Day-Ahead Photovoltaic Power Forecasting with Limited Data

Diana Estefanía Chérrez Barragán
School of Electrical Engineering
Universidade Estadual de Campinas
Campinas - SP, Brazil
d227041@dac.unicamp.br

Byron Alejandro Acuña Acurio
School of Electrical Engineering
Universidade Estadual de Campinas
Campinas - SP, Brazil
b209428@dac.unicamp.br

Juan Camilo López
Faculty of Electrical Engineering, Mathematics and Computer Science,– EEMCS
University of Twente,
Enschede, The Netherlands
j.c.lopezamezquita@utwente.nl

Hugo Morais
INESC-ID, Department of Electrical and Computer Engineering,
Instituto Superior Técnico-IST
Universidade de Lisboa
Lisboa, Portugal
hugo.morais@tecnico.ulisboa.pt

Cindy P. Guzman
INESC-ID, Department of Electrical and Computer Engineering,
Instituto Superior Técnico-IST
Universidade de Lisboa
Lisboa, Portugal
cindy.lascano@tecnico.ulisboa.pt

Luiz Carlos Pereira da Silva
School of Electrical Engineering
Universidade Estadual de Campinas
Campinas - SP, Brasil
lui@unicamp.br

***Abstract*—Forecasting algorithms for photovoltaic (PV) power generation play an important role in energy management systems. Nevertheless, the precision of machine learning models is significantly compromised when historical data is limited. This situation is challenging for new plants for which a long history of measurements is not yet available. The unpredictable nature of the weather gives the perception that a competitive forecast requires a substantial amount of data and a very complicated algorithm. However, in this manuscript, it was found that using five historical days for the inverse quantification of uncertainty, can implicitly describe complex non-linear relationships between last five-day records and day-ahead PV power generation. The proposed approach learns the emerging patterns across various seasons throughout the year without relying on exogenous data such as air temperature, wind speed, pressure, cloud cover, and relative humidity. Results using real-world data collected at the microgrid of the University of Campinas (UNICAMP) confirm that our proposed model outperforms previous state-of-the-art deep learning models as Long short-term memory (LSTM), Gated Recurrent Unit (GRU) and traditional Autoregressive Integrated Moving Average (ARIMA) statistical model, using limited data. The proposed approach is flexible and can be easily adapted to other PV power generation systems with limited data.The source code is available at https://github.com/byronacunia/Day-Ahead-Photovoltaic-Power-Forecasting-with-Limited-Data.git**

***Index Terms*—Forecasting, Adaptive learning, Solar, photovoltaic, Statistical approach.**

I. Introduction

Photovoltaic (PV) power generation is uncertain, mainly due to the dynamic and stochastic nature of cloud formation and movement [1]. Nevertheless, accurate forecast models bring more economic and environmental benefits to microgrid owners and utilities [2]. Forecasting methods for PV power generation can be broadly categorized [3] into short-term forecasting (from 1 hour to 7 days ahead), medium-term forecasting (one week to one month ahead), and long-term forecasting (from one month to one year). Based on the aforementioned time horizons, the approach proposed in this manuscript is classified as a short-term forecasting method. A variety of day-ahead short-term forecasting methods for PV power generation have been explored [4], such as: (i) meteorological models, (ii) statistical models, (iii) machine learning models, and (iv) hybrid models. In this case, the proposed approach can be considered a statistical approach. **Meteorological models**, also known as, indirect methods, use numerical weather prediction techniques [5], such as satellite image processing, to forecast the intensity of solar radiation and then convert it into PV power generation. However, there are several disadvantages and challenges associated with meteorological models. For instance, small errors in weather forecasts can lead to significant errors in PV power generation forecasts [6]. In [7] was reported that meteorological models such as numerical weather prediction (NWP) models are limited in terms of accuracy by the non-linearity of the domain equations and the spatial resolution, which generally lies between 16-50 km. Therefore, NWP models are highly dependent on the availability of meteorological records. Other well-known meteorological models use digital cameras or satellites [8] to analyze sky images and capture cloud movements. However, the efficiency of cloud detection and tracking techniques is significantly influenced by camera setup [9].

