Proceedings of the 11th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'25)
Paris, France - August, 2025
Paper No. EEE 134
DOI: 10.11159/eee25.134

Effectiveness of Time Series Models in Consumption Forecasting for Photovoltaic Energy with Limited Historical Data: SARIMA and Prophet

Henrique Teixeira Silva[1], Elcio Franklin de Arruda[1], Luiz Felipe Pugliese[1], Rodrigo Aparecido da Silva Braga[2]
[1]Federal University of Itajuba – *campus* Itabira
Rua Irmã Ivone Drumond, 200 - Distrito Industrial II, Itabira, Brazil
d2021007160@unifei.edu.br; elcio.arruda@unifei.edu.br; puglies@unifei.edu.br
[2]Federal University of Lavras
R. Antonio Carlos Pinheiro Alcântara, n° 855, São Sebastião do Paraíso, Brazil
rodrigobraga@ufla.br

Abstract - This study compares the effectiveness of SARIMA and Prophet models in forecasting business electricity consumption in Brazil, focusing on scenarios with limited historical data. The research analyses data from various companies that are photovoltaic energy consumers, applying a methodology that includes data treatment, characteristic analysis, and parameter optimization. The results indicate that Prophet generally outperforms SARIMA in data with strong seasonality, while SARIMA may be more effective in less defined consumption patterns. The study concludes that model choice should consider the specific characteristics of each company's data, offering valuable insights for photovoltaic energy companies to optimize their consumption forecasts.

Keywords**:** Photovoltaic energy. Consumption forecasting. SARIMA. Prophet. Time series.

1. Introduction

Energy consumption forecasting applies historical data to predict future demand, with Energy Demand Forecasts (EDFs) being crucial for optimal resource allocation [1]. Brazil's photovoltaic sector has grown exponentially, exceeding 30 GW of installed capacity by 2023, representing over 10% of the country's electric matrix [2]. However, consumption forecasting faces challenges including varied customer profiles and limited historical data [3], along with grid modernization needs, solar intermittency, and variable energy demand.

Companies' consumption profiles vary significantly (Figure 1), necessitating personalized forecasts. Limited historical data further complicates forecasting, making optimization of conventional models like SARIMA and Prophet particularly relevant.

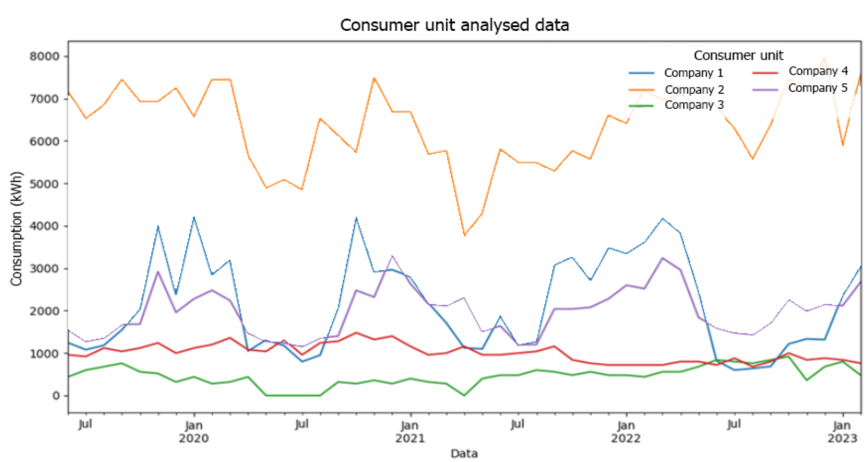

Fig. 1. Monthly energy consumption between the analyzed companies

This research compares these models in forecasting electrical energy consumption of consumer units with limited historical data. We analyze performance across different consumption profiles, evaluate seasonality and stationarity impacts, investigate data augmentation effects, and propose selection guidelines based on company-specific factors. The study addresses a literature gap by focusing on model optimization in data-scarce contexts, providing insights for photovoltaic companies to enhance forecasting strategies, operational efficiency, and strategic planning.

2. Theoretical Framework

This paper explores SARIMA and Prophet models. SARIMA extends ARIMA by incorporating seasonal components, decomposing time series into trend, seasonality, and error components. It's represented as SARIMA(p,d,q)(P,D,Q)[S], where p,d,q are non-seasonal parameters, P,D,Q are seasonal counterparts, and S represents periods per seasonal cycle [4]. SARIMA is widely used for medium and long-term energy forecasting [1].

