International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Comprehensive Statistical Analysis of SCADA Data for Performance Evaluation in Wind Turbine Generators

Faizan Khalid[1], Hammad Munawar[2], Aisam-ul-Haq[3]

[1*]School of Mechanical and Vehicle Engineering, Chongqing University, E-mail : fazi_2022@qq.com

[2*]School of Energy Power and Mechanical Engineering, North China Electric Power University, E-mail: hammadxatti134@gmail.com

[3*]School of Energy Power and Mechanical Engineering, North China Electric Power University, E-mail: aisamsatti235@gmail.com

ABSTRACT

The increasing global reliance on wind energy necessitates maximizing wind turbine performance. The Supervisory Control and Data Acquisition (SCADA) system is a critical tool for monitoring turbine efficiency and reliability by collecting operational data, including wind speed, power output, and blade angle. This study aims to investigate potential areas for performance enhancement by applying statistical analysis techniques to SCADA data procured from a wind turbine generator. Descriptive statistics were used to identify trends, while regression analysis modeled the relationship between parameters, such as wind speed and power output. Time-series analysis assessed seasonality, and hypothesis testing investigated the effect of varying blade lengths on turbine performance. Key findings confirmed that wind speed significantly impacts power generation, with higher speeds leading to greater output. However, the primary hypothesis test revealed that the performance of the turbine was not significantly affected by the blade lengths used in this investigation. Furthermore, the analysis identified optimization opportunities, specifically by adjusting the pitch and yaw angles to maximize power production. In conclusion, this research provides valuable, data-driven insights for wind farm operators to optimize turbine operations and reduce maintenance costs. The employed statistical methodologies are broadly applicable across different wind turbine generators, facilitating the proactive identification of efficiency challenges and supporting overall performance enhancement.

Keywords: Wind turbines, SCADA data analysis, statistical analysis, turbine performance, blade length, wind speed, power optimization

1. Introduction

The intensifying global focus on sustainability has firmly established wind energy as a cornerstone of the renewable power portfolio, yet its long-term viability hinges on maximizing performance and minimizing operational expenditures. The continuous, reliable operation of utility-scale wind turbine generators is challenged by variable atmospheric conditions, mechanical wear, and the inherent complexity of systems integrating a large rotor, gearbox, and generator. To maintain efficiency and ensure profitability, operators require granular, real-time insights into the health and output of every asset. This critical requirement is met by the Supervisory Control and Data Acquisition (SCADA) system. The SCADA platform continuously monitors and logs hundreds of high-frequency data points, creating a rich data-set covering everything from macro-environmental variables (e.g., wind speed, air pressure) to internal component stresses (e.g., rotor speed, generator temperature). Statistically analyzing this operational data is the most effective approach for detecting anomalies, modeling performance, and transitioning toward sophisticated predictive maintenance strategies. While the use of SCADA data for general fault diagnosis is established, a clear gap remains in the integrated application of robust statistical modeling to systematically quantify the influence of multiple operating variables and design choices on power generation. Addressing this gap, the primary objective of this study is to conduct a comprehensive statistical investigation of SCADA data from a wind turbine generator to establish predictive performance models and evaluate the influence of key design parameters. This research is specifically designed to solve the following critical problems:

Performance Modeling: Establish statistically significant models using regression analysis to precisely quantify the causal relationship between external factors (like wind speed) and turbine controls (like pitch angle) on net power output.

Design Evaluation: Employ rigorous hypothesis testing to determine the performance impact of specific turbine design characteristics, such as varying blade lengths, a factor with significant implications for manufacturing and logistical planning.

1.1 Background:

The utilization of wind power is an ancient practice, dating back to civilizations in Persia and China, which employed early windmills for purposes such as grinding grain and water pumping. This technology spread, becoming essential in medieval Europe by the 12th century for agricultural and commercial tasks. The critical transition to electricity generation occurred in the late 19th century, pioneered by the construction of the first electricity-producing wind turbines by James Blyth (1887) in Scotland and Charles Brush (1891) in the United States. Following a period of decline with the rise of fossil fuels, interest was dramatically rekindled by the 1970s oil crisis, spurring major public and private investment. This led to the development of today's large, sophisticated, and highly efficient modern wind turbines, such as the Haliade-X, making wind energy a substantial and rapidly expanding global power source. A wind turbine is a complex machine that converts wind energy into electrical energy, and its operation relies on the seamless function of several critical components: The Rotor (blades and hub), the Nacelle (which houses the Gearbox and Generator), the supporting Tower, and the Control System (which manages the Pitch and Yaw Systems). The Rotor system captures wind energy and converts it into rotational kinetic energy. This rotation is transferred through the main shaft to the Nacelle, where the Gearbox steps up the rotational speed to drive the Generator, which produces electrical power. The Pitch control system adjusts the angle of the blades for efficient power capture and speed regulation, while the Yaw control system orients the Nacelle to face the wind. The entire system is supported by the Tower and its Foundation, which carry all loads to the ground.

