Analysis of Wind Energy Generation in Brazil Using a SARIMA-MLP Hybrid Model with Weight Optimization

1st Pedro Natanael Firmino da Silva
Programa de Pós-Graduação em Engenharia da Computação (PPGEC)
Universidade de Pernambuco (UPE)
Garanhuns, Brazil
pnfs@ecomp.poli.br

2nd Morgan Oliveira Melo de Siquiera
Programa de Pós-Graduação em Engenharia da Computação (PPGEC)
Universidade de Pernambuco (UPE)
City, Country
moms@ecomp.poli.br

***Abstract*—This paper presents a hybrid forecasting model combining the Seasonal AutoRegressive Integrated Moving Average (SARIMA) and Multi-Layer Perceptron (MLP) with weight optimization for wind energy generation in Brazil. The model aims to improve predictive accuracy by capturing both linear and nonlinear patterns in the time series data. Our approach involves data preprocessing, model type determination, trend and seasonality tests, and fitting the SARIMA and MLP models. The performance of the proposed model was compared to existing models using RMSE as the evaluation metric, showing significant improvement. The SARIMA-MLP hybrid model demonstrated an 8.16% improvement in predictive accuracy over the SARIMA model, indicating its effectiveness in capturing complex patterns in wind energy data. The optimized weight calculation further fine-tuned the model, minimizing prediction error and maximizing accuracy. These results highlight the potential of the SARIMA-MLP hybrid model as a robust tool for wind energy forecasting in Brazil, providing valuable insights for future planning and management of wind energy resources.**

***Index Terms*—Wind Energy, Time Series Forecasting, SARIMA, MLP, Hybrid Model, Weight Optimization.**

I. Introduction

The increasing demand for electricity in Brazil is a growing concern for both the public and private sectors. In 2014, the National System Operator (ONS) reported a new peak in electricity consumption, reaching 85,708 megawatts (MW). From January to May 2021, electricity consumption rose by 6.9% compared to the same period in 2020, according to the Ministry of Mines and Energy [1].

Brazil's electricity matrix relies heavily on hydroelectric power, which accounts for 62.8% of the installed capacity, followed by thermoelectric plants at 28.2%. Other sources, including wind, small hydroelectric plants, hydroelectric generating stations, and solar, contribute minimally to the overall energy mix [2]. Thermoelectric plants, which use finite and environmentally harmful fossil fuels, present additional challenges.

The National Electric Energy Agency (ANEEL) highlights Brazil's considerable potential for wind power generation, estimated at 143 gigawatts (GW). This potential is particularly notable in coastal regions of the north and northeast, the southern coast of Rio Grande do Sul, the southeast of Paraná, and the São Francisco River valley [3]. Brazil's wind resources, characterized by twice the global average wind volume and low speed oscillation, offer significant predictability for electricity generation.

Time series forecasting plays a vital role in many economic, financial, and scientific applications, where precise predictions can lead to significant improvements in decision-making and operational efficiency. Traditional models, such as the Auto-Regressive Integrated Moving Average (ARIMA), have been widely used due to their robustness in capturing linear relationships and trends in time series data. However, these models often fall short when dealing with non-linear patterns that are increasingly present in complex systems like financial markets, weather forecasting, and energy demand management.

To address these limitations, researchers have turned to machine learning techniques, such as Multi-Layer Perceptrons (MLP), which can model complex non-linear relationships hidden in the data. Despite their flexibility, MLPs alone can sometimes overlook the seasonal and trend components that ARIMA models handle well. This has led to a growing interest in hybrid models that combine the strengths of statistical and machine learning approaches to enhance forecasting accuracy and reliability.

Hybrid models, particularly those integrating SARIMA (Seasonal ARIMA) and MLP, offer a promising approach for capturing both linear and non-linear dynamics in time series

Identify applicable funding agency here. If none, delete this.

data. The SARIMA model extends ARIMA by accounting for seasonality, an essential feature for many practical applications such as electricity load forecasting and inventory management, where patterns repeat over a known cycle. By coupling SARIMA with MLP, the hybrid approach aims to leverage SARIMA's effectiveness in handling seasonal variations and the powerful pattern recognition capabilities of MLP to improve overall forecast performance.

This paper builds on this foundation by exploring specific hybrid combinations of SARIMA-MLP to improve wind energy forecasting in Brazil. Each configuration is designed to test how different sequences of these models affect forecasting accuracy, providing insights into the optimal structuring of hybrid models for various time series forecasting challenges. Furthermore, this work introduces an innovative approach by optimizing a weight on the MLP residuals, aiming to further enhance the model's accuracy. The goal is to establish a robust framework that can be effectively applied across different domains, offering substantial improvements over existing methodologies.