Statistical models, also known as, direct forecast methods, traditionally use statistical methods [10] such as Autoregressive Moving Average (ARMA) [11], Autoregressive Integrated Moving Average (ARIMA), and Exponential Smoothing to predict PV power generation directly without the need to forecast the solar irradiance firstly. However, traditional statistical methods [12], such as ARIMA, require extensive historical datasets (e.g., five years approx.) for calibration. This is because ARIMA-like models assume that the time series is stationary [13], but the time series of PV power generation is non-stationary.

Machine learning models can be considered as a regression problem [14]. Therefore, it is possible to develop a model for mapping available measurements to day-ahead PV power generation forecast values using supervised learning models, such as support vector regression [15] and neural networks [16]. Grouping several machine learning models is known as an ensemble model. Slightly different from the ensemble model, hybrid models combine models from the aforementioned three categories, such as meteorological models with machine learning and statistical models. However, the precision of machine learning models is significantly compromised when historical data of the time series is insufficient for training the models adequately for each season and/or weather condition [17].

Different from previous works that require an important amount of data, this paper presents a new adaptive learning approach that uses only five days of historical power generation to provide accurate day-ahead production forecasts.

In summary, the main contributions of this paper are as follows:

- A practical approach that provide feasible and accurate day-ahead PV power forecasting method with limited data.
- A recursive average formulation for uncertainty quantification.
- A closed-form expression for input correlational analyses.

II. Proposed Approach

The proposed adaptive learning approach is based on a sun-earth geometric point-of-view. In which the earth's relative position with respect to the sun is represented by the solar declination δ [18]. The approximate value of solar declination can be determined by the approximate equation of Cooper as follows [18]:

$$\delta = 23.45 \sin\left[\frac{360}{365}(284+m)\right] \text{ degrees} \quad (1)$$

where, m represents the day number and is defined as, $m = \{1, 2, ..., 365\}$. PV power generation, weather patterns, and the seasons depend on solar declination, as shown in Fig. 1

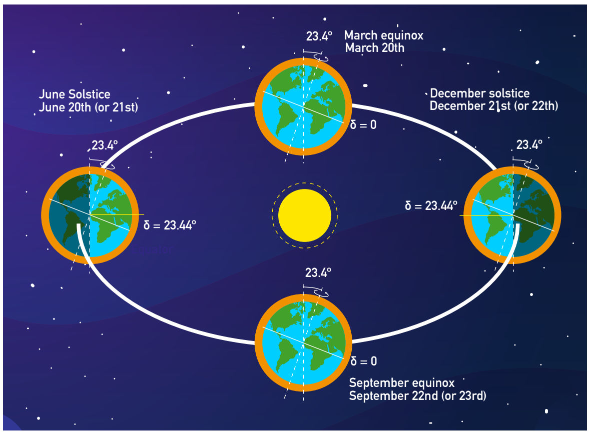

Fig. 1. Solar declination along one solar year

Therefore, solar power generation can be considered a cyclo-stationary process, which can be modeled as follows:

$$X(m) = A \sin\left[\frac{360}{365}(284+m)\right] + N(m) \quad (2)$$

Where, $N(m)$ represents the unwanted random variations caused by the stochastic formation and movement of clouds or dust particles which scatter or disperse solar radiation, and A is the maximum PV power generation value. In this case, the cyclo-stationary process means that the data distribution changes periodically with time. But using the deterministic part of Equation (2) $= A \sin\left[\frac{360}{365}(284+m)\right]$ is possible to find its auto-correlation function, as follows:

$$R_{XX}(\tau) = \mathrm{E}[X(m)X(m+\tau)] \quad (3)$$

$$\begin{aligned} R_{XX}(\tau) = &\frac{1}{365}\int_0^{365} A \sin\left[\frac{360}{365}(284+m)\right] \\ &A \sin\left[\frac{360}{365}(284+m+\tau)\right] dm \end{aligned} \quad (4)$$

Expanding Equation (4) and after performing some minor manipulations, the proposed auto-correlation closed-form expression can be rewritten as,

$$R_{XX}(\tau) = \frac{A^2}{2}\cos\left(\frac{360}{365}\tau\right) \quad (5)$$

In this case, τ is the amount of lag, indicating how many days the signal is shifted when comparing it to its original form. For instance, if $\tau = 0$, it means there is no delay. Hence, the autocorrelation function will return its maximum value, indicating perfect correlation with no shift. As τ increases, the autocorrelation function will drop over different lags of days until zero with a $\tau = 91$ days. Based on the auto-correlation function analysis, it is possible to conclude that a lag(τ) of five days possesses a lower variance. This means that observations within $(0 < \tau \leq 5)$ days from a specific

TABLE I
PV Power Plant Main Components.

