Performance of Regression Techniques in Predicting Alternating Current Power (ACP) in Solar Plants: Case Study at IFPE Pesqueira

[1]**Ricardo Ataíde de Lima,** [2]**Manoel Henrique de Oliveira Pedrosa Filho**
[1]Pernambuco University (UPE - POLI), Recife (Brazil)
[2]Federal Institute of Education, Science and Technology (IFPE Campus Pesqueira), Pesqueira (Brazil)

Abstract

This study evaluates machine learning techniques for predicting Alternating Current Power (ACP) output in a photovoltaic plant at IFPE Pesqueira campus, Brazil. Models including Multiple Regression, Support Vector Regression (SVR), Decision Tree, Random Forest, and XGBoost were compared using solar irradiance, ambient temperature, and Sandia cell module temperature as inputs. All models achieved high accuracy ($R^2 > 0.8$, $RMSE < 0.17$), with SVR demonstrating the best performance ($R^2 = 0.8517$), albeit with a higher computational cost. Faster methods such as Decision Tree (0.0022 s forecast time) proved viable for real-time applications. This work offers practical insights for model selection based on the trade-off between accuracy and efficiency in operational scenarios.
*Keywords:*Photovoltaic power forecasting, Machine Learning applications, Regression models comparison.

1. Introduction

In the design of photovoltaic plants, energy generation is directly influenced by meteorological variables and equipment operating conditions (Sharma et al., 2021). Key factors such as solar irradiance, ambient temperature, and photovoltaic module temperature models significantly affect system performance (Sun et al., 2022). However, the intermittent nature of solar power poses considerable operational challenges, particularly for stable integration into the electrical grid. To address these issues, accurate forecasting methods are essential for optimizing energy dispatch, storage, and demand management (Kim et al., 2023).
In this context, Machine Learning (ML) techniques have become established tools for modeling the relationships between environmental variables and photovoltaic energy generation (Subramanian et al., 2023). Unlike traditional physical models that rely on deterministic equations, ML algorithms can learn nonlinear patterns from historical data, enhancing the accuracy of short- and medium-term forecasts. Promising approaches include Multiple Regression (Kim et al., 2021), Support Vector Regression (SVR) (Kasireddy et al., 2022), Decision Trees (Shorabeh et al., 2022), Random Forests (Wang et al., 2023), and gradient boosting methods such as XGBoost, which effectively capture temporal dependencies in historical series (Phan et al., 2023).
In addition to operational optimization, forecasting photovoltaic generation via machine learning allows monitoring plant performance over time, identifying discrepancies between the energy generated and theoretical expectations (e.g., losses due to module degradation, dirt, failures, or unforeseen inefficiencies). This analysis is essential to validate manufacturers' warranties, which typically specify annual efficiency losses (e.g., 0.5% to 1% per year) (Şevik and Aktaş, 2022), and to support preventive maintenance actions or contractual compensations. Machine learning models, by incorporating historical production data and environmental variables, can isolate the impact of equipment aging from climate fluctuations, offering a more accurate assessment of the actual degradation of the system.
This study compares ML techniques for predicting energy generation in a photovoltaic plant installed at the Federal Institute of Education, Science and Technology (IFPE), Pesqueira Campus. The following algorithms were evaluated: Multiple Regression, Support Vector Regression (SVR), Decision Tree, Random Forest and XGBoost, modeling the relationship between Alternating Current Power (ACP) and environmental variables — solar radiation, ambient temperature and module temperature (according to the Sandia Cell mathematical model). The results indicate that all models achieved high accuracy, with Pearson's coefficient ($R^2 > 0.8$) and a maximum Mean Absolute Error (MAE) of 0.1670. The main difference between the techniques resided in

10.18086/swc.2025.03.54 Available at http://proceedings.ises.org

the Training and Prediction Time. The objective of this work is to develop robust predictive models to evaluate and optimize the efficiency of the photovoltaic plant, supporting more assertive energy planning.

2. Metodology

The main objective of this study is to predict ACP, which constitutes the target variable. Data were obtained from the Pernambuco Solarimetric Network (Pedrosa Filho, 2022), an online repository developed by researchers at IFPE Pesqueira Campus, aiming to centralize solar radiation data sources throughout the state of Pernambuco, facilitating and promoting research in this field.