II. Related Work

Numerous studies have employed statistical techniques to analyze historical time series data on wind energy across different regions. Notable among these, Alencar (2018) [3] developed a hybrid wind speed prediction model combining SARIMA and neural networks, while Fernandes (2018) [5] focused on short-term wind power forecasting using spectral analysis and time series decomposition. Silva (2018) [6] compared three dimensionality reduction methods for analyzing extreme wind values, and Silva (2017) [7] performed statistical modeling for short-term wind forecasting in Northeast Brazil. Recent literature also highlights the growing use of hybrid models to enhance forecasting accuracy. For instance, Hajirahimi and Khashei (2016) [8] demonstrated the superiority of hybrid ARIMA and ANN models in financial forecasting. Liu et al. (2019) [9] introduced a model integrating EMD, MAF, LSSVR, and QES for forecasting electrical energy consumption, showing its effectiveness over single models. Hu et al. (2019) [10] optimized short-term electric load forecasting using a GA-PSO and BPNN hybrid model. These studies and their proposed hybrid models are summarized in Table 1.

TABLE I: Recent studies established hybrid models for time series forecasting

Authors	Field of Study	Proposed Hybrid Model
Silva, 2017	Short-term wind analysis and forecasting	Statistical modeling
Hajirahimi and Khashei, 2016	Financial forecasting	ARIMA - Artificial Neural Networks (ANN)
Alencar, 2018	Wind speed prediction for wind energy generation	SARIMA - Artificial Neural Networks
Fernandes, 2018	Short-term wind power forecasting	Spectral analysis - Time series decomposition
Silva, 2018	Analysis of extreme values in wind series	Three methods of dimensionality reduction
Zhang et al., 2019	Short-term wind speed forecasting	VMD - WT - PCA - BPNN - RBFNN
Yuan et al., 2019	Prediction interval of wind power	Beta distribution function based PSO - LSTM neural network
Lan et al., 2019	Day-ahead spatiotemporal solar irradiation forecasting	DFT - PCA - Elman Neural Network (ENN)
Kumar Handar, 2019	Stock market prediction	Wavelet-adaptive network-based fuzzy inference system (WANFIS)
Liu et al., 2019	Electrical energy consumption forecasting	EMD - MAF - LSSVR - QES
Hu et al., 2019	Short-term electric load forecasting	GA-PSO - BPNN

III. Proposed Method

In time series forecasting, various models and techniques are employed to improve the accuracy and reliability of predictions. This study is exploratory in nature and quantitative, as it utilizes statistical techniques for data analysis and modeling [11]. The process begins with data collection and preprocessing, essential steps to ensure the quality and suitability of the data for analysis. The subsequent stages involve determining the model type—whether additive or multiplicative—using the amplitude versus mean plot, and applying necessary transformations like the Box-Cox transformation for multiplicative models [12].

To evaluate trends and seasonality, the Cox-Stuart test and the Kruskal-Wallis test are utilized [13]. These tests are crucial for understanding the underlying patterns in the data, which inform the selection and parameterization of forecasting models. The ARIMA model, known for its effectiveness in capturing linear relationships, is then developed. This involves analyzing autocorrelation function plots to determine the need for differencing and evaluating the p, d, and q parameters along with seasonal factors to build a SARIMA model. It is necessary for the time series to be stationary, meaning it propagates randomly around a constant mean and variance [14].

A. ARIMA Model

The ARIMA model, also known as the Box-Jenkins approach, is among the most effective classical time series models for short-term forecasting. This model, represented as ARIMA (p, d, q), consists of three components: autoregression (AR), which indicates the series' dependence on its past values and is denoted by the parameter p; moving average (MA), which relates to the dependency of error terms on past lags and is denoted by q; and the integrated part, used when the series is non-stationary, denoted by d. The Box-Jenkins methodology involves four procedures: model identification, parameter estimation, diagnostic checking, and forecasting. Stationarity of the series is tested using specific tests, and the model is identified through the autocorrelation function. The estimation process follows, during which the parameters of the model are evaluated using diagnostic checks. If the

candidate model meets the criteria, it is used for forecasting [15]. The general form of the ARIMA model is represented by the equation 1:

$$\begin{aligned} Z_t = \mu &+ (\phi_1 - \theta_1)Z_{t-1} + (\phi_2 - \phi_1)Z_{t-2} \\ &+ \ldots + (\phi_p - \phi_{p-1})Z_{t-p} \\ &+ a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \\ &- \ldots - \theta_q a_{t-q} \end{aligned} \tag{1}$$

Where Z_t is the data at period t, μ is a constant value, $\phi_1, \ldots, \phi_p$ are the autoregressive parameters, and $\theta_1, \ldots, \theta_q$ are the moving average parameters. Identifying the model involves plotting time series data and verifying the stationarity of the mean and variance [15].

However, linear models like ARIMA may not fully capture the complexities in the data, which is why this research incorporates Multi-Layer Perceptrons (MLP), a type of neural network capable of modeling non-linear relationships.

