

Engineering Research Express



PAPER

Enhancing the reliability of solar PV power modelling through HBA-optimized variational autoencoder and GRU neural networks

RECEIVED
24 January 2026

REVISED
20 July 2026

ACCEPTED FOR PUBLICATION
29 July 2026

PUBLISHED
10 August 2026

Mariappan D¹ , Jain Vinith P R² and Garkki B^{3,*}

¹ Department of Electronics and Communication Engineering, Periyar Maniammai Institute of Science and Technology, Thanjavur, Tamil Nadu 613403, India

² Department of Electrical and Electronics Engineering, PSG Institute of Technology and Applied Research, Coimbatore, Tamil Nadu 641062, India

³ Department of Energy science and Engineering, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India

* Author to whom any correspondence should be addressed.

E-mail: garkki98@gmail.com

Keywords: solar photovoltaic power (SPV), forecasting, imputation, grid stability, missing data

Abstract

Solar photovoltaic (SPV) technology is gaining prominence due to its sustainability and minimal carbon footprint. However, the intermittent nature of SPV systems poses a challenge to widespread adoption. Accurate forecasting of SPV power is essential, relying heavily on high-quality data. Meteorological parameters from the National Renewable Energy Laboratory (NREL) often contain missing values, impacting the reliability of power system planning. These missing values can stem from various issues such as sensor malfunctions, communication errors, or database issues. Therefore, this paper introduces a variational autoencoder technique to construct an SPV power imputation model, aiming to fill in gaps in unobserved data. Moreover, the filled SPV power data is used to train a gated recurrent unit neural network for forecasting the SPV power. The hyperparameters of the imputation and forecasting models are optimized using the hummingbird optimization algorithm. Furthermore, the NovoGrad optimizes the hyperparameters of the imputation and forecasting models during training, thereby enhancing their performance and convergence. Furthermore, the proposed imputation and forecasting models are tested for various missing patterns. The effectiveness of the proposed imputation and forecasting models is demonstrated with real-world SPV power data for imputation and forecasting tasks.

1. Introduction

Solar photovoltaic (SPV) power generation is one of the most prominent renewable energy technologies, offering an environmentally sustainable solution to address the growing global demand for electricity. Nevertheless, the inherent intermittency of solar energy introduces significant challenges for power system operators, energy planners, and policymakers. The variability and uncertainty associated with solar power must be carefully accounted for in operational and planning decisions [1]. Moreover, uncertainty in solar generation influences the participation of solar farms in electricity markets. Consequently, accurate SPV power forecasting plays a crucial role in mitigating these challenges. In contrast, poor forecasting accuracy can adversely affect system security and operational reliability [2]. Therefore, improving the accuracy of SPV power forecasting is of critical importance for both power system operators and solar power plant owners. To ensure grid stability, optimize resource allocation, and mitigate the effects of SPV power intermittency on energy markets, accurate forecasting techniques are required to handle fluctuations in SPV power caused by meteorological parameters, seasonal changes, or geographical factors.

Generally, there are two ways to forecast SPV power such as direct and indirect over the long-term horizon [3]. The direct forecasting model uses data gathered from operational SPV generators. Conversely, the indirect forecasting model relies on data derived from meteorological parameters. It

should be noted that collecting data from existing SPV generators is critical for developing a direct forecasting model. Therefore, this paper focuses on the development of an indirect long-term SPV power forecasting model.

In general, satellite images offer a broad spectrum of meteorological parameters. However, their spatial dispersion leads to variations in actual meteorological conditions at specific locations. The data provided by ground station radiometers can also be used as an alternative source of collecting meteorological parameters. Nonetheless, issues such as communication failures and instrumentation problems often result in significant gaps in meteorological data. Moreover, historical meteorological data frequently contain missing data, with approximately 40% of meteorological parameters missing in real-world scenarios. Consequently, an accurate imputation technique is necessary for filling in missing historical SPV power data. Thus, this paper concentrates on developing indirect long-term SPV power imputation and forecasting models.

