5	4	1-2	2-3
Required data volume	3	4	4-5	5
Forecast period	Short-term (minutes to hours)	Medium and long-term forecasts (hours, days, weeks)	Short-term and medium-term forecasts	Applicable across all time horizons, from short- to long-term
Applicability	Suitable for newly installed systems with limited data and for basic comparisons	Used in large installations and where detailed meteorological data is available	Suitable for systems with extensive historical databases	Applicable to both local systems and power grids

The evaluation scale (1-5) used in the table allows a clear comparison between the different groups of methods for predicting production from PV systems.

The statistical models obtain values around 3 for accuracy and 2 for computational complexity. This confirms their basic profile - they are easy to implement and have low computational requirements, but their accuracy remains limited under high weather dynamics. These results indicate that statistical methods are primarily suitable for short-term forecasts and for systems with limited historical data.

Physical methods are rated 4 in accuracy and 5 in computational complexity. This reflects their high reliability in the medium to long term, owing to the inclusion of NWP predictions and PV simulations. However, it also highlights the significant resources required for implementation. As a result, these methods are preferred for large installations and regional forecasts, provided sufficiently detailed input data is available.

Machine learning methods show values in the range 4-5 for accuracy and 3-4 for computational complexity. The results confirm their capacity to model complex non-linear dependencies but also point to higher requirements for data and computational resources. Despite being evaluated as balanced and flexible, their practical application remains limited when adequate input data is unavailable.

Hybrid models obtain the highest ratings (5 for accuracy and 5 for computational complexity), positioning them as the most promising category by integrating the strengths of statistical, physical, and machine learning approaches. However, substantial resource requirements and the complexity of integrating diverse methods remain key obstacles to their large-scale implementation.

The analysis of the evaluations highlights that at short-term horizons statistical models are sufficient, for long-term forecasts physical approaches dominate, while machine learning provides a balance between accuracy and resources. Hybrid solutions demonstrate the highest potential for future development, especially in the integration of heterogeneous data and their application in large-scale energy systems. Furthermore, the integration of Geographic Information Systems (GIS) can enhance both accuracy and applicability by providing spatial context and real-time data analysis, opening opportunities for combining forecasting methods with spatial analyses in future developments [19,20].

4. Conclusion

The evaluation framework presented in this study demonstrates that each forecasting method has specific applications in PV production prediction. The overall trend favors combined approaches, with hybrid models emerging as the most effective option.

Future research should prioritize the integration of diverse data sources - such as satellite observations, IoT sensors, and real-time measurements - alongside improvements in algorithmic accuracy and computational efficiency. These advancements will enhance the robustness and real-time applicability of forecasts, thereby facilitating the seamless integration of PV systems into modern power grids.

Acknowledgments

This article is supported by internal research project RD-08-124/07.02.2025 "Renewing the research environment for collecting empirical data in measurement processes", Konstantin Preslavsky University of Shumen,

Faculty of Technical Sciences, Department of Communication and Computer Engineering and Security Technologies.

References:

[1] Qaisar, S., Awan, A.B., Rehman, S., Al-Hebshi, A., & Al-Sumaiti, A.S. Solar Radiation Forecasting: A Systematic Meta-Review of Current Methods and Emerging Trends. Energies, 17(13), 3156, 2024, https://doi.org/10.3390/en17133156.

[2] Energies. Forecasting Solar Photovoltaic Power Production: A Comprehensive Review and Innovative Data-Driven Modeling Framework. Energies 17, no. 16 (2024): 4145, https://www.mdpi.com/1996-1073/17/16/4145.

[3] Energies. Advances in Short-Term Solar Forecasting: A Review and Benchmark of Machine Learning Methods and Relevant Data Sources. Energies 17, no. 1 (2024): 97, https://www.mdpi.com/1996-1073/17/1/97.

[4] Mellit, A., & Pavan, A. (2010). A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. Solar Energy, 84(5), 807–821. Elsevier, https://doi.org/10.1016/j.solener.2010.02.006.

[5] Voyant, C., Notton, G., Kalogirou, S., Nivet, M.L., Paoli, C., Motte, F., & Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. Renewable Energy, 105, 569–582. Elsevier, https://doi.org/10.1016/j.renene.2016.12.095.