B. MLP Model

The multilayer perceptron (MLP) machine learning model is highly regarded for its versatility in various applications and accuracy in time series forecasting. The MLP model excels at approximating continuous, nonlinear, differentiable, and constrained functions. It consists of an input layer, an output layer, and one or more hidden layers. Information is transferred from one layer to the next by artificial neurons. Depending on the specific application, hidden layers gather data from input layers and process it through a nonlinear function before passing it to other layers [16]. The combination of input, hidden, and output layers forms the MLP model. This model employs neural network operations with adaptive weighting to handle input data. The output of the MLP ANN is determined by the activation function, input weights, and bias [17]. This is represented in equation 2.

$$Y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{2}$$

The neuron's output is represented by Y, the input by x, the weights by w, and the bias by b. The activation function is defined in equation 2. This study utilizes sigmoid, hyperbolic tangent, and rectified linear unit functions. Equations 3, 4, and 5 illustrate these functions.

$$F_s = f(x) = \frac{1}{1 + e^x} \tag{3}$$

$$F_t = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{4}$$

$$F_r = f(x) = \max(0, x) \tag{5}$$

F_s refers to the sigmoid function. F_t denotes the hyperbolic tangent function. F_r signifies the rectified linear unit function. The process of establishing an MLP ANN model encompasses data selection, model training, and validation [17].

C. Model Evaluation

The accuracy of the forecasting model was evaluated using the Root Mean Square Error (RMSE), which measures the average magnitude of the errors between predicted and observed values. RMSE is widely used due to its straightforward interpretation and sensitivity to large errors. It is defined as equation 6:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2} \tag{6}$$

where X_i represents the observed values, Y_i represents the predicted values, and m is the number of observations. Lower RMSE values indicate better model performance, as they signify that the differences between predicted and observed values are smaller. The RMSE calculation was performed to identify the optimal model for wind energy generation forecasting, ensuring minimal prediction error and maximizing accuracy [18].

D. ARIMA-MLP Model

This section discusses the hybrid forecasting model combining autoregressive integrated moving average (ARIMA) with multilayer perceptron (MLP). The ARIMA model, a linear model, is often used to forecast time-series data, but it can produce nonlinear residuals. To address these residuals, multiple layer perceptron networks (MLPs), a type of nonlinear model, are employed. By applying ARIMA to the linear components of the time-series data, the residuals generated are then modeled using MLP. Figure 1 illustrates the framework of the ARIMA-MLP hybrid model.

In the ARIMA-MLP hybrid model, ARIMA is used initially to capture the linear aspects of the data. The residuals from this step, denoted as e_t, represent the difference between the observed values y_t and the fitted values $\hat{L}_t$ from the ARIMA model. This is shown in equation 7.

$$e_t = y_t - \hat{L}_t \tag{7}$$

Here, $\hat{L}_t$ is the ARIMA model's forecast at time t. The nonlinear relationships present in the residuals are then captured using an MLP. The MLP is trained on these residuals, where the residuals from ARIMA are used as input nodes to the MLP. Equation 8 describes this process.

$$\begin{aligned} e_t &= F(e_{t-1}, e_{t-2}, \ldots, e_{t-n}) + \epsilon_t \\ &\Rightarrow \hat{N}'_t = F(e_{t-1}, e_{t-2}, \ldots, e_{t-n}) \end{aligned} \tag{8}$$

In this equation, $\hat{N}'_t$ is the value predicted by the MLP at time t based on the residuals e_t, and ϵ_t represents the random error term. The function F captures the nonlinear patterns learned by the MLP.

The final prediction $\hat{y}_t$ is obtained by summing the SARIMA output $\hat{L}_t$ and the weighted MLP residual prediction $w \cdot \hat{N}'_t$. The formula can be represented in equation 9:

$$\hat{y}_t = \hat{L}_t + w \cdot \hat{N}'_t \tag{9}$$

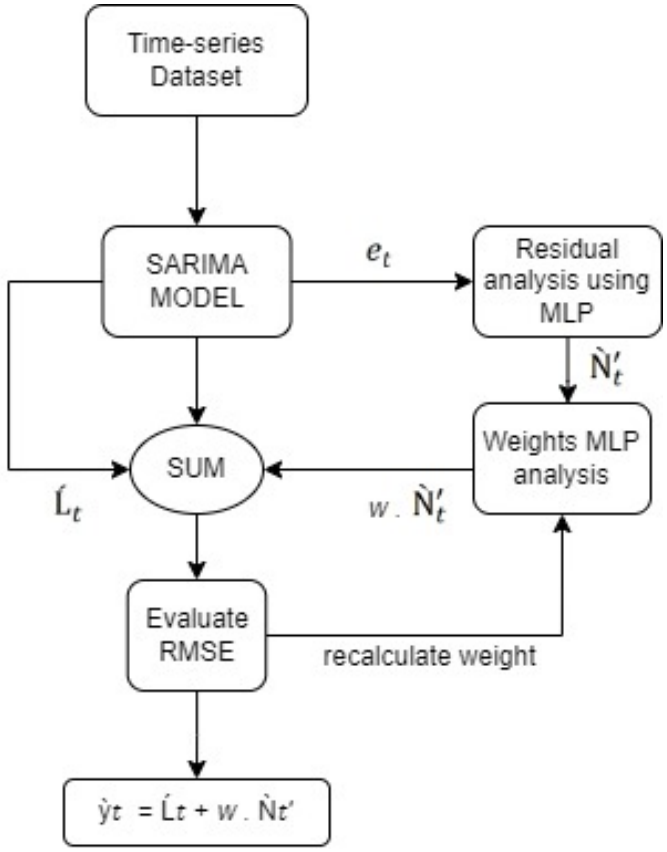


Fig. 1: The Framework of ARIMA-MLP Hybrid Model.