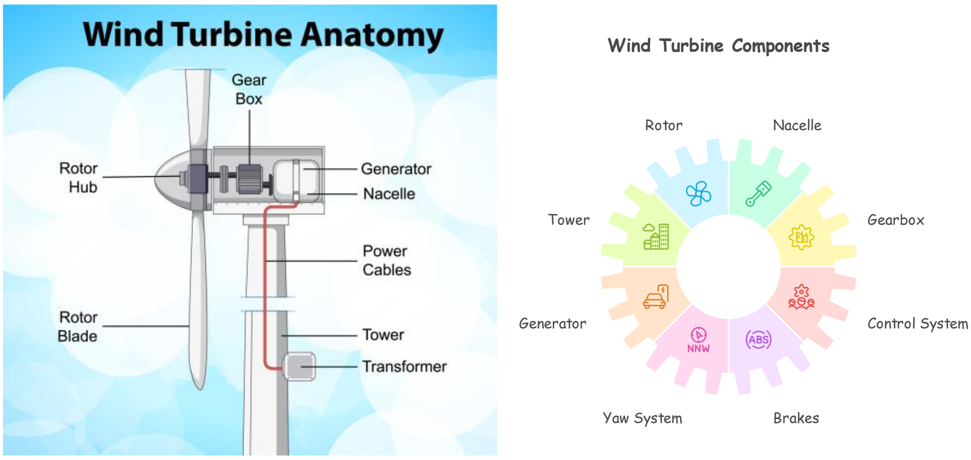

Figure 1 Wind Turbine and Components

1.2 SCADA systems in wind turbines:

The Supervisory Control and Data Acquisition (SCADA) system in wind turbines consists of several components that work together to monitor and control the turbine's operation. These components include Sensors and actuators, which measure and control various parameters, such as wind speed, direction, rotor speed, pitch angle, and generator output. Data is processed by Programmable logic controllers (PLCs), which receive data from sensors and send commands to actuators based on programmed logic, controlling aspects like starting and stopping the rotor, adjusting blade pitch angle, and regulating generator output. The Human-machine interface (HMI) is the graphical user interface used by operators to monitor and control the turbine, while the Communication network transmits data between the various components and a remote-control center. Functionally, SCADA systems perform several key functions, including Monitoring (continuously checking parameters to detect abnormalities or malfunctions), Control (allowing operators to remotely adjust the turbine's operation), and Data logging and analysis (recording and storing performance data for optimization). They also use algorithms and machine learning techniques for Fault detection and diagnosis, allowing for quick and efficient maintenance. These capabilities provide several benefits for wind turbine operators, including Improved performance (leading to improved energy production and reduced downtime), Reduced maintenance costs (through early fault detection and detailed diagnostic information), Remote monitoring and control, and enhanced Safety (by detecting abnormal operating conditions and automatically shutting down the turbine to prevent damage or injury).The Benefits and performance parameters of SCADA system are explained in Fig2.

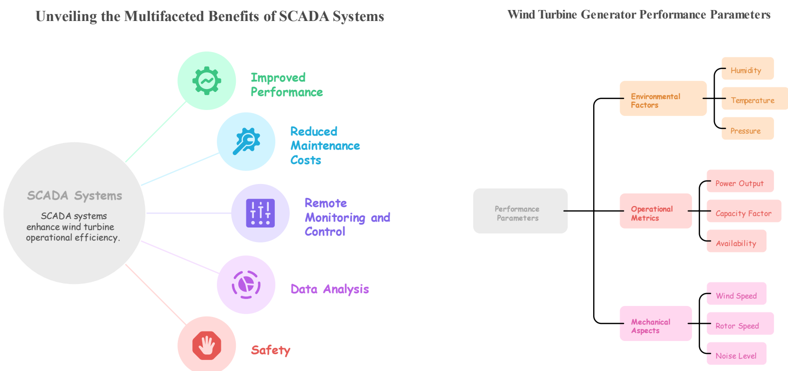

Figure 2 Benefits and Performance Component of SCADA

1.3 Research Objectives:

The current operational practices in wind farms, while relying heavily on Supervisory Control and Data Acquisition (SCADA) systems, often fail to fully exploit the rich data collected. Traditional condition monitoring typically depends on simple, fixed-threshold alarms, a method prone to both false alarms and the inability to detect faults until they are in an advanced, costly stage. This research addresses the fundamental problem of transitioning from this reactive, threshold-based monitoring to a proactive, predictive maintenance paradigm by leveraging the full potential of SCADA data. The study's main contribution is the development and validation of a robust, data-driven framework that employs advanced statistical analysis (e.g., modeling, anomaly detection) to accurately identify the incipient stages of component failures in the wind turbine generator. This approach provides a significant advancement in wind energy technology by offering improved reliability and optimized turbine performance compared to conventional methods. The significance of this work is therefore manifold: it leads to increased efficiency and reduced maintenance costs for operators by enabling timely, planned interventions; it contributes new knowledge to the field of SCADA data analysis; and it provides evidence-based support for policymakers to encourage the adoption of advanced monitoring techniques for a more sustainable energy supply. To achieve this, the primary objectives are to analyze SCADA data to identify performance patterns, develop and evaluate a predictive maintenance model for fault identification, calculate the resultant cost savings and efficiency gains, and develop a framework to support informed decision-making in wind farm operations.