Component	Units	Power [kW]	Total Power [kW]
PV module	1248	0.27	336.96
String	48	7.02	336.96
Inverter	5	55.3	276.5

TABLE II
Comparison of Model Performances.

Model	RMSE	MAE
Proposed Approach	0.02701	0.027000
ARIMA [20]	0.08200	0.046510
LSTM [20]	0.07304	0.040300
Xgboost [20]	0.07340	0.039240

point y_t are more useful for predicting y_t than more distant observations ($\tau = 91$). Therefore, the proposed local estimator in this work was defined as follows:

$$X(m+1) = X(m) + \gamma_N(\tau) \tag{6}$$

where, $\gamma_N(\tau)$ is the measure of how much the random variable X changes between two consecutive days, m and $m+1$. To obtain $\gamma_N(\tau)$ term from a collection of random variables $\{X(m), X(m-1), X(m-2), ..., X(m-\tau)\}$. First was obtained a new collection of consecutive days differences$\{\Delta(m), \Delta(m-1), \Delta(m-2), ..., \Delta(m-\tau+1)\}$, using the following equation,

$$\Delta(m) = X(m) - X(m-1) \tag{7}$$

With the aforementioned collection of differences is possible to compute $\gamma_N(\tau)$ based on past trends, using our novel proposed recursive average uncertainty quantification model, as follows:

1) $\gamma_N(k=0) = \Delta(m-\tau+1)$
2) $\gamma_N(k+1) = \frac{1}{2}(\Delta(m-\tau+1+k)+\gamma_N(k))\forall 0 \leq k < \tau$

III. Materials and Methods

This section presents the validation and demonstration of the proposed approach explained in Section II. The discussions are based on the real-life data collected at the UNICAMP - microgrid [19]. The technical information and main components of this case study's PV power plant are presented in Table I.

A. Database: UNICAMP-dataset

This study employs data collected at the output of the inverter of the PV power generation plant of UNICAMP microgrid. The dates range from 2019-04-05/ 00:00:00 to 2021-07-30/ 23:45:00, with a sampling interval of 15 minutes. The 819 days were used to perform exhaustive assessments of our proposed approach in different scenarios.

B. Experiments

The 814 different real-world scenarios to test the proposed approach were generated using the data detailed in Section III-A. Each test scenarios has $N = 96$ samples. In all test scenarios, the number of **observed days** was five days $= m = 5$.

IV. Results and Discussion

The proposed approach for day-ahead PV power generation forecasting, presented in Section II was validated on CampusGrid-60 microgrid at University of Campinas (UNICAMP).

A. Quality

To evaluate the quality of the proposed approach, three different evaluation metrics were employed: mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE). MAPE represents the relative percentage error between the prediction and the actual value. MAE is the average of the absolute difference between the estimation and the actual value of PV power generation. It aims at measuring the average magnitude of errors of the proposed method. RMSE indicates the deviation of the estimation value and the actual value, and thus represents the quality of estimation. The results were summarized in Fig. 2.

Uncertainty in Root Mean Square Error (RMSE) is caused by weather conditions, but also the quality of historical data. However, it is important to note that our proposed approach has a competitive RMSE without exogenous variables. In this case, all RMSE values are below 4%. In seasons with more predictable weather patterns and clear skies, the MAE is lower compared to rainy seasons with highly variable or unpredictable weather. However, on average the error is stable and competitive with state-of-the-art approaches.

Our proposed approach has a MAPE that ranges from 1 to 10%. However in all cases on average the proposed approach shows a MAPE lower of 2% The non-linearities are effectively captured by our proposed approach, as demonstrated in Figures 3 and 4.

B. Computational Efficiency

All tests were performed using a workstation with an AMD Ryzen 7 PRO 5850U with Radeon Graphics 1.90 GHz processor and 16 GB RAM. In this case, the average elapsed time to obtain a forecast of PV power generation forecasting was 3×10^{-6} seconds.

C. Comparison with Prior Art Works

Table II compares our proposed approach with previous works, a hybrid deep learning approach (LSTM model), a machine learning model (Xgboost), and a widely used statistical model (ARIMA).