IV. Experiments

A. Time Series Data Visualization

"Fig. 2" presents the time series graph of electricity generated by the wind power matrix in Brazil, measured in megawatt-hours (MWh), from January 2007 to June 2024. A noticeable trend in energy production can be seen starting in 2015, the year when the National Bank for Economic and Social Development (BNDES) allocated R$ 6.6 billion for wind power generation projects. Additionally, a periodic pattern indicating a possible seasonal effect is evident. However, visual inspection alone cannot confirm the presence of trend and seasonality components; the Cox-Stuart and Kruskal-Wallis tests are required for verification.

, even at the beginning of a sentence.

B. Model Type Determination

Before analyzing the trend and seasonality components, it is necessary to determine whether the model is additive or multiplicative. For this purpose, the amplitude versus mean graph was plotted, as shown in "Fig. 3". This graph revealed a positive relationship between the subgroup means of the original series observations and their amplitude, with a slope coefficient of 0.7507, estimated through simple linear regression. A t-test was conducted to statistically verify if the slope coefficient of the fitted line was significant. Since the p-value was less than 0.01, the null hypothesis that the slope

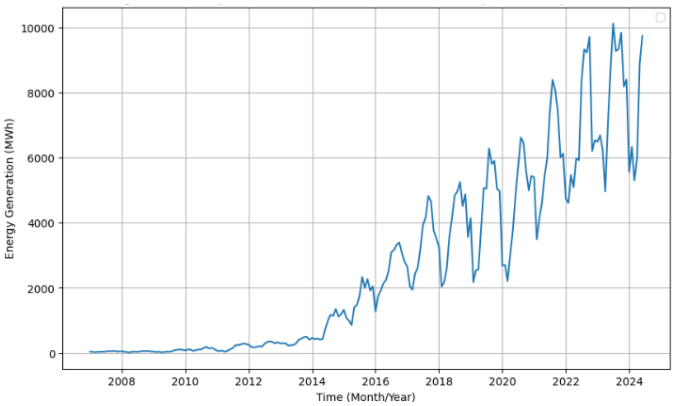


Fig. 2: Monthly Series of Wind Power Generation from Jan/2007 to Jun/2024.

coefficient is zero was rejected at a 5% significance level, indicating a multiplicative model. To make it additive and stabilize variance, a Box-Cox transformation was applied to the data. The estimated value of λ was 0.1706, and the data transformation was performed.

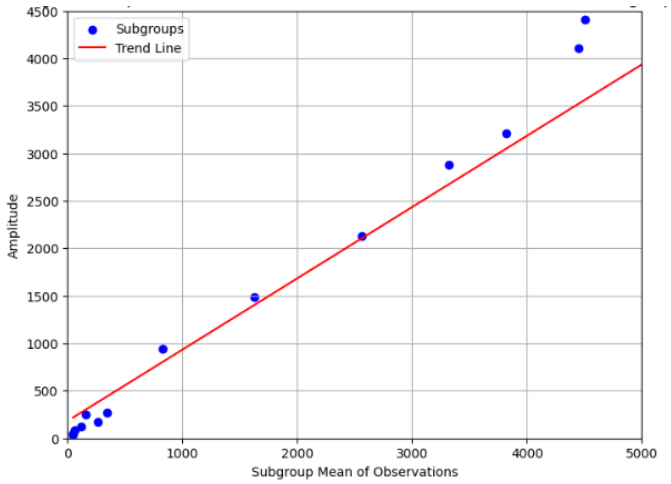


Fig. 3: Amplitude vs. Mean Plot for Wind Power Generation Time Series Subgroups.