Using a script developed in SQL and Python, the data were extracted and preprocessed. Although the accessed database contains several parameters, preliminary correlation and feature importance analyses identified ambient temperature (°C) and solar irradiance (W/m^2) as the most relevant variables for predicting ACP (W). Hourly data from 2022, comprising measurements collected at 5-minute intervals, were used for the analysis, resulting in 45,871 valid records after processing. A hold-out validation strategy was adopted, allocating 70% of the data for training and the remaining 30% for testing.

The PV module temperature was estimated using the library pvlib python (Holmgren and The pvlib Python Community, 2024), which provides validated implementations of several empirical models. A comparative analysis was conducted between the Sandia (Lynch et al., 2023), Sandia Cell (Anderson et al., 2021), Jacques (Jacques et al., 2013), and PVsyst Cell (Khan et al., 2024) models, using irradiance and ambient temperature data extracted from the solarimetric network. The performance of the temperature models was quantified by comparing their estimates with measured module data, using the Coefficient of Determination (R^2) and Mean Absolute Error (MAE) as metrics. Given these metrics, the Sandia Cell model was selected for the study because it presented the highest correlation with the experimental data, the smallest errors, and the greatest subsequent influence on the PCA prediction.

Using a Python script, regression techniques (Multiple Regression, SVR, Decision Tree, Random Forest, and XGBoost) were evaluated to predict ACP from temperature data (ambient and Sandia Cell). The implementation was performed through a Python script, using libraries such as Scikit-learn (Pedregosa et al., 2011) and XGBoost (Chen and Guestrin, 2016). The input data were normalized using StandardScaler to ensure that all features were on a similar scale before training. The models' hyperparameters were optimized through 5-fold cross-validation on the training set to prevent overfitting.

The evaluation of the machine learning models was based on a double-criteria analysis, considering both predictive accuracy and computational efficiency. Predictive accuracy was quantified using the R^2 and MAE, metrics calculated on the test set to ensure an unbiased assessment of the models' generalization ability. Computational efficiency was measured by recording the total training time and the average prediction time per sample for each model. This analysis is crucial for determining the feasibility of deployment in real-time prediction systems.

All experiments were conducted on a standardized hardware platform, consisting of a computer equipped with an Intel i9-12900F processor, an NVIDIA RTX 4070 graphics card, and 36 GB of RAM, ensuring consistent and reproducible performance measurements.

3. Results

Figure 1(a)-(d) present the temporal comparison between actual and predicted ACP values for each regression method, revealing consistent patterns across the evaluation period. Figure 2(a)-(b) detail a representative 24-hour timeframe, highlighting the models' performance during high-variability periods. Quantitative results in Table 1 summarize key metrics including R^2, MAE, and computational times. Although all models achieved satisfactory performance ($R^2 > 0.83$, $MAE < 0.17$) for ACP prediction, the optimal method selection depends on the specific application requirements. SVR achieved the highest R^2 (0.8517) and lowest MAE (0.1503), but its substantial computational overhead (26.7 s training, 11.0 s prediction)—approximately three orders of magnitude higher than faster alternatives—limits its practical deployment. For real-time applications or large-scale implementations, Linear Regression (0.026 s training) and Decision Tree (0.0022 s prediction) offer compelling alternatives with comparable accuracy ($R^2 \sim 0.83 - 0.84$). Random Forest and XGBoost present a balanced compromise, delivering robust performance with moderate computational demands, making them suitable for scenarios requiring both generalization capability and reasonable processing times. The choice ultimately hinges on whether the operational context prioritizes prediction accuracy or computational efficiency.