V. Conclusions

Numerical results in Table II show that our proposed approach is more adaptable, resilient, and accurate than the existing approaches using limited data. Our proposed day-ahead PV power forecasting method is

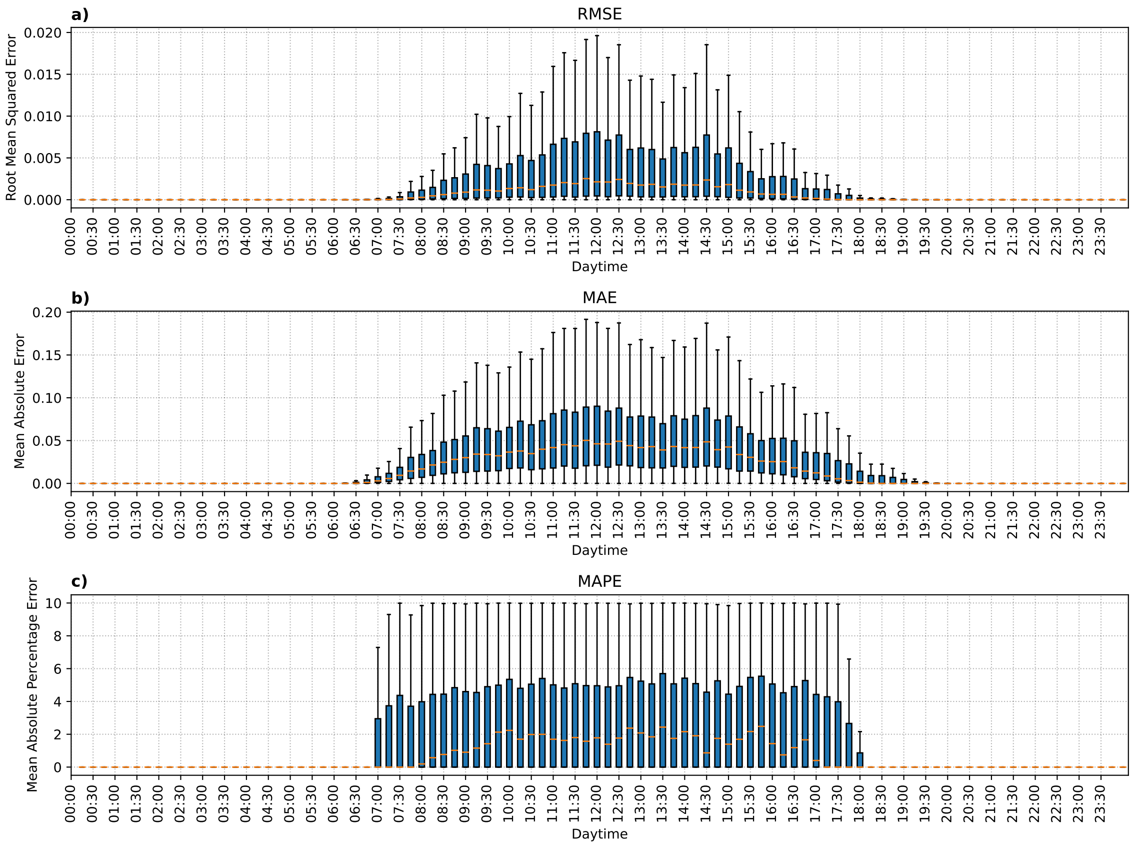

Fig. 2. a) Root mean square error (RMSE), b) Mean absolute error (MAE), and c) Mean absolute percentage error (MAPE).

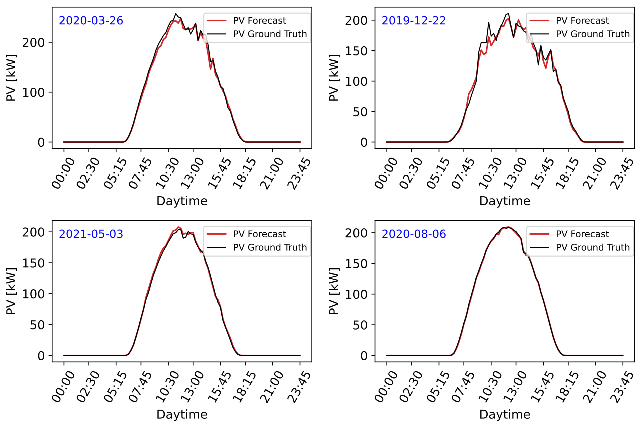

Fig. 3. The proposed approach shows the capacity to adapt autonomously to varying seasonal conditions throughout the year.