C. Application of Cox-Stuart and Kruskal-Wallis Tests

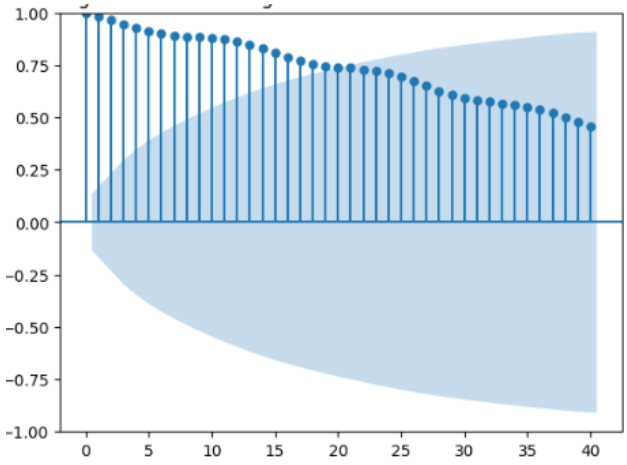
To assess the presence of trends and seasonality in the data, the Cox-Stuart test for trend analysis and the Kruskal-Wallis test for seasonality assessment were applied. The results of these tests are summarized in Table 2. The Cox-Stuart test confirmed the existence of a trend component with a p-value of 4.9304×10^{-32}, rejecting the null hypothesis of no trend at a 5% significance level. Therefore, it was necessary to perform a first-order difference on the transformed series to remove the trend effect. The Kruskal-Wallis test resulted in a p-value of 0.87878, not rejecting the null hypothesis of no deterministic seasonality at a 5% significance level, meaning the series does not require a seasonal difference.

TABLE II: Cox-Stuart and Kruskal-Wallis Tests

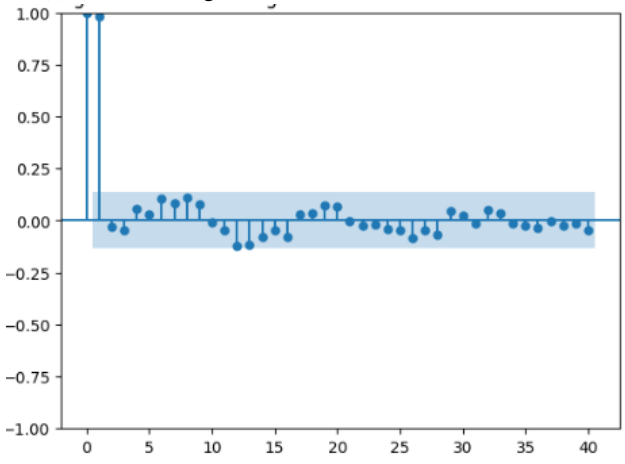
Test	P-value
Cox-Stuart	4.9304×10^{-32}
Kruskal-Wallis	8.7878×10^{-1}

D. ARIMA Model Fitting

"Fig. 4a" shows the correlogram of the autocorrelation function (ACF) for the transformed time series, indicating that the series is not stationary since the lags decay slowly to zero. The partial autocorrelation function (PACF) "Fig. 4b" shows significant correlations at lags 1 and 3, with a tendency towards zero in subsequent lags, suggesting a probable moving average process.



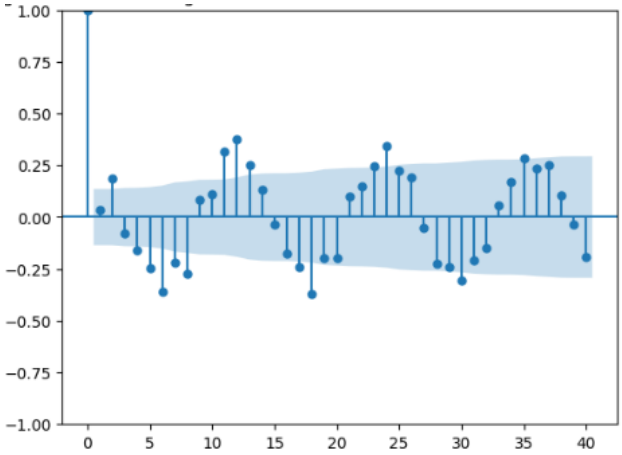
(a) ACF Correlogram for the Transformed Time Series.



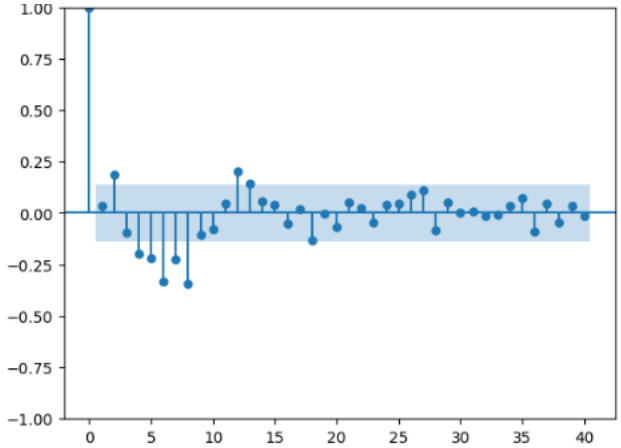
(b) PACF Correlogram for the Transformed Time Series.

Fig. 4: Correlograms for the Transformed Time Series: (a) ACF Correlogram, (b) PACF Correlogram.