Mostly, SPV power forecasting techniques are commonly categorized into four primary groups such as mathematical technique, machine, deep, and hybrid learning techniques. Among the mathematical techniques, the persistence model, auto-regression model, auto-regressive moving average (ARMA) model, autoregressive integrated moving average (ARIMA) model, and multi-variant ARMA (ARMAX) model [4] are used to forecast SPV power through linear prediction and regression techniques. However, SPV power typically exhibits non-linear behaviour in response to changes in its meteorological parameters. Usually, linear forecast models may struggle to accurately capture the non-linear relationship of the data.

In recent years, machine learning-based approaches for SPV power forecasting have attracted considerable attention due to their ability to achieve satisfactory predictive performance [5]. This class of methods includes techniques such as extreme learning machines [6], multilayer perceptrons [7], support vector regression [8], radial basis function networks, and gradient boosting machines [9], all of which are well suited for capturing the nonlinear characteristics of SPV power generation. As a result, these models have been widely applied to SPV power forecasting. However, as reported in [10], a comparative study of several machine learning methods for hourly SPV power prediction indicates that these models are susceptible to overfitting. This issue often arises from limited training data or overly complex model structures, leading to reduced generalization performance during the testing phase.

Further, deep learning techniques such as convolutional neural networks (CNNs) [11], long short-term memory (LSTM) networks [12], and deep belief networks [13] have been used for developing the SPV power forecasting model. Certain investigations suggest that deep neural networks could outperform traditional neural networks in SPV power forecasting tasks. Nonetheless, there are also discussions, as seen in [14], suggesting that deep learning neural networks might not perform efficiently with limited data.

Hybrid techniques hybridize different forecast techniques to enhance their performance for SPV power forecasting [8]. For instance, ensemble models have been combined with a back propagation neural network model for forecasting SPV power [15]. In [16], a hybrid of support vector regression model with a grasshopper optimization algorithm and the Boruta-based feature selection algorithm is used for SPV power forecasting. An integration of two deep neural networks, including LSTM and CNNs, is presented for SPV forecasting [17]. The SPV power uncertainty, caused by the intermittent behaviour of meteorological parameters is discussed. Then [18], presents a forecasting technique, composed of wavelet transform, support vector machine, and particle swarm optimization, for long-term SPV power forecasting. Although PSO is often effective, it can have premature convergence problems and struggles with search spaces with high dimensions. It also requires careful parameter tuning.

Another important classification for SPV power forecasting techniques is based on the forecasting horizon, which divides these techniques into four categories such as very short-term, short-term, medium-term, and long-term [19]. Very short-term SPV power forecasting techniques are useful for allocating optimal reserves, electricity dispatch, and power smoothing; short-term SPV power forecasting techniques are mainly applied for unit commitment and grid security evaluation; medium-term SPV power forecasting techniques are used for smoothing power system maintenance schedules; and long-term SPV power forecasting helps in the system planning [20]. This paper focuses on the development of a long-term SPV power forecasting model.

Various studies have been conducted to replace missing values in SPV power forecasting. In [21], imputation techniques were applied to training and test data to analyse the impact of missing SPV power on forecasting performance, and a forecasting model was developed and compared with a perfect information-based model. In [22], the SPV power at locations that were not directly measured was estimated using multiple linear regression—a statistical-based technique to replace the missing values in SPV power data measured at nine weather stations. Then SolarGAN [23] was used to replace missing

7. Conclusion

This paper presented an integrated optimization-driven framework for robust SPV power imputation and forecasting under incomplete data conditions. The proposed methodology combines LPI for initial missing-data estimation, a VAE for probabilistic reconstruction of missing observations, the HBA for automatic hyperparameter optimization, the NovoGrad optimizer for efficient network parameter learning, and a GRU for temporal forecasting. Unlike conventional approaches that treat imputation and forecasting as independent tasks, the proposed framework establishes a sequential learning pipeline in which reconstructed SPV data are utilized to improve the reliability of the subsequent forecasting model.

The proposed framework was evaluated under multiple missing-data mechanisms, including MCAR, MNAR, and MP, with missing rates ranging from 30% to 90%. Comprehensive experimental studies conducted on datasets collected from diverse geographical locations and under different temporal resolutions demonstrated that the proposed framework consistently achieved lower imputation and forecasting errors than the benchmark methods considered in this study. The ablation study further confirmed that each component of the framework contributes to the overall performance, while the hierarchical optimization strategy combining HBA and NovoGrad improves training stability and model convergence. Sensitivity analyses also indicated that the proposed framework maintains stable performance across different missing-data scenarios and practical hyperparameter settings.