Developed by Facebook, Prophet is designed for business time series forecasting [5]. It uses a decomposable model with three components: trend function g(t), seasonal changes s(t), holiday effects h(t), plus an error term εt. Prophet offers advantages including multiple seasonality flexibility, handling of missing data, and intuitive parameters [5].

Limited data increases overfitting risk, as models may excessively fit training data, compromising generalization [6]. This risk increases in photovoltaic forecasting with scarce data, requiring careful validation [7]. Data augmentation can help but must preserve fundamental characteristics like seasonality [8]. For energy forecasts, duplicating seasonal cycles or generating synthetic data may be appropriate, with careful performance evaluation [9].

For the performance evaluation, common performance metrics will be used, which includes RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and MAPE (Mean Absolute Percentage Error),

3. Description of Developed Activities

A procedural flow was created to analyse individual company, consisting of several key steps (Figure 2). Data acquisition and treatment involved importing and formatting consumption data into appropriate dataframes, excluding Consumer Units (CUs) with insufficient historical data. Data characteristics analysis, by extracting features like seasonality (using seasonal_decompose from the StatsModels library) and stationarity (using the Augmented Dickey-Fuller (ADF) test, which performs a unit root test for stationarity [10]). An optimization function was also used to help choose other parameters and corroborate those already found.

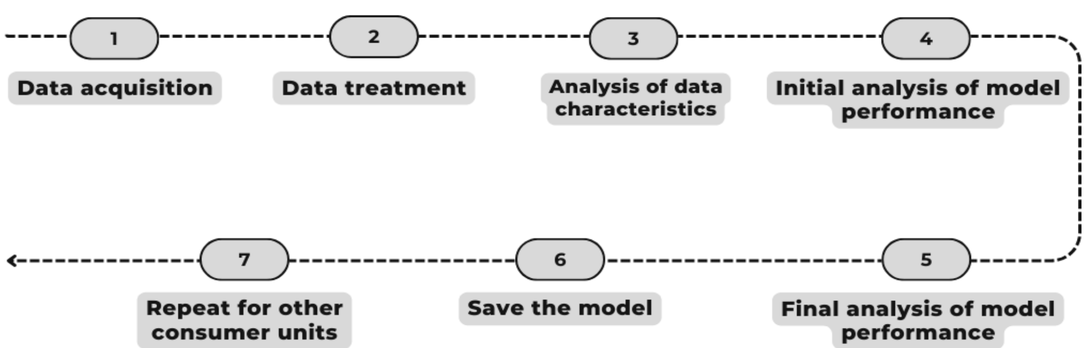

Fig. 2. Model creation flow chart.

Model performance was initially tested using training/test split (90%/10%), while final models were created using all available data and saved for future predictions and integration into other systems.

4. Results and Analysis

After performing the model creation steps, the test was performed for each company until a satisfactory error was obtained. It's worth noting that not always a satisfactory error is a smaller error, due to the risk of overfitting. Due to this problem and data limitation, it was often necessary to compare future values with the entire database and choose parameters that returned higher errors but predictions that better followed the consumption trend.

Based on all analysis, Table 1 and Table 2 summarize the results, with Table 1 presenting the best models obtained for SARIMA and Table 2 detailing the top Prophet models.

To address the issue of limited data, data duplication was employed, followed by model retraining and analysis. However, this approach generally led to a deterioration in error metrics, likely due to the need to prevent overfitting.

Table 1. Error metrics for the models created with SARIMA.

	SARIMA			SARIMA - Augmented		
CU	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE
Company 1	838.51	42.28	692.01	1073.57	54.82	822.93
Company 2	705.83	9.13	589.66	849.05	10.81	661.72
Company 3	140.61	Erro	112.09	188.91	Erro	112.09
Company 4	111.94	10.44	88.95	111.94	11.62	139.63
Company 5	433.01	15.54	351.91	138.72	22.77	115.63

Analysing Prophet's performance, Table 2 shows it excelled with Company 1's seasonal data, outperforming SARIMA significantly. Prophet effectively handled both seasonal and stationary data, maintaining reasonable predictions. However, data augmentation caused some overfitting. Overall, Prophet demonstrated greater versatility for photovoltaic consumption forecasting with limited historical data.