"Fig. 5" presents the correlograms of the ACF and PACF for the transformed series with a first-order difference. From the PACF correlogram, significant autocorrelations different from zero are observed at lags 2, 4, 5, 6, 7 and 8 indicating possible orders for the autoregressive part "p". Since there are significant seasonal lags in the PACF, the seasonal order "P" is one. In the ACF correlogram, significant lags are 2, 4, 5, 6, 7 and 8, indicating possible orders for the moving average part "q". The presence of significant seasonal lags in the ACF suggests a probable seasonal order "Q" of two. Thus, various SARIMA (p,d,q)×(P,D,Q)S models were proposed.



(a) ACF Correlogram for Transformed and Differentiated Time Series.



(b) PACF Correlogram for Transformed and Differentiated Time Series.

Fig. 5: Correlograms for the Transformed Time Series: (a) ACF Correlogram, (b) PACF Correlogram.

The ARIMA model parameters p, d, and q, along with the seasonal parameters P, D, Q, and S, were then estimated. The optimal values of these parameters were determined by minimizing the Root Mean Square Error (RMSE) and Akaike Information Criterion (AIC). The final fitted ARIMA model is summarized in Table 3.

The models with uncorrelated residuals and lower AIC values, shown in Table 3, were used to make predictions for the period from January 2023 to June 2024. From these predictions and the actual observations during this period, the root mean squared error (RMSE) was calculated. As one of the objectives of this work is to make predictions for electricity production generated by the wind power matrix, the SARIMA (7,1,2)×(1,0,2)12 model was chosen because it presented the

TABLE III: Summary of ARIMA Model Fitting

ARIMA Order	Seasonal Order	RMSE	AIC
(7, 1, 2)	(1, 0, 2, 12)	1002.994354	300.429472
(2, 1, 8)	(1, 0, 2, 12)	1044.494717	297.472426
(2, 1, 7)	(1, 0, 2, 12)	1047.731456	296.803581
(8, 1, 2)	(1, 0, 2, 12)	1062.534420	301.503035
(2, 1, 2)	(1, 0, 2, 12)	1064.216903	296.786647

lowest root mean square prediction error value.

E. MLP Model Training and Residual Analysis

The residuals from the ARIMA model were analyzed using a Multilayer Perceptron (MLP) model. The MLP was trained on these residuals to capture any nonlinear patterns not accounted for by the ARIMA model.

To optimize the MLP model, a grid search approach was employed, testing different combinations of hyperparameters to identify the best configuration. The parameters included in the grid search were the sizes of the hidden layers, the alpha values for regularization, the solvers, and the activation functions. The grid search used cross-validation with a time series split, ensuring that the model was validated on different segments of the data to avoid overfitting. The grid search process involved the following steps:

a) Model Definition: The MLPRegressor was defined with a maximum of 1000 iterations and a random state of 42 for reproducibility.

b) Parameter Grid: A parameter grid was created, specifying the possible values for the hidden layer sizes, alpha, solver, and activation function.

c) Scoring Function: A custom scoring function was defined to evaluate the models based on the RMSE.

d) Cross-Validation: Time series split cross-validation was used to ensure the model was validated on different time segments, with 5 splits.

e) Grid Search Execution: The GridSearchCV function was used to perform the grid search, fitting the model on the training data and evaluating it using the defined scoring function.

The performance of the MLP model in predicting the residuals is shown in Table 3.

TABLE IV: Summary of MLP Model Results

Hidden Layer Sizes	Alpha	Solver	Activation	Mean Test RMSE	Std Test RMSE
(100, 50)	0.1	sgd	relu	0.559031	0.087310
(100, 50)	0.01	sgd	relu	0.559221	0.087170
(100, 50)	0.001	sgd	relu	0.559243	0.087153
(100, 50)	0.0001	sgd	relu	0.559246	0.087152
(150, 150)	0.1	sgd	relu	0.572078	0.073071

The final prediction $\hat{y}_t$ is obtained by combining the ARIMA model output $\hat{L}_t$ with the weighted MLP residual prediction $w \cdot \hat{N}'_t$.

F. Optimization of Weight Calculation

In this study, the optimal weight for the ARIMA-MLP hybrid model was determined through an exhaustive search within a specified range of weights. The process was designed to identify the weight that minimized the Root Mean Square Error (RMSE) of the combined model predictions. Here's a detailed description of the steps taken:

1) **Initialization of Parameters**: The search range for the weights was set from 0 to 5000, with an incremental step of 1. This range was chosen based on prior experimentation and domain knowledge to ensure a comprehensive search.
2) **Variables for Tracking**: Two variables were initialized to keep track of the best RMSE value (`best_rmse`) and the corresponding weight (`best_weight`). The initial `best_rmse` was set to infinity to ensure that any calculated RMSE would be lower.
3) **Iterative Search**:
   a) A loop was employed to iterate over the possible weight values within the specified range.
   b) For each weight, the combined predictions were calculated by adding the weighted residual predictions from the MLP model to the ARIMA model predictions.
4) **RMSE Calculation**: The RMSE for each set of combined predictions was computed against the actual test data. The RMSE is a standard metric for measuring the accuracy of predictions in time series analysis.
5) **Best Weight Selection**:
   a) If the RMSE of the current weight was lower than the previously recorded best RMSE, the current weight was stored as the best weight.
   b) The best RMSE was updated accordingly.
6) **Application of Best Weight**: Once the loop concluded, the best weight identified was used to combine the ARIMA and MLP predictions for future forecasts. This final combination utilized the optimized weight to ensure the lowest possible RMSE, thereby enhancing the predictive accuracy of the hybrid model.