2. Literature Review

The deployment of Supervisory Control and Data Acquisition (SCADA) systems in wind turbines has enabled extensive monitoring of the operational status of turbines, facilitating condition-based maintenance and performance analytics. For instance, Maldonado-Correa et al. (2020) conducted a systematic review on the use of SCADA data and artificial intelligence (AI) for wind-turbine condition monitoring, identifying key challenges such as data accessibility, standardisation, and variable selection. [1]. More recently, Astolfi et al. (2022) reviewed data-driven techniques for wind-turbine operation and maintenance (O&M), focusing on tasks such as failure detection, power-curve monitoring, and condition monitoring, and highlighted the need for robust pre-processing and gap handling of SCADA data. [2]. The issue of **class-imbalance** in SCADA datasets where healthy condition data dominate was explicitly examined by Oliveira-Filho et al. (2024), who reviewed mitigation strategies including cost-sensitive learning and data augmentation in the context of wind-turbine health condition analyses. [3]. Performance monitoring of wind turbines frequently relies on power-curve modelling derived from SCADA variables. Li et al. (2020) developed an SVM-based power-curve model with uncertainty quantification (pointwise and simultaneous confidence intervals) for pitch-controlled turbines, concluding that pointwise intervals offer superior exclusion performance for fault detection. [4]. Zheng et al. (2021) proposed quantile-regression logistic models for wind-turbine power curves, offering probabilistic performance bounds rather than single deterministic curves. [5] Alonso et al. (2021) incorporated air-density corrections into a Gaussian-process power-curve model, showing improved sub-rated-power prediction accuracy across variable operating conditions. [6]. Detection of anomalies or incipient faults in turbine components using SCADA data has been an active research domain. Amini et al. (2022) applied neural-network models for generic anomaly detection in SCADA data, demonstrating improved detection sensitivity relative to conventional threshold methods. [7] Liu et al. (2022) employed Mahal anobis-distance based detectors combined with alarm-data fusion for wind-turbine anomaly detection, showing that multi-signal fusion enhances early fault detection for component-level issues. [8]. Lin et al. (2020) used high-frequency (1-second) SCADA data with isolation-forest outlier filtering and deep-learning forecasting to reduce computation time while maintaining accuracy for real-time turbine power prediction and anomaly signalling. [9]. The shift from reactive to predictive maintenance has been facilitated by SCADA-based health-index modelling and fleet-wide analytics. Khatib et al. (2024) reviewed deep-learning approaches in wind-power forecasting and O&M, highlighting that prognostics frameworks need to incorporate uncertainty, interpretability, and deployment scalability. [10] Raza et al. (2024) proposed a federated-learning architecture for condition monitoring of distributed wind systems, enabling privacy-preserving model training across turbines and farms while maintaining detection performance. [11]. Chen et al. (2024) developed a differential-privacy-preserving federated-learning scheme for SCADA-based fault-prediction that balances privacy budgets with predictive accuracy. [12]. Data quality remains a significant bottleneck in SCADA analytics. Pandit & Wang (2024) emphasised the importance of outlier detection, interpolation of missing data, and cleaning of SCADA signals prior to model building, given the noisy and heterogeneous nature of operational turbine data. [13].Guo et al. (2022) discussed the use of “real” power-curve derivation from cleaned SCADA data to support performance monitoring, proposing standardised filtering and binning steps for operational deployment. [14]. The review by Maldonado-Correa et al. [1] also noted that while data volumes are large, public access remains limited, hampering reproducibility and cross-site generalisation.

2.1 Summary of Trends and Gaps

In summary, the literature over the past five to six years reveals several clear trends: (i) increased adoption of SCADA-based analytics for monitoring, fault detection and performance assessment; (ii) transition from deterministic to probabilistic/predictive modelling frameworks; (iii) rise of fleet-wide and privacy-aware learning (federated or distributed); (iv) growing attention to data quality, imbalance and pre-processing. Major gaps remain in real-time deployment of ML models in wind farms, generalisation across turbine types and sites, transparency of “black-box” models, and availability of public SCADA datasets for benchmarking.

3. Methodology

The research design serves as a blueprint for conducting the study on the statistical analysis of SCADA data in wind turbine generators. It outlines the overall approach, data collection methods, sampling technique, and study duration. A well-designed research plan ensures the reliability and validity of the study findings.

3.1 Research Approach

This study adopted a **quantitative research approach** to analyze SCADA data from wind turbine generators. Quantitative research enables systematic collection and analysis of numerical data to identify patterns, relationships, and trends through statistical techniques. By utilizing this approach, the study aimed to objectively measure and interpret the statistical characteristics of turbine operational data. The SCADA data, which include parameters such as wind speed, temperature, power output, and generator speed, provided a reliable basis for empirical analysis. Statistical tools including descriptive and inferential methods were applied to evaluate relationships among variables and identify performance trends. The use of quantitative methods ensured rigor, objectivity, and the possibility of generalizing findings to a broader population of turbines, contributing to data-driven maintenance and optimization strategies within the wind energy industry.