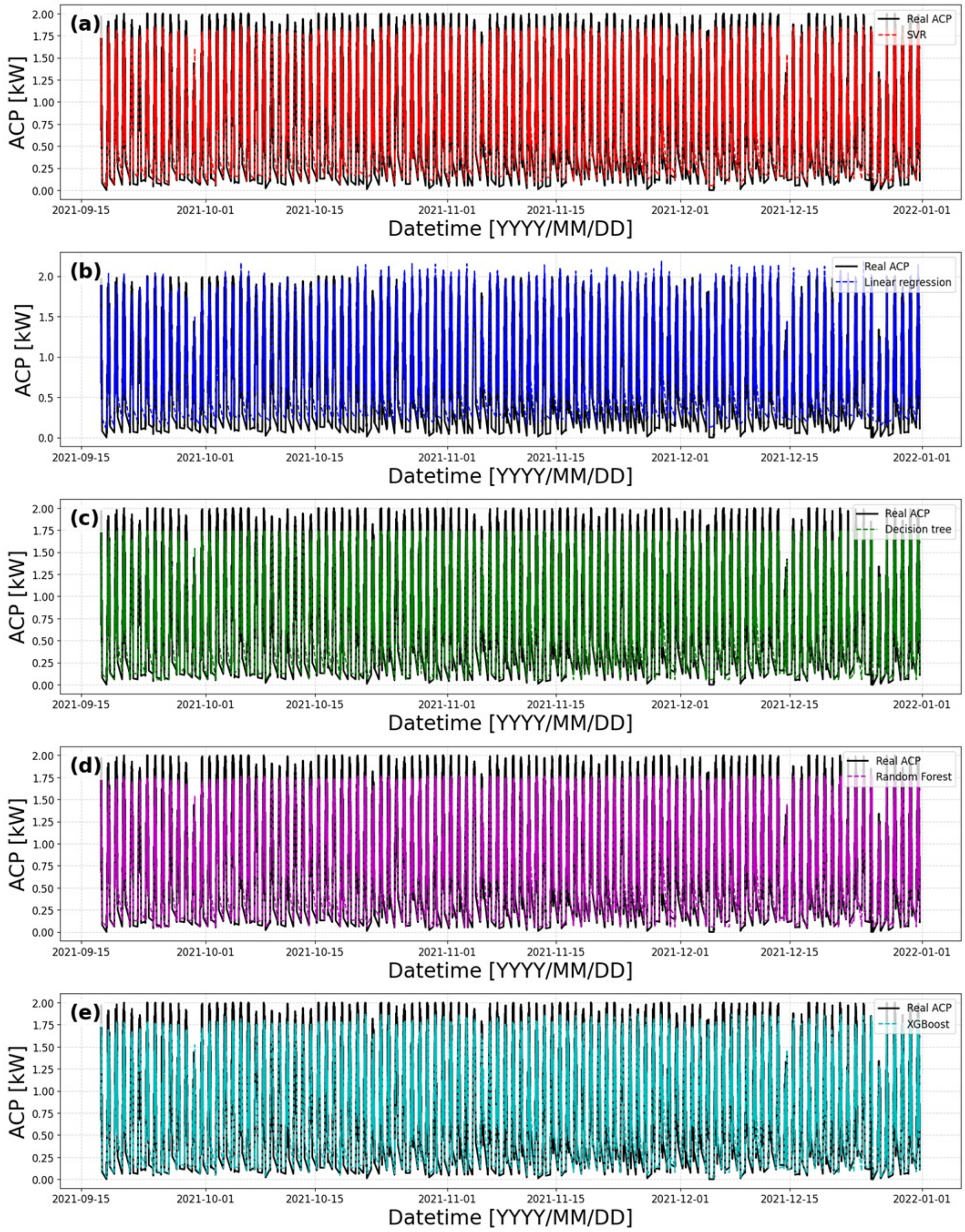

Figure 1: Comparison between the actual PCA values and the predictions of the (a) SVR, Linear Regression (b), Decision Tree (c), Random Forest (d) and XGBoost models (e).

4. Discussions and Conclusion

The results presented in this manuscript reinforce the viability of using regression-based approaches for the operational forecasting of photovoltaic systems. It was possible to verify the accuracy of the models for forecasting the Alternating Current Power (ACP) of the IFPE Campus Pesqueira photovoltaic plant. A significant improvement in statistical parameters is observed as the time interval decreases, especially in the transition from 1 month to 1 week. This behavior is explained by the greater temporal autocorrelation and stability of operational conditions in hourly/daily windows, where the system state strongly depends on the immediately preceding state. Among the models evaluated, the SVR stood out as the most robust and accurate, achieving a Pearson correlation coefficient of 0.9667 and an R^2 of 0.9330 in the weekly forecast.