By methodically exploring the weight range and evaluating the RMSE for each, the study effectively optimized the weight parameter, leading to improved performance of the ARIMA-MLP hybrid model.

This methodical approach ensures that the hybrid model leverages the strengths of both ARIMA and MLP, providing more accurate and reliable forecasts for wind energy generation.

V. Discussion and Conclusion

A. Model Performance Comparison

The performance of the proposed ARIMA-MLP hybrid model was compared to the model applied by Silva [19], who used the same dataset. The comparison, shown in Table V, reveals that the SARIMA-MLP model outperforms the SARIMA model by a notable margin. Specifically, the RMSE for the SARIMA-MLP model is 710.69, compared to 773.79 for the SARIMA model, representing an 8.16% improvement in predictive accuracy. This enhancement demonstrates the effectiveness of integrating SARIMA with MLP, particularly

in capturing the complex patterns inherent in wind energy generation data. The optimized weight calculation further fine-tunes the model, ensuring minimal prediction error and maximizing accuracy.

TABLE V: Comparison of Model Performance

Model	RMSE
SARIMA	773.790000
SARIMA-MLP	710.689507

The SARIMA-MLP model showed an RMSE of 710.69, which is lower than the RMSE of 773.79 obtained by the SARIMA model. This represents an improvement of approximately 8.16% in predictive accuracy. The improvement percentage is calculated as follows:

$$\text{Improvement}\% = \left(\frac{\text{RMSE}_{\text{SARIMA}} - \text{RMSE}_{\text{SARIMA-MLP}}}{\text{RMSE}_{\text{SARIMA}}} \right) \times 100$$

$$\text{Improvement}\% = \left(\frac{773.79 - 710.69}{773.79} \right) \times 100 \approx 8.16\%$$

This improvement demonstrates the effectiveness of the hybrid approach in reducing prediction error and enhancing the accuracy of wind energy generation forecasts.

B. Forecasting Future Wind Power Generation

Finally, the improved hybrid model was used to forecast future wind power generation from January 2024 to December 2024. The forecasts are presented in "Fig. 6" and the referenced values on Table 6.

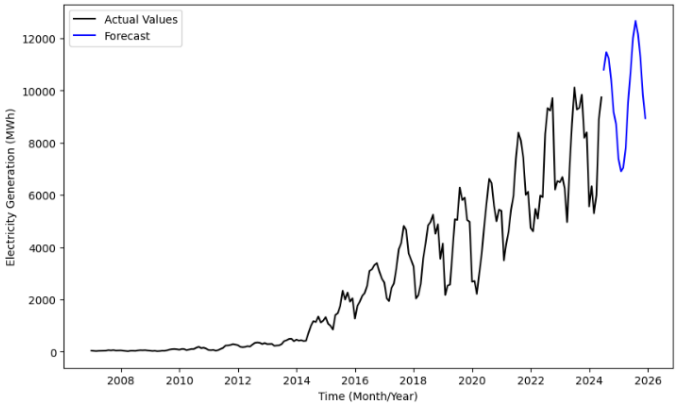


Fig. 6: Predictions for Wind Power Generation.

The proposed hybrid model's forecasts from January 2024 to December 2025, shown in Table VI, reveal periods of high and low wind energy generation. For instance, peak generation is predicted in August 2024 and August 2025, while the lowest generation occurs in February 2025. These predictions are crucial for managing energy resources, planning maintenance, and ensuring a stable energy supply. Accurately predicting these fluctuations can help mitigate the risks associated with energy shortages and optimize the use of wind energy, which is inherently variable.

TABLE VI: Predictions for Wind Energy Generation (MWh)

Date	Predictions (MWh)
2024-07-01	10802.102
2024-08-01	11476.124
2024-09-01	11239.953
2024-10-01	10423.430
2024-11-01	9183.023
2024-12-01	8724.493
2025-01-01	7391.026
2025-02-01	6906.526
2025-03-01	7047.664
2025-04-01	7812.844
2025-05-01	9568.568
2025-06-01	10752.590
2025-07-01	12016.607
2025-08-01	12674.114
2025-09-01	11774.026
2025-10-01	11259.817
2025-11-01	9829.588
2025-12-01	8943.126

The SARIMA model, known for its robustness in capturing seasonality and trends in time series data, offers simplicity and ease of implementation. However, it may fall short when dealing with nonlinear patterns. On the other hand, the MLP model excels in modeling complex nonlinear relationships but can struggle with seasonality and trend components. By combining these models, the SARIMA-MLP hybrid approach leverages the strengths of both models, addressing their individual limitations.