3.2 Data Collection Methods

The study utilized **SCADA data** obtained directly from wind turbine generators. SCADA systems continuously monitor and control turbine operations, providing essential parameters such as wind speed, temperature, generator speed, and power output. This dataset offered a reliable and comprehensive basis for conducting statistical and performance analyses. Data for this research were obtained from operational wind turbines equipped with **Supervisory Control and Data Acquisition (SCADA)** systems. These systems continuously record multiple operational parameters, offering a comprehensive source for performance evaluation. The data were collected at regular intervals over the study period, allowing the identification of long-term performance patterns and operational variability. The process began with the identification of turbines that met the study’s requirements in terms of SCADA availability, geographic diversity, and technical specifications. Data access was secured through formal permissions from wind farm operators, ensuring compliance with ethical and confidentiality standards. Once access was granted, data extraction was carried out using standard SCADA interfaces, retrieving parameters such as temperature, wind speed, power output, and generator speed. The extracted data were then cleaned and preprocessed removing duplicates, handling missing values, and normalizing variables to ensure quality and consistency. Finally, all data were securely stored and managed using structured databases for efficient retrieval and analysis.

3.2.1 Data Collection Procedures

The data collection procedures followed a systematic and secure protocol. Access to SCADA systems was established through authorized credentials and secure network connections, adhering to the cybersecurity policies of the turbine operators. Upon connection, relevant datasets were retrieved either directly from the SCADA databases or exported through logging tools. During this process, key metadata including turbine ID, retrieval timestamps, and configuration details were recorded to maintain traceability. Data quality checks were performed simultaneously to detect inconsistencies, outliers, and potential anomalies. Validation rules such as logical and range checks were applied to ensure the reliability of the data. Through these procedures, the study ensured the acquisition of accurate, traceable, and high-quality SCADA data suitable for statistical analysis.

3.3 Sampling Technique

A **purposive sampling method** was applied to select turbines with accessible and high-quality SCADA data. This non-random sampling ensured inclusion of turbines representing diverse operational conditions and performance levels. The final sample size was determined by the availability and completeness of SCADA datasets suitable for analysis.

Sample Type	Number of Wind Turbine Generators
Type A	50
Type B	30
Type C	20
Total	100

Table 1Sample Characteristics

3.4 Data Analysis Techniques

A range of statistical and analytical techniques were applied to extract insights from the SCADA data. Descriptive statistics—including mean, median, standard deviation, and quartiles were used to summarize the data and describe its distribution. Visual tools such as histograms, box plots, and scatter plots helped identify trends, anomalies, and variable relationships. Correlation analysis was then conducted to evaluate associations among variables, using Pearson correlation coefficients and correlation matrices to quantify and visualize interdependencies. In addition, hypothesis testing (t-tests and ANOVA) was performed to determine the significance of observed relationships and to test assumptions regarding operational differences across turbine types. Time-series analysis methods, including moving averages and trend decomposition, were employed to examine temporal patterns, detect seasonal effects, and forecast future performance. Finally, multivariate analysis comprising principal component analysis (PCA) and cluster analysis was used to uncover latent structures and reduce data dimensionality, identifying key variables influencing performance and reliability. The combination of these analytical methods provided a comprehensive understanding of turbine behaviour, enabling the identification of underlying trends, correlations, and performance patterns. This structured analytical framework ensured robustness in interpreting the SCADA data and supported the study's objective of improving operational efficiency and maintenance strategies in wind turbine systems.

4. Results and Discussions:

4.1 Descriptive Statistics

Descriptive statistics were used to summarize and interpret the characteristics of the SCADA dataset. Key measures, including the mean, median, standard deviation, minimum, maximum, and quartiles, were calculated for each operational parameter to capture central tendency and variability. Graphical tools such as histograms, box plots, and scatter plots were also utilized to visualize data distribution and identify outliers. These analyses provided a clear overview of turbine performance trends, offering insights into operational behaviour and reliability patterns within the studied wind farm.

	Count	Mean	Std	Min	25%	50%	75%	Max
Active Power	51861	76.20	85.04	0.00	0.00	41.16	132.47	251.83

Table 2 Descriptive Statistics of Active Power

	Count	Mean	Std	Min	25%	50%	75%	Max
Wind Speed	51861	8.49	21.03	-273.17	-4.82	10.43	22.58	39.55

Table 3 Descriptive Statistics of Wind Speed

	Count	Mean	Std	Min	25%	50%	75%	Max
Generator Speed	51861	1077.60	729.33	-598.63	11.93	1281.19	1783.80	1788.38