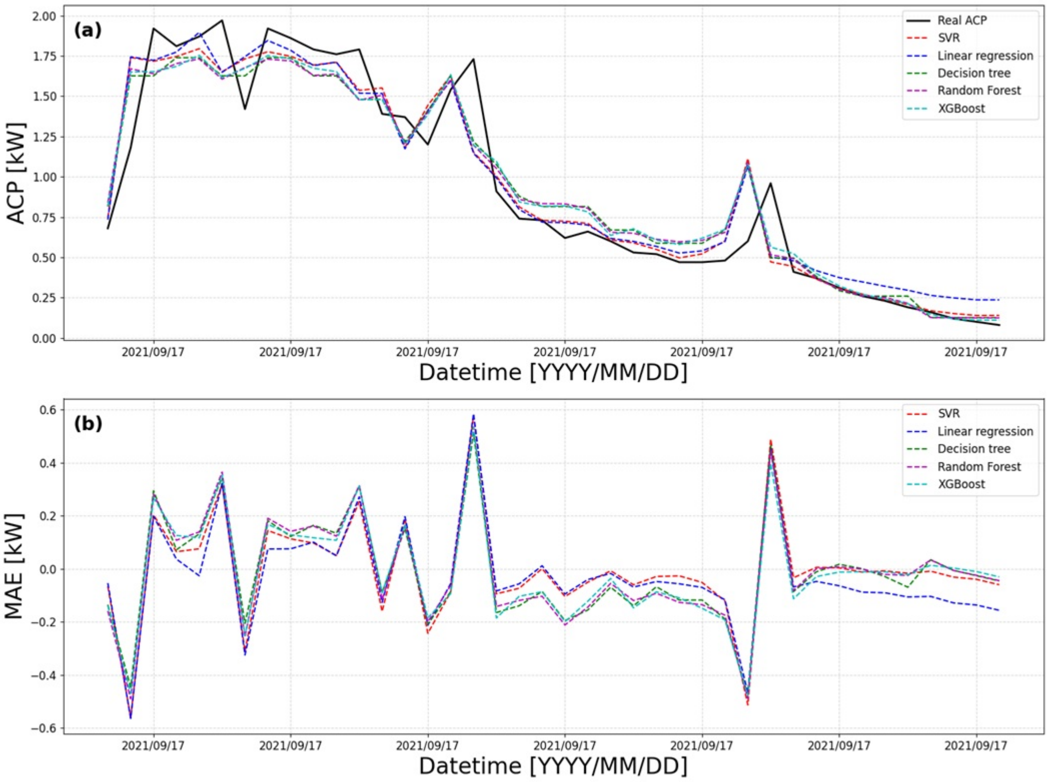

Figure 2: A 24-hour period with calculation of (a) R^2 and (h) the MAE.

Table 1: Performance comparison of ACP prediction models.

Model	R²	MAE (kW)	Training time (s)	Prediction time (s)
SVR	0.8517	0.1503	26.5088	10.4124
Linear Regression	0.8329	0.1670	0.0286	0.0021
Decision Tree	0.8423	0.1649	0.0728	0.0240
Random Forest	0.8436	0.1641	5.8301	0.0782
XGBoost	0.8398	0.1663	0.2465	0.0606

The results reinforce the viability of using data-driven approaches for the operational forecasting of photovoltaic systems, especially in short horizons. Although all models achieved satisfactory performance, the selection of the appropriate method depends on the specific requirements of the application. As mentioned earlier, SVR achieved the best performance, but its substantial computational overhead is approximately three orders of magnitude greater than faster alternatives. For real-time applications or large-scale implementations, Linear Regression and Decision Tree offer attractive alternatives with comparable accuracy. Random Forest and XG-Boost represent a balance, offering robust performance with moderate computational demands, making them suitable for scenarios requiring both generalization capability and reasonable processing times.
Future studies could explore the inclusion of other meteorological variables, such as cloud cover and humidity, as well as the incorporation of physical constraints into the models. Another alternative would be the application of hybrid modeling and forecasting methods. These possibilities allow avoiding non-physical forecasts in low irradiance conditions and improving robustness in rapidly varying atmospheric scenarios.