[6] Tsai, W.-C., Huang, C.-M., & Hsu, C.-T. A Review of State-of-the-Art and Short-Term Forecasting of Solar PV Power. Energies, 16(14), 5436, 2023, https://doi.org/10.3390/en16145436.

[7] Yang, D., Kleissl, J., Gueymard, C. A., Pedro, H. T. C., & Coimbra, C. F. M. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. Solar Energy 168 (2018): 60–101. Elsevier, https://doi.org/10.1016/j.solener.2017.11.023.

[8] Lara-Fanego, V., Ruiz-Arias, J. A., Pozo-Vázquez, D., Santos-Alamillos, F. J., & Tovar-Pescador, J. Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain). Solar Energy 86, no. 8 (2012): 2200–2217. Elsevier, https://doi.org/10.1016/j.solener.2011.02.014.

[9] Singh, J. G., et al. A comprehensive review of machine learning applications in forecasting solar PV and wind turbine power output. Journal of Electrical Systems and Information Technology, 12, Article 54, 2025. SpringerOpen, https://doi.org/10.1186/s43067-025-00239-4.

[10] Gupta, M., Sharma, R., & Singh, A. A review of PV power forecasting using machine learning. Renewable Energy, (2025). Elsevier, https://www.sciencedirect.com/science/article/abs/pii/S2950425225000106

[11] Alzahrani, A., Alshammari, R., & Alotaibi, M. Machine learning and deep learning models-based grid search for short-term solar irradiance forecasting. Journal of Big Data, 11, 52 (2024). SpringerOpen, https://doi.org/10.1186/s40537-024-00991-w.

[12] Raza, M. Q., & Khosravi, A. A review on artificial intelligence based load demand and renewable energy forecasting. International Journal of Forecasting 35, no. 1 (2019): 136–174. Elsevier, https://doi.org/10.1016/j.rser.2015.04.065.

[13] Benti, N.E., Chaka, M.D., & Semie, A.G. Forecasting Renewable Energy Generation with Machine Learning and Deep Learning: Current Advances and Future Prospects. Sustainability 15, 7087 (2023), https://doi.org/10.3390/su15097087.

[14] Liu, H., Cai, C., Li, P., Tang, C., Zhao, M., Zheng, X., Li, Y., Zhao, Y., & Liu, C. "Hybrid prediction method for solar photovoltaic power generation using normal cloud parrot optimization algorithm integrated with extreme learning machine." Scientific Reports, 15, Article number: 6491 (2025), https://www.nature.com/articles/s41598-025-89871-8.

[15] A. Hussain, M., et al. A Hybrid Deep Learning-Based Network for Photovoltaic Power Forecasting. Scientific Reports 12 (2022): 7040601, https://doi.org/10.1155/2022/7040601.

[16] Salman, D., Direkoglu, C., Kusaf, M., et al. Hybrid deep learning models for time series forecasting of solar power. Neural Computing and Applications, 36 (2024): 9095–9112, https://doi.org/10.1007/s00521-024-09558-5.

[17] Ladjal, B., Nadour, M., Bechouat, M., Hadroug, N., Sedraoui, M., Rabehi, A., Guermoui, M. et al. Hybrid deep learning CNN-LSTM model for forecasting direct normal irradiance: A study on solar potential in Ghardaia, Algeria. Scientific Reports, 15, Article 15404 (2025), https://doi.org/10.1038/s41598-025-94239-z.

[18] Attya, M., Abo-Seida, O., Mohamed, H., Mohammed, A., & others. A hybrid deep learning framework for solar irradiation prediction based on regional satellite images and data. Neural Computing & Applications (2025). Springer, https://doi.org/10.1007/s00521-025-11197-3.

[19] Dimanova, D., Geographical information systems and their application in various fields, Journal Scientific and Applied Research, Vol. 27 No. 1, 2024. DOI: https://doi.org/10.46687/jsar.v27i1.409

[20] Dimanova, D., Kuzmanov, Z., Real-time GIS for monitoring and security, Journal Scientific and Applied Research, Vol. 27 No. 1, 2024. DOI: https://doi.org/10.46687/jsar.v27i1.408