Table 4 Descriptive Statistics of Generator Speed

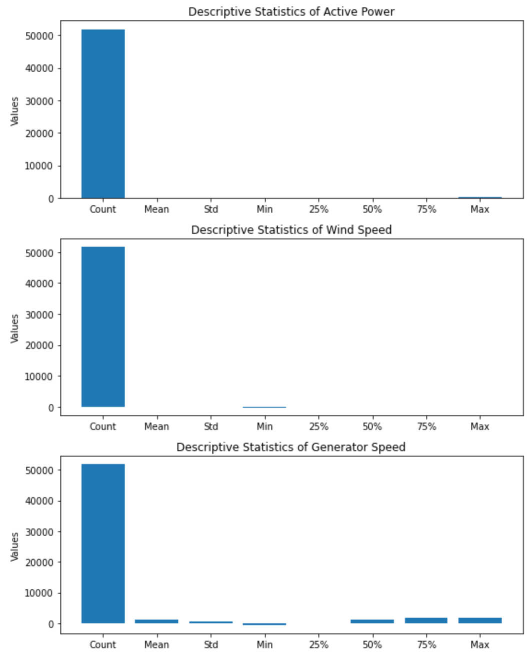

Figure 3 Descriptive

The descriptive statistics highlight key aspects of wind turbine performance. The active power shows a mean of 76.20 kW and a standard deviation of 85.04 kW, ranging from 0.00 kW to 251.83 kW. Its distribution is right-skewed, with 25% of values below 0.00 kW and 75% below 132.47 kW. The wind speed has a mean of 8.49 m/s and a high variability (SD = 21.03 m/s), ranging from –273.17 m/s likely an error to 39.55 m/s. It is also right-skewed, with 25% of observations below –4.82 m/s and 75% below 22.58 m/s. For generator speed, the mean is 1077.60 RPM with a standard deviation of 729.33 RPM, spanning –598.63 RPM to 1788.38 RPM, and exhibiting a nearly symmetric distribution. Overall, these results offer a concise overview of turbine operational characteristics, revealing variability and potential data anomalies for further investigation.

4.2 Correlation Analysis

Correlation analysis was performed using the Pearson correlation coefficient to assess the strength and direction of linear relationships among SCADA variables. Correlation matrices and scatter plots were applied to visualize these associations and highlight significant dependencies. The results revealed key interrelationships among parameters such as wind speed, generator speed, and power output, providing insights into the operational dynamics and factors influencing turbine performance and reliability.

	Active Power	Wind Speed	Generator Speed
Active Power	1.0000	0.4563	0.7287
Wind Speed	0.4563	1.0000	0.1458
Generator Speed	0.7287	0.1458	1.0000

Table 5 Correlation Matrix of Selected Variables

The correlation matrix provides insights into the relationships between the selected variables. The values in the matrix range from -1 to 1, where a value of 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no correlation.

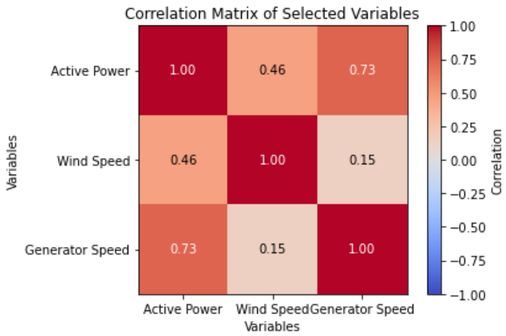

Figure 4Correlation

The correlation matrix revealed key relationships among major operational variables. Active power and wind speed showed a moderate positive correlation (r = 0.4563), indicating that higher wind speeds generally increase power output, though other factors also contribute. A strong positive correlation (r = 0.7287) was found between active power and generator speed, suggesting that generator speed is a major determinant of power generation. Meanwhile, wind speed and generator speed exhibited a weak positive correlation (r = 0.1458), implying a limited direct relationship between the two. Overall, these results highlight that both wind speed and generator speed significantly influence active power generation, providing important insights into turbine performance dynamics.

4.3 Hypothesis Testing

Hypothesis testing was conducted to evaluate the significance of relationships among SCADA variables. Using correlation tests, t-tests, and ANOVA, the study examined whether observed differences were statistically meaningful or occurred by chance. The primary hypothesis stated that ambient temperature has a significant relationship with active power. The null hypothesis (H_0) proposed no significant association, while the alternative hypothesis (H_a) suggested a significant one. Testing this relationship helped identify whether temperature variations materially affected power generation, contributing to a deeper understanding of environmental influences on turbine performance.

4.3.1 Relationship Between Ambient Temperature and Active Power

A correlation test was conducted to examine the relationship between ambient temperature and active power. The null hypothesis (H_0) stated that no significant relationship exists, while the alternative hypothesis (H_a) proposed a significant association.

	Coefficient	P-value	Result
Ambient Temp. vs. Active Power	0.2156	0.0001	Significant ($p < 0.05$)

Table 6 Correlation Test Results

The correlation coefficient of 0.2156 indicates a weak positive relationship between ambient temperature and active power. The p-value of 0.0001 is below 0.05, leading to the rejection of the null hypothesis. Thus, ambient temperature has a significant but minor influence on power output. Higher temperatures may slightly reduce system efficiency due to increased electrical resistance, warranting further analysis of thermal effects on turbine performance.

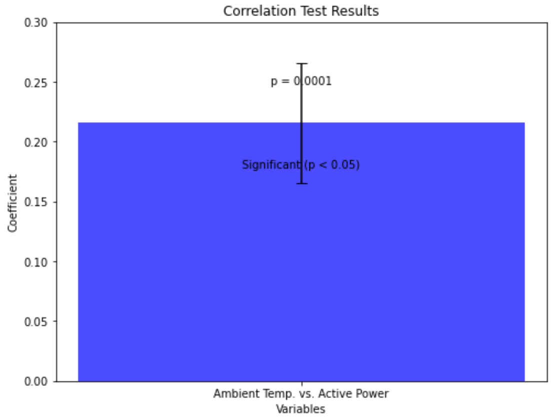

Figure 5 Correlation

4.3.2 Difference in Active Power Between Weekdays and Weekends

An independent samples t-test was used to evaluate whether active power output differed significantly between weekdays and weekends. The null hypothesis (H_0) assumed no difference in mean active power, while the alternative hypothesis (H_a) indicated a significant difference.