Table 2. Error metrics for the models created with Prophet.

	Prophet			Prophet - Augmented		
CU	RMSE	MAPE (%)	MAE	RMSE	MAPE (%)	MAE
Company 1	170.77	5.58	170.77	486.45	12.66	486.45
Company 2	661.61	8.75	661.61	320.02	4.23	320.02
Company 3	144.57	Erro	144.57	115.83	Erro	115.83
Company 4	257.24	33.84	257.24	79.71	10.48	79.71
Company 5	310.13	11.57	310.13	159.97	7.54	159.97

The study demonstrated that Prophet performed better with strong seasonality data, while SARIMA was more effective with less defined consumption patterns. Further analysis revealed Prophet also effectively handled stationary data without strong trends.

Overfitting risks were significant, especially with limited data. Primary causes include training database noise (sudden trend changes that disappear after one or two months) [6], hypothesis complexity [7], and multiple comparison procedures [11].

5. CONCLUSIONS

Results demonstrated that model performance depends on data characteristics. Prophet excelled with both seasonal and stationary data, while SARIMA performed better with less defined consumption patterns. Data augmentation produced mixed results, highlighting forecasting challenges with limited historical data.

Individual company analysis proved essential due to varied consumption patterns. Both models performed well with stable consumption data, with Prophet showing greater overall reliability.

These findings offer practical value for the photovoltaic sector, enabling companies to optimize resources through model selection based on specific data characteristics. Understanding limitations in data-scarce scenarios helps manage forecast risks and improve operational efficiency.

This study establishes a foundation for enhancing energy consumption forecasting in the photovoltaic sector with limited data, potentially improving operational efficiency and contributing to solar energy advancement in Brazil.

Acknowledgements

I would like to express my deep gratitude to the Federal University of Itajubá (UNIFEI) for providing the environment and resources necessary for conducting this research. I also thank CNPq [129512/2023-0] for the financial support that allowed this work to be carried out.

References

[1] S. Chaturvedi, E. Rajasekar, S. Natarajan and N. McCullen, "A comparative assessment of SARIMA, LSTM RNN and FB Prophet models to forecast total and peak monthly energy demand for India," Energy Policy, vol. 160, p. 112648, 2022.

[2] ABSOLAR, "Retrospectiva ESG 2023: O ano da virada para a energia solar no Brasil," 2024. [Online]. Available: https://www.absolar.org.br/noticia/retrospectiva-esg-2023-o-ano-da-virada-para-a-energia-solar-no-brasil/. [Accessed: Sep. 13, 2024].

[3] S. Wei, C. Li, X. Peng, F. Zeng and X. Lu, "Conventional models and artificial intelligence-based models for energy consumption forecasting: A review," *Journal of Petroleum Science and Engineering*, vol. 181, p. 106187, 2019.

[4] J. L. T. García, E. C. Calderón, G. G. Ávalos, E. R. Heras and A. M. Tshikala, "Forecast of daily output energy of wind turbine using SARIMA and nonlinear autoregressive models," *Advances in Mechanical Engineering*, vol. 11, no. 12, p. 1687814019891159, 2019.

[5] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37-45, 2017.

[6] X. Ying, "An overview of overfitting and its solutions," *Journal of Physics: Conference Series*, vol. 1168, no. 2, p. 022022, 2019.

[7] G. Paris, D. Robilliard, and C. Fonlupt, "Exploring overfitting in genetic programming," in *Lecture Notes in Computer Science*, 2004, pp. 267-277.

[8] B. K. Iwana and S. Uchida, "An empirical survey of data augmentation for time series classification with neural networks," *PLoS One*, vol. 16, no. 7, p. e0254841, 2021.

[9] A. Le Guennec, S. Malinowski, and R. Tavenard, "Data augmentation for time series classification using convolutional neural networks," in *ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data*, 2016.

[10] H. Ramenah, P. Casin, M. M. Ba, M. Benne and C. Tanougast, "Accurate determination of parameters relationship for photovoltaic power output by augmented dickey fuller test and engle granger method," *AIMS Energy*, vol. 6, no. 1, pp. 19-48, 2018.

[11] D. Jensen and P. Cohen, "Multiple comparisons in induction algorithms," *Machine Learning*, vol. 38, no. 3, pp. 309-338, 2000.