Beyond prediction accuracy, the proposed framework offers practical value for photovoltaic energy management by improving the reliability of historical data reconstruction and supporting more accurate power forecasting in the presence of incomplete measurements. These capabilities can facilitate energy scheduling, grid operation, renewable energy integration, and decision-making in intelligent photovoltaic monitoring systems.

Despite these encouraging results, several limitations remain. Although the framework was validated using multiple datasets with different climatic characteristics, further evaluation on larger and more diverse publicly available datasets would provide a broader assessment of its generalization capability. In addition, comparisons with recently developed transformer-based forecasting models, advanced imputation methods, and foundation models for time-series analysis would further strengthen the experimental validation. Future work will therefore focus on extending the proposed framework to larger-scale real-time photovoltaic systems, investigating lightweight architectures for edge deployment, incorporating uncertainty quantification and explainable artificial intelligence techniques, and evaluating the framework against emerging state-of-the-art forecasting and imputation models under a wider range of operating conditions.

Overall, the proposed optimization-driven imputation and forecasting framework provides an effective and reproducible solution for addressing missing-data challenges in SPV power prediction while offering a promising foundation for reliable renewable energy forecasting in practical smart grid applications.

Data availability statement

The data cannot be made publicly available upon publication because no suitable repository exists for hosting data in this field of study. The data that support the findings of this study are available upon reasonable request from the authors.

Author contributions

Mariappan D  0000-0003-4682-8260

Writing – original draft (equal), Writing – review & editing (equal)

Jain Vinith P R  0000-0003-0197-0001

Visualization (equal), Writing – review & editing (equal)

Garkki B  0009-0007-0563-5266

Methodology (equal), Writing – original draft (equal), Writing – review & editing (equal)