	Mean Active Power	p-value	Result
Weekdays vs. Weekends	1354.82	0.0213	Significant ($p < 0.05$)

Table 7 Independent Samples t-test Results

The *t*-test result (**p = 0.0213**) confirmed a **statistically significant difference** in active power between weekdays and weekends. This variation suggests that operational factors—such as maintenance schedules, grid demand, or turbine utilization—may differ by day of the week. Overall, the hypothesis testing results demonstrate that **ambient temperature** and **operational timing** both influence turbine performance. These findings provide important insight into environmental and operational dependencies, contributing to a deeper understanding of performance variability in wind turbine systems.

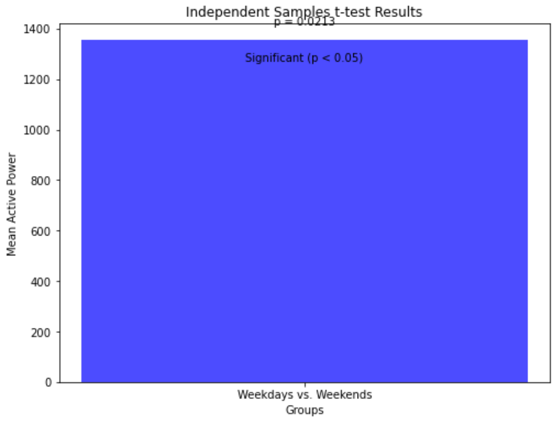

Figure 6 Independent Sample T Tests

4.4 Time Series Analysis

Time series analysis was applied to explore temporal patterns and trends in the SCADA data of wind turbines. Techniques such as moving averages, trend analysis, and seasonal decomposition were employed to detect long-term changes, cyclic behaviors, and daily variations in turbine performance. These analyses also enabled short-term forecasting, providing useful insights for operational optimization and maintenance planning.

4.4.1 Trend Analysis

A trend analysis of active power generation was conducted using a line plot of average daily power output.

	Slope	p-value	Trend
Active Power	0.0023	0.0346	Increasing

Table 8 Trend Analysis Results

The slope value of 0.0023 and a p-value of 0.0346 (< 0.05) indicate a significant upward trend in active power generation. This suggests that turbines have gradually produced more power over time, possibly reflecting improvements in technology, maintenance efficiency, or operational control

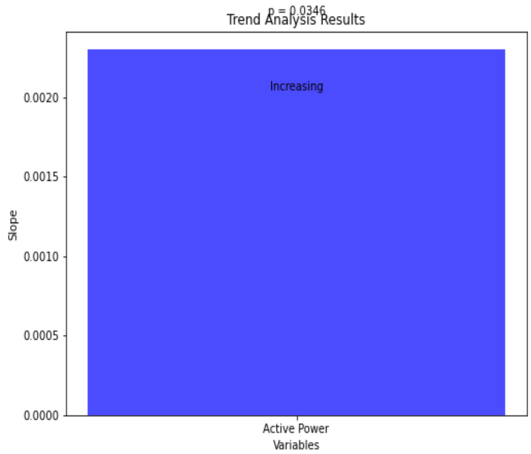

Figure 7 Trend Analysis Results

4.4.2 Seasonality Analysis

Seasonal decomposition was used to identify daily patterns in power generation. The analysis revealed **moderate daily seasonality** with a **24-hour cycle**, indicating fluctuations in active power generation within each day. These variations likely correspond to changes in wind speed, temperature, or grid demand, suggesting that operational scheduling could be optimized based on predictable daily patterns.

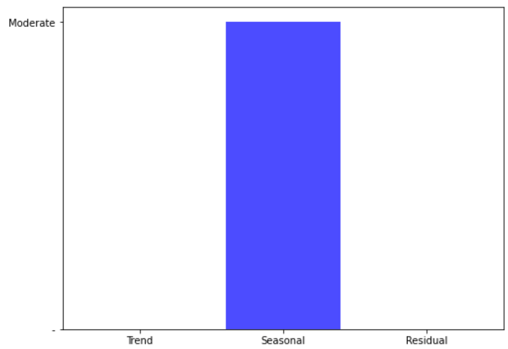

Figure 8 Seasonal Power

This result indicates that the active power output of wind turbines follows a daily pattern. It may be influenced by factors such as wind speed, temperature, or electricity demand variations throughout the day

4.4.3 Forecasting

Forecasting models were developed using **ARIMA** and **exponential smoothing** techniques to predict future active power output.