Future work could explore further enhancements to the hybrid model, such as incorporating other machine learning algorithms or optimizing additional parameters. Additionally, expanding the model to include exogenous variables, such as weather data, could improve forecast accuracy. These advancements could provide more robust and reliable predictions, aiding in the efficient management and planning of wind energy resources.

REFERENCES

[1] ONS, "Operador Nacional do Sistema Elétrico," 2021. [Online]. Available: http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/geracao_energia.aspx.

[2] MME, "Ministério de Minas e Energia," 2021. [Online]. Available: https://static.poder360.com.br/2021/08/1-Boletim-Mensal-de-Energia-Maio-2021-Portugues.pdf. Accessed: Sep. 10, 2021.

[3] D. B. Alencar, "Modelo híbrido baseado em séries temporais e redes neurais para prevenção da geração de energia eólica," Ph.D. dissertation, Univ. Fed. do Pará, Belém, PA, Brazil, 2018.

[4] L. C. Morais, "Estudo sobre o panorama da energia elétrica no Brasil e tendências futuras," M.S. thesis, Univ. Estadual Paulista, Bauru, SP, Brazil, 2015.

[5] L. C. Fernandes, "Previsão de potência eólica de curtíssimo prazo baseada na análise espectral e decomposição da série temporal," M.S. thesis, Univ. Fed. de Pernambuco, Recife, PE, Brazil, 2018.

[6] H. B. Silva, "Técnicas para redução de dimensionalidade de séries temporais e detecção de velocidades extremas do vento para geração de energia eólica," M.S. thesis, Univ. Fed. de Pernambuco, Recife, PE, Brazil, 2018.

[7] P. K. O. Silva, "Análise e previsão de curto prazo do vento através de modelagem estatística em áreas de potencial eólico no Nordeste do Brasil," Ph.D. dissertation, Univ. Fed. de Campina Grande, Campina Grande, PB, Brazil, 2017.

[8] Z. Hajirahimi and M. Khashei, "Improving the performance of financial forecasting using different combination architectures of ARIMA and ANN models," Journal of Industrial Engineering and Management Studies, vol. 3, pp. 17-32, 2016.

[9] Z. Liu, X. Wang, Q. Zhang, and Ch. Huang, "Empirical mode decomposition based hybrid ensemble model for electrical energy consumption forecasting of the cement grinding process," Measurement, vol. 138, pp. 314–324, 2019.

[10] Y. Hu et al., "Short term electric load forecasting model and its verification for process industrial enterprises based on hybrid GA-PSO-BPNN algorithm - A case study of papermaking process," Energy, vol. 170, pp. 1215-1227, 2019.

[11] A. S. Pereira, D. M. Shitsuka, F. J. Parreira, and R. Shitsuka, "Metodologia da pesquisa científica," UFSM, 2018. [Online]. Available: https://repositorio.ufsm.br/bitstream/handle/1/15824/Lic_Computacao_Metodologia-Pesquisa-Cientifica.pdf?sequence=1.

[12] G. E. P. Box and G. M. Jenkins, "Time Series Analysis: Forecasting and Control," Holden-Day, 1976.

[13] P. H. A. B. Santos, O. A. S. Delfino, R. V. R. Santos, and M. Nascimento, "Ajuste de um modelo de séries temporais para prever a precipitação pluviométrica," Research, Society and Development, vol. 10, no. 6, pp. 1-11, 2021.

[14] P. A. Morettin and C. M. C. Toloi, "Análise de Séries Temporais," 2nd ed., Edgard Blucher, 2006.

[15] I. Syahrini, R. Radhiah, W. F. Damanik, N. Sciences, U. S. Kuala, and B. Aceh, "Application of the Autoregressive Integrated Moving Average (ARIMA) Box-Jenkins Method in Forecasting Inflation Rate in Aceh," Transcendent J. Math. Appl. ISSN, vol. 2, no. 1, pp. 27–33, 2023.

[16] M. Qureshi, M. Daniyal, and K. Tawiah, "Comparative Evaluation of the Multilayer Perceptron Approach with Conventional ARIMA in Modeling and Prediction of COVID-19 Daily Death Cases," J. Healthc. Eng., vol. 2022, 2022, doi: 10.1155/2022/4864920.

[17] H. S. Wicaksana et al., "Air Temperature Sensor Estimation on Automatic Weather Station Using ARIMA and MLP," vol. 14, no. 2, 2022.

[18] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," PeerJ Comput. Sci., vol. 7, pp. 1–24, 2021, doi: 10.7717/PEERJ-CS.623.

[19] M. dos S. Silva, P. H. A. B. Santos, R. V. R. dos Santos, M. do Nascimento, M. A. R. de Pascoa, R. N. Pereira, T. A. de Oliveira, "Análise da energia eólica no Brasil usando Séries Temporais," *Research, Society and Development*, vol. 11, no. 1, e26611124827, Jan. 2022, doi: http://dx.doi.org/10.33448/rsd-v11i1.24827.