References

- [1] Kothamasu T 2024 Solar power generation in India (Accessed 10 January 2024) (available at: www.mercomindia.com)
- [2] McGlade C 2024 World Energy Outlook 2024 International energy agency report (Accessed 12 January 2024) (available at: www.iea.org)
- [3] Araujo M A D S G, Oliveira F L C and Souza R C 2026 Missing data imputation using climate data reanalysis in solar energy generation time series *Energy Convers. Manage.* **30** 101653
- [4] Zhang W, Luo Y, Zhang Y and Srinivasan D 2020 SolarGAN: multivariate solar data imputation using generative adversarial network *IEEE Trans. Sustain. Energy* **12** 743–6
- [5] Phan Q T, Wu Y K, Phan Q D and Lo H Y 2022 A study on missing data imputation methods for improving hourly solar dataset” in *2022 8th Intern. Conf. on Appl. Syst. Innov. (ICASI)*, pp. 21–24
- [6] De-paz-centeno I, García-ordás M T, García-olalla Ó and Alaiz-moretón H 2023 Imputation of missing measurements in PV production data within constrained environments *Expert Syst. Appl.* **217** 119510–28
- [7] Liu W, Ren C and Xu Y 2022 Missing-data tolerant hybrid learning method for solar power forecasting *IEEE Trans. Sustain. Energy* **13** 1843–52
- [8] Zhang C, Fu X, Qiu D, Badihi H and Gu H 2025 Robust imputation of missing photovoltaic power data using a weather-and context-aware hybrid transformer framework *Renew. Energy* **256** 124576
- [9] Nayak A and Heistrene L 2020 Hybrid machine learning model for forecasting solar power generation in 2020 *Intern. Conf. on Smart Grids and Energy Syst. (SGES)* pp 910–5
- [10] Lee D-S and Son S-Y 2024 PV forecasting model development and impact assessment via imputation of missing pv power data *IEEE Access* **12** 12843–52
- [11] Yan K, Liu J, Zhang J, Yang F, Gao Y and Du Y 2025 Robust photovoltaic forecasting under severe data missingness via multi-domain collaboration and covariate interaction *Appl. Energy* **401** 126771
- [12] Liu W, Ren C and Xu Y 2020 PV generation forecasting with missing input data: a super-resolution perception approach *IEEE Trans. Sustain. Energy* **12** 1493–6
- [13] Başakın E E, Ekmekcioğlu Ö and Özger M 2023 Developing a novel approach for missing data imputation of solar radiation: a hybrid differential evolution algorithm based eXtreme gradient boosting model *Energy Convers. Manage.* **280** 116780–98
- [14] Sundararajan A and Sarwat A I 2020 Evaluation of missing data imputation methods for an enhanced distributed PV generation prediction in *Proc. Future Technol. Conf. (FTC) 2019* vol 1 pp 590–609
- [15] Iao H W, Lao K W and Kang J 2022 Learning-based solar power and load forecasting in DC net-zero energy building with incomplete data in *2022 7th Intern. Conf. on Power and Renewable Energy (ICPRE)* pp 583–9
- [16] Denhard A, Bandyopadhyay S, Habte A and Sengupta M 2021 *Evaluation of Time-series Gap-filling Methods for Solar Irradiance Applications* (No. NREL/TP-5D00-79987) (Nat. Renewable Energy Lab. (NREL), Golden, CO (United States))
- [17] Polasek T and Čadík M 2023 Predicting photovoltaic power production using high-uncertainty weather forecasts *Appl. Energy* **339** 120989–98
- [18] Mohamad Radzi P N L, Akhter M N, Mekhilef S and Mohamed Shah N 2023 Review on the application of photovoltaic forecasting using machiner learning for very short-to long-term forecasting *Sustainability* **15** 2942–60
- [19] Zhang T, Zhang X, Chau T K, Chow Y, Fernando T and Iu H H C 2023 Highly accurate peak and valley prediction short-term net load forecasting approach based on decomposition for power systems with high PV penetration *Appl. Energy* **333** 120641–61
- [20] Díaz-bedoya D, González-rodríguez M, Clairand J M, Serrano-guerrero X and Escrivá-escrivá G 2023 Forecasting univariate solar irradiance using machine learning models: a case study of two Andean cities *Energy Convers. Manage.* **296** 117618
- [21] Tsai W C, Tu C S, Hong C M and Lin W M 2023 A review of state-of-the-art and short-term forecasting models for solar PV power generation *Energies* **16** 5436–54
- [22] Abedi S, Moradi M H and Shirmohammadi R 2023 Real-time photovoltaic energy assessment using a GSM-based smart monitoring system: addressing the impact of climate change on solar energy estimation software *Energy Rep.* **10** 2361–73
- [23] Abou Houran M, Bukhari S M S, Zafar M H, Mansoor M and Chen W 2023 COA-CNN-LSTM: coati optimization algorithm-based hybrid deep learning model for PV/wind power forecasting in smart grid applications *Appl. Energy* **349** 121638–56
- [24] Kumar S S and Sam K N 2023 Multi-hybrid STL-LSTM-SDE-MA model optimized with IWOA for solar PV-power forecasting in *2023 IEEE Intern. Conf. on Power Electron., Smart Grid, and Renewable Energy (PESGRE)* pp 1–6
- [25] Kim E, Akhtar M S and Yang O B 2023 Designing solar power generation output forecasting methods using time series algorithms *Electr. Power Syst. Res.* **216** 109073–91
- [26] Xiong B, Chen Y, Chen D, Fu J and Zhang D 2025 Deep probabilistic solar power forecasting with Transformer and Gaussian process approximation *Appl. Energy* **382** 125294
- [27] Zameer A, Jaffar F, Shahid F, Muneeb M, Khan R and Nasir R 2023 Short-term solar energy forecasting: integrated computational intelligence of LSTMs and GRU *PLoS One* **18** 0285410–21
- [28] Panda S K 2025 Electrical load and solar power forecasting using machine learning techniques *J. King Saud Univ.—Eng. Sci.* **37** 11
- [29] Tejaswi P and Swathika O G 2023 Machine learning algorithms-based solar power forecasting in smart cities *Artif Intell. And Mach. Learn. In Smart City Planning* pp 171–9 (Elsevier)