5. References

Anderson, K., Kemnitz, J., Boyd, M., 2021. Evaluating cell temperature models and the effect of wind speed in pv system capacity testing, in: 2021 IEEE 48th Photovoltaic Specialists Conference (PVSC), IEEE. pp. 1663–1669.

Chen, T., Guestrin, C., 2016. Xgboost documentation. `https://xgboost.readthedocs.io/en/stable/`. Accessed: 2023-10-25.

Holmgren, W.F., The pvlib Python Community, 2024. pvlib python: a python package for modeling solar energy systems. URL: `https://pvlib-python.readthedocs.io/en/stable/`. version 0.10.3.

Jacques, S., Caldeira, A., Ren, Z., Schellmanns, A., Batut, N., et al., 2013. Impact of the cell temperature on the energy efficiency of a single glass pv module: thermal modeling in steady-state and validation by experimental data. Renewable Energy and Power Quality Journal 11, 1–4.

Kasireddy, I., Padmini, K., Ramarao, R., Seshagiri, B., Rani, B.V.N., 2022. Solar energy prediction using machine learning with support vector regression algorithm, in: International Conference on Cognitive Computing and Cyber Physical Systems, Springer. pp. 17–25.

Khan, M.M., Ahmad, S., Raza, A., Haroon, S., Hassan, R.G., Shafi, M.A., 2024. Performance analysis of pvsyst based grid connected photovoltaic systems in pakistan compared to saarc countries. Mehran University Research Journal Of Engineering & Technology 43, 112–122.

Kim, E., Akhtar, M.S., Yang, O.B., 2023. Designing solar power generation output forecasting methods using time series algorithms. Electric Power Systems Research 216, 109073.

Kim, Y.S., Joo, H.Y., Kim, J.W., Jeong, S.Y., Moon, J.H., 2021. Use of a big data analysis in regression of solar power generation on meteorological variables for a korean solar power plant. Applied Sciences 11, 1776.

Lynch, K.P., Grasser, T., Spillers, R., Downing, C., Daniel, K., Jans, E., Kearney, S., Morreale, B., Wagnild, R., Wagner, J., 2023. Design and characterization of the sandia free-piston reflected shock tunnel. Shock Waves 33, 299–314.

Pedregosa, F., et al., 2011. Scikit-learn: Machine learning in python. `https://scikit-learn.org/stable/`. Accessed: 2023-10-25.

Pedrosa Filho, M.H.O., 2022. Rede Solarimétrica de Pernambuco, in: IX Congresso Brasileiro de Energia Solar., CBENS, Florianópolis.

Phan, Q.T., Wu, Y.K., Phan, Q.D., 2023. Application of a new transformer-based model and xgboost to improve one-day-ahead solar power forecasts, in: 2023 IEEE/IAS 59th Industrial and Commercial Power Systems Technical Conference (I&CPS), IEEE. pp. 1–7.

Şevik, S., Aktaş, A., 2022. Performance enhancing and improvement studies in a 600 kw solar photovoltaic (pv) power plant; manual and natural cleaning, rainwater harvesting and the snow load removal on the pv arrays. Renewable Energy 181, 490–503.

Sharma, P., Goyal, P., et al., 2021. Analysing the effects of solar insolation and temperature on pv cell characteristics. Materials Today: Proceedings 45, 5539–5543.

Shorabeh, S.N., Samany, N.N., Minaei, F., Firozjaei, H.K., Homaee, M., Boloorani, A.D., 2022. A decision model based on decision tree and particle swarm optimization algorithms to identify optimal locations for solar power plants construction in iran. Renewable Energy 187, 56–67.

Subramanian, E., Karthik, M.M., Krishna, G.P., Prasath, D.V., Kumar, V.S., 2023. Solar power prediction using machine learning. arXiv preprint arXiv:2303.07875 .

Sun, C., Zou, Y., Qin, C., Zhang, B., Wu, X., 2022. Temperature effect of photovoltaic cells: a review. Advanced Composites and Hybrid Materials 5, 2675–2699.

Wang, X., Xiao, X., Zhang, X., Ye, H., Dong, J., He, Q., Wang, X., Liu, J., Li, B., Wu, J., 2023. Characterization and mapping of photovoltaic solar power plants by landsat imagery and random forest: A case study in gansu province, china. Journal of Cleaner Production 417, 138015.