a valuable tool for resource-constrained environments such as remote microgrids with new PV power plants where data are sparse. In our research, we systematically evaluated our model over the entire year in different weather conditions. The results reveal the capacity of our model to adapt autonomously to varying seasonal conditions throughout the year. In addition, our proposed model does not require measurement conversion and can be easily adapted to other microgrids with different PV power configurations.

Acknowledgment

This work has been supported by the following Brazilian Research Agencies: FAPESP, CAPES, CNPq, INEP, FINEP, etc. The authors are funded by the grant #2022/16881-5, #2020/03069-5, #2021/11380-5, and #2016/08645-9, Centro Paulista de Estudos da Transição Energética (CPTEn), São Paulo Research Foundation (FAPESP). This work was also developed under the Electricity Sector Research and Development Program PD-00063-3058/2019 - PA3058: "MERGE - Microgrids for Efficient, Reliable and Greener Energy", regulated by the National Electricity Agency (ANEEL in Portuguese), in partnership with CPFL Energia (Local Electricity Distributor). This work has also been developed under the activities of the U2DEMO project, funded by the European Union's Horizon Innovation Actions under grant agreement no. 101160684. Views and opinions expressed in this document are those of the authors only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them. H.M, C.G. and D.C. was also supported by national funds through FCT, Fundação para a Ciên-

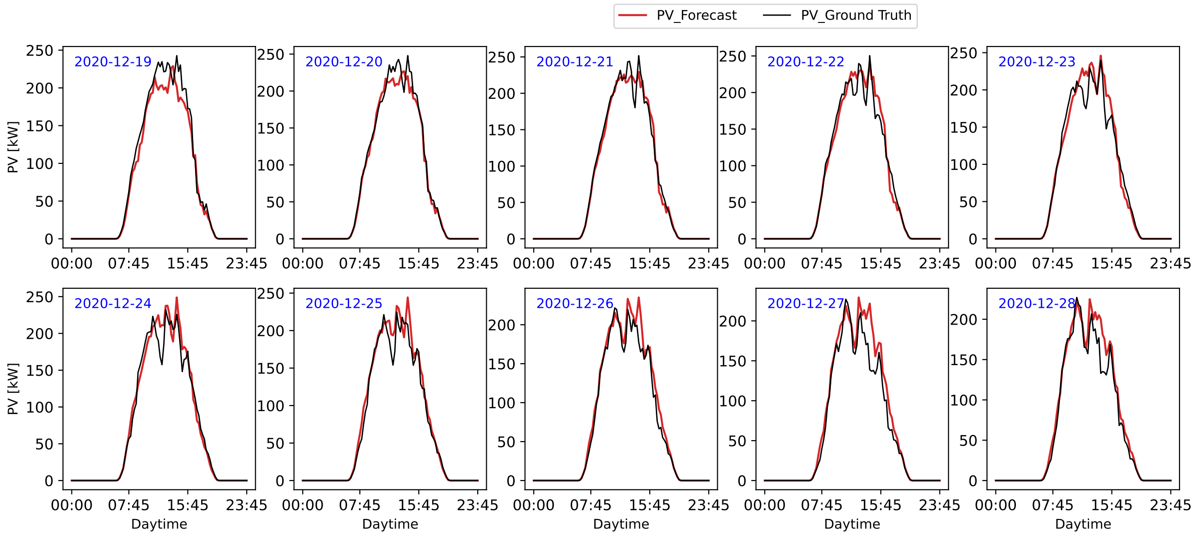

Fig. 4. Moving from sunny days to cloudy days breaks the continuity and production patterns. This transition alters the daily mean and variance of PV power generation. However, the proposed forecasting approach is designed to adapt effectively to these changes in weather conditions, maintaining its accuracy and reliability.

cia e a Tecnologia, under project UIDB/50021/2020 (DOI:10.54499/UIDB/50021/2020).