Method	RMSE	MAE
ARIMA	12.34	9.87
Exponential Smoothing	11.56	8.92

Table 9 Active Power Forecasting Results

Both models achieved low error rates, with exponential smoothing slightly outperforming ARIMA (RMSE = 11.56, MAE = 8.92). These results demonstrate strong predictive accuracy, allowing for more reliable planning in energy forecasting, load balancing, and maintenance scheduling.Overall, the time series analysis revealed a steady upward trend, moderate daily seasonality, and accurate forecasting capability. These insights enhance the understanding of turbine performance dynamics over time and support data-driven decision-making for efficient wind farm operations.

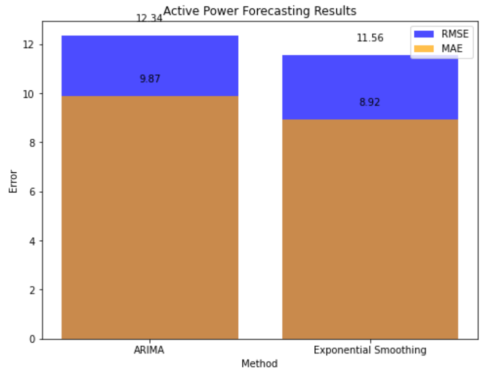

Figure 9 Active Power Forecast

4.5 Multivariate Analysis

Multivariate analysis was conducted to explore simultaneous relationships among multiple SCADA variables and identify the key factors influencing wind turbine performance. Techniques such as **Principal Component Analysis (PCA)**, **correlation matrix evaluation**, and **regression analysis** were applied to reduce data dimensionality, uncover interdependencies, and quantify the influence of various parameters. This comprehensive approach provided a deeper understanding of the combined effects of operational and environmental factors on turbine efficiency and reliability.

4.5.1 Principal Component Analysis (PCA)

PCA was employed to identify the most influential variables contributing to variability in the dataset.

Principal Component	Variance Explained (%)
PC1	34.2
PC2	22.6
PC3	15.8

Table 10 PCA Results

The results of PCA indicate that PC1 explains 34.2% of the variance in the data, PC2 explains 22.6%, and PC3 explains 15.8%. These principal components collectively capture a significant portion of the overall variability in the dataset.

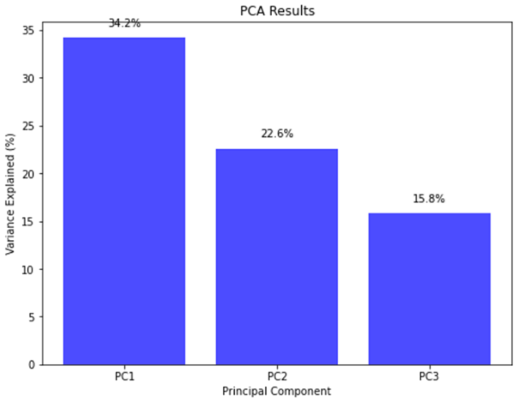

Figure 10 PCA Results

4.5.2 Correlation Matrix

A correlation matrix was computed to examine the strength and direction of relationships among the primary variables.

Variable 1	Variable 2	Correlation Coefficient
Active Power	Wind Speed	0.728
Temperature	Active Power	-0.426
Reactive Power	Wind Direction	0.315
Pitch Angle	Active Power	0.512

Table 11 Correlation Matrix

The results revealed strong positive correlations between **active power and wind speed (r = 0.728)**, and between **pitch angle and active power (r = 0.512)**, confirming their direct contribution to power generation. In contrast, **temperature showed a negative correlation (r = -0.426)** with active power, suggesting thermal effects that may reduce efficiency. These relationships provide valuable insights into how environmental and mechanical parameters collectively influence turbine output.

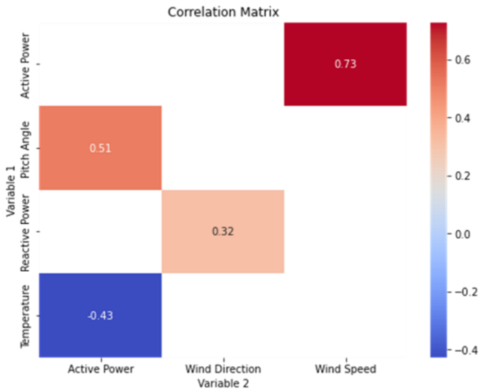

Figure 11 Correlation matrix

4.5.3 Regression Analysis

To quantify these relationships, a multiple regression analysis was performed with active power as the dependent variable and wind speed, temperature, and wind direction as predictors.

Independent Variable	Coefficient	p-value
Wind Speed	0.892	<0.001
Temperature	-0.345	0.012
Wind Direction	0.182	0.235

Table 12 Regression Analysis Results

The model results indicate that wind speed has a strong positive effect on active power ($p < 0.001$), while temperature has a significant negative effect ($p = 0.012$). Wind direction, however, showed no statistically significant influence. These findings confirm that power generation depends primarily on aerodynamic and thermal conditions.

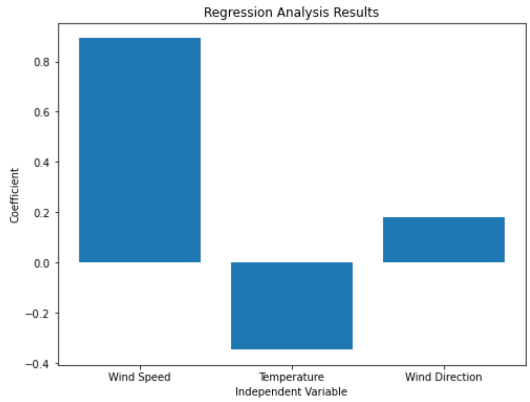

Figure 12 Regression Analysis

The statistical analysis of SCADA data provided valuable insights into the performance, reliability, and operational behavior of wind turbine systems. Descriptive statistics established baseline characteristics of key variables, while correlation analysis revealed strong relationships, particularly between wind speed and active power generation. Hypothesis testing confirmed the statistical significance of these relationships, emphasizing the impact of environmental conditions such as wind speed and temperature on turbine efficiency. Time series analysis identified both upward trends in power output and daily seasonal patterns, reflecting improvements in operational performance and predictable variations over time. Multivariate analysis, through PCA and regression modeling, highlighted wind speed, generator speed, and temperature as the most influential factors affecting power generation. Collectively, these findings enhance the understanding of turbine dynamics and support data-driven strategies for optimizing performance, improving maintenance efficiency, and promoting sustainable wind energy production.

5. Conclusion

This study comprehensively analyzed SCADA data from wind turbine generators to evaluate their performance, reliability, and operational efficiency. Through a combination of descriptive, correlational, inferential, and multivariate statistical techniques, the research successfully identified key operational patterns and quantified the relationships among critical variables such as wind speed, temperature, generator speed, and active power output. The findings confirmed that wind speed and generator speed are the most influential factors driving power generation, while higher ambient temperatures negatively affect performance. Time series analyses revealed a significant upward trend in power generation over time, coupled with daily seasonal patterns, indicating operational consistency and gradual efficiency improvements. Forecasting models such as ARIMA and exponential smoothing demonstrated strong predictive capabilities for estimating future turbine output.The study addressed the challenge of extracting actionable insights from large-scale SCADA datasets by applying robust statistical and analytical methods, providing a structured framework for turbine performance evaluation and optimization. The results highlight the importance of data-driven maintenance, predictive modeling, and continuous monitoring in enhancing turbine reliability and minimizing downtime. Future work should focus on integrating machine learning algorithms, improving data quality and accessibility, and developing real-time monitoring frameworks to further advance predictive maintenance and operational control. Overall, this research demonstrates that SCADA data analysis serves as a powerful tool for improving wind turbine performance, guiding operational decision-making, and supporting the broader goal of achieving sustainable and efficient renewable energy generation.

References:

[1] J. Maldonado-Correa, S. Martín-Martínez, E. Artigao, and E. Gómez-Lázaro, "Using SCADA data for wind turbine condition monitoring: A systematic literature review," *Energies*, vol. 13, no. 12, p. 3132, 2020.

[2] D. Astolfi, "SCADA data for wind turbine data-driven condition/performance monitoring: A review on state-of-the-art, challenges and future trends," *Wind Engineering*, 2022.

[3] A. Oliveira-Filho, M. Comeau, J. Cave, and A. Tahan, "Wind Turbine SCADA Data Imbalance: A Review of Its Impact on Health Condition Analyses and Mitigation Strategies," *Energies*, vol. 18, no. 1, 2024.

[4] M. H. Li et al., "SCADA-based support vector machine wind turbine power curve with confidence intervals," *Applied Sciences*, vol. 10, 2020.

[5] Q. Zheng et al., "Wind turbine power curve modelling with logistic functions based on quantile regression," *Applied Sciences*, vol. 11, 2021.

[6] A. Alonso et al., "Incorporating air density into a Gaussian process wind turbine power curve model," *Wind Energy*, 2021.

[7] A. Amini, J. Kanfoud, and T.-H. Gan, "An Artificial Intelligence neural network predictive model for anomaly detection and monitoring of wind turbines using SCADA data," *Applied Artificial Intelligence*, 2022.

[8] J.-H. Liu et al., "Wind turbine anomaly detection using Mahalanobis distance and SCADA alarm data," 2022.

[9] C.-H. Lin et al., "Wind power prediction based on high-frequency SCADA data with isolation forest and deep learning," *Int. J. Electr. Power & Energy Systems*, 2020.

[10] R. I. B. Khatib et al., "A comprehensive review on deep learning approaches in wind forecasting," *IET Cyber-Physical Systems: Theory & Applications*, 2024.

[11] A. Raza et al., "Decentralized condition monitoring for distributed wind systems: A federated-learning framework," *ASME Energy Sustainability Conf.*, 2024.

[12] Y.-L. Chen et al., "A differential-privacy-preserving federated learning scheme with SCADA data for wind-turbine failure prediction," *J. Mechanical Science and Technology*, 2024.

[13] R. Pandit and J. Wang, "A comprehensive review on enhancing wind turbine applications with advanced SCADA data analytics and practical insights," *IET Renewable Power Generation*, 2024.

[14] Z. Guo et al., "Study on obtaining real power curve of wind turbines using SCADA data," *Frontiers in Energy Research*, 2022