References

[1] Y. Kamarianakis, Y. Pantazis, E. Kalligiannaki, T. D. Katsaounis, K. Kotsovos, I. Gereige, M. Abdullah, A. Jamal, and A. Tzavaras, "Day-ahead forecasting of solar irradiance & pv power output through statistical machine learning methods," in *2022 Saudi Arabia Smart Grid (SASG)*. IEEE, 2022, pp. 1–5.

[2] J. Antonanzas, D. Pozo-Vázquez, L. Fernandez-Jimenez, and F. Martinez-de Pison, "The value of day-ahead forecasting for photovoltaics in the spanish electricity market," *Solar Energy*, vol. 158, pp. 140–146, 2017.

[3] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. Idris, W. Van Deventer, B. Horan, and A. Stojcevski, "Forecasting of photovoltaic power generation and model optimization: A review," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 912–928, 2018.

[4] P. Singla, M. Duhan, and S. Saroha, "A comprehensive review and analysis of solar forecasting techniques," *Frontiers in Energy*, pp. 1–37, 2021.

[5] P. Bauer, A. Thorpe, and G. Brunet, "The quiet revolution of numerical weather prediction," *Nature*, vol. 525, no. 7567, pp. 47–55, 2015.

[6] T. Polasek and M. Čadík, "Predicting photovoltaic power production using high-uncertainty weather forecasts," *Applied Energy*, vol. 339, p. 120989, 2023.

[7] C. Cornaro, M. Pierro, and F. Bucci, "Master optimization process based on neural networks ensemble for 24-h solar irradiance forecast," *Solar Energy*, vol. 111, pp. 297–312, 2015.

[8] A. Tuohy, J. Zack, S. E. Haupt, J. Sharp, M. Ahlstrom, S. Dise, E. Grimit, C. Mohrlen, M. Lange, M. G. Casado *et al.*, "Solar forecasting: methods, challenges, and performance," *IEEE Power and Energy Magazine*, vol. 13, no. 6, pp. 50–59, 2015.

[9] H. Yang, B. Kurtz, D. Nguyen, B. Urquhart, C. W. Chow, M. Ghonima, and J. Kleissl, "Solar irradiance forecasting using a ground-based sky imager developed at uc san diego," *Solar energy*, vol. 103, pp. 502–524, 2014.

[10] D. C. Montgomery, C. L. Jennings, and M. Kulahci, *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.

[11] E. B. Ssekulima, M. B. Anwar, A. Al Hinai, and M. S. El Moursi, "Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: a review," *IET Renewable Power Generation*, vol. 10, no. 7, pp. 885–989, 2016.

[12] A. K. Chattopadhyay and T. Chattopadhyay, *Statistical methods for astronomical data analysis*. Springer, 2014, vol. 3.

[13] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 65–76, 2013.

[14] C. Molnar, G. Casalicchio, and B. Bischl, "Interpretable machine learning–a brief history, state-of-the-art and challenges," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2020, pp. 417–431.

[15] S. Qijun, L. Fen, Q. Jialin, Z. Jinbin, and C. Zhenghong, "Photovoltaic power prediction based on principal component analysis and support vector machine," in *2016 IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia)*. IEEE, 2016, pp. 815–820.

[16] F. Almonacid, P. Pérez-Higueras, E. F. Fernández, and L. Hontoria, "A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a pv generator," *Energy conversion and Management*, vol. 85, pp. 389–398, 2014.

[17] K. Wang, X. Qi, and H. Liu, "A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network," *Applied Energy*, vol. 251, p. 113315, 2019.

[18] J. R. Brownson, *Solar energy conversion systems*. Academic Press, 2013.

[19] R. Quadros, J. L. Jucá, J. G. I. Cypriano, R. P. B. D. Silva, L. C. P. D. Silva, and R. G. Bento, "Implementation of microgrid on the university campus of UNICAMP - Brazil: Case study," *Journal of Electronics and Advanced Electrical Engineering*, vol. 1, no. 2, pp. 21–25, 2021.

[20] J. C. Cortez, L. Z. Terada, B. V. B. Bandeira, J. Soares, Z. Vale, and M. J. Rider, "Comparative analysis of arima, lstm, and xgboost for very short-term photovoltaic forecasting," in *2023 15th Seminar on Power Electronics and Control (SEPOC)*. IEEE, 2023, pp. 1–6.