FEYZİYE SCHOOLS FOUNDATION
IŞIK UNIVERSITY

Statistical Analysis of the Relationship Between Wind Speed and Capacity Factor at Kelmarsh Wind Farm

Instructor: Dr. Habibe AKTAY

Bilgesu ÇAKIR, 22MISY1016

Can Deniz KOÇAK, 22MISY1026

Nov, 2025

Introduction and Literature Review

Identification of Specific Problem Area

The transition to renewable energy requires precise monitoring of power generation efficiency. A critical challenge in wind energy management is understanding the variability of power output in relation to meteorological conditions. Specifically, operators must understand how wind speed, the primary kinetic driver, translates into normalized efficiency (Capacity Factor) rather than just raw power output, to assess the reliability of specific wind farm sites.

Prevalence and Scope of Problem

The inability to accurately characterize the performance curve of wind turbines can lead to grid instability and inefficient energy dispatch. This problem is global in scope but requires site-specific analysis, as local wind profiles differ significantly. This study focuses on the operational scope of the Kelmarsh Wind Farm in Northamptonshire, UK. 52°24'05"N 0°56'35"W

Fig. 1: Kelmarsh Wind Farm

Overview of Existing Knowledge (Literature Review)

Previous research has established that wind speed is the dominant variable affecting energy production.

- Akdağ and Dinler (2009) utilized the Weibull function to model wind speed distributions, demonstrating that statistical modeling of wind speed is a prerequisite for accurate energy estimation.

- Kusiak, Zheng, and Song (2009) applied data mining techniques to Supervisory Control and Data Acquisition (SCADA) data. Their findings confirmed a strong, positive, and non-linear relationship between wind speed and generated power. They identified that the relationship typically follows a cubic power curve before saturating at the turbine's rated capacity.
- Bouabdallaoui et al. (2025) conducted a specific multi-temporal forecasting study on the Kelmarsh Wind Farm using data from 2016 to 2021. Using methods such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN), they found a Pearson correlation coefficient (r) of 0.93 between wind speed and power output, confirming the strength of this relationship. Conversely, they noted a very low correlation ($r = 0.08$) between wind direction and power output.

Critique of Previous Research

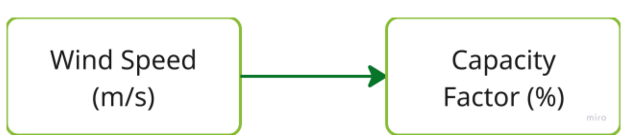
While Bouabdallaoui et al. (2025) provided robust forecasting models, their study focused on complex Artificial Intelligence (AI) predictions for future grid planning. There is a tendency in the literature to focus on “black box” predictive algorithms rather than fundamental descriptive statistical profiling of the site's operational efficiency (Capacity Factor). Furthermore, many studies analyze raw power (kW), which makes it difficult to compare performance across turbines with different rated capacities.

Gap in the Literature

The existing literature on the Kelmarsh site focuses heavily on complex machine learning forecasting. There is a lack of simplified, descriptive statistical analysis that focuses specifically on Capacity Factor as a standardized dependent variable. This study addresses this gap by applying fundamental descriptive statistics to a longitudinal dataset (2016-2024) to characterize the efficiency profile of the site without the complexity of AI modeling.

Purpose of Study and Research Questions

The purpose of this study is to statistically quantify the strength and shape of the relationship between wind speed and wind energy efficiency.



Research Question: To what extent does wind speed explain the variation in the Capacity Factor of turbines at the Kelmarsh Wind Farm?

Hypothesis 1 (H1): Wind speed has a strong, positive effect on the Capacity Factor.

Hypothesis 2 (H2): The relationship between wind speed and Capacity Factor is non-linear, following a sigmoid distribution where efficiency saturates at rated wind speeds.

Methodology

Design

Type of Design: This study utilizes a Quantitative research design. The rationale is that the variables of interest (Wind Speed and Capacity Factor) are numerical (Ratio scale) and derived from objective measurement systems, requiring statistical analysis rather than qualitative interpretation.

Specific Design: The study employs a Descriptive and Correlational Design. The rationale is that the goal is to summarize the characteristics of the dataset (mean, variance, dispersion) and to test the strength of the association between two variables using bivariate analysis.

Specific Research Methods: The method used is Secondary Data Analysis. The rationale is that high-frequency historical operational data is already available via the Zenodo repository, eliminating the need for expensive and time-consuming primary data collection.

Sample and Procedures

1. Population and Sample:
    a. Population: The theoretical energy generation performance of all wind turbines at the Kelmarsh wind farm over their entire operational lifespan.

    b. Sample: The sample consists of historical SCADA data records spanning from 2016 to 2024 for the six turbines located at the Kelmarsh site.

2. Selection Procedures: The sample was selected using Non-probability Convenience Sampling. The data was chosen because it is a publicly available, open-access dataset

provided by Cubico Sustainable Investments. Once selected, the data will be filtered to remove timestamps containing sensor errors or maintenance periods (where power output is zero despite wind presence).

Measurement

1. Instruments and Procedures:
    a. Instrument: The data was originally collected using the site's SCADA system.

    b. Independent Variable (Wind Speed): Measured in meters per second (m/s) using anemometers mounted on the turbine nacelles. This is a continuous, ratio-level variable.

    c. Dependent Variable (Capacity Factor): This variable measures efficiency. It is calculated using the formula:

    $$\frac{\text{Actual Power Output}}{\text{Rated Peak Power}} x\ 100$$

    This is a continuous, ratio-level variable.

    d. Validity and Reliability: SCADA systems are the industry standard for wind farm monitoring. The sensors are calibrated instruments, ensuring high reliability. To ensure validity, we will follow the preprocessing steps outlined by Bouabdallaoui et al. (2025), removing non-physical values (e.g., negative power output).

Analysis Plan

The data analysis will proceed in the following steps using Microsoft Excel and Python:

1. Data Cleaning: Remove outliers and rows with missing values or negative readings.
2. Descriptive Statistics: Calculate the Mean, Median, Mode, Standard Deviation, and Coefficient of Variation (CV) for both Wind Speed and Capacity Factor to summarize the central tendency and dispersion.
3. Visualization: Generate Histograms to analyze the distribution shape (skewness) of the variables.
4. Bivariate Analysis: Create Scatter Plots to visualize the relationship between Wind Speed (X-axis) and Capacity Factor (Y-axis). Calculate the Pearson Correlation Coefficient (r) and the Coefficient of Determination (R^2) to test the hypotheses.

Limitations

Sampling Limitation: Convenience sampling limits the generalizability of the findings. Results from Kelmarsh (an onshore site in the UK) may not apply to offshore sites or different climatic regions.

Variable Limitation: The study focuses only on wind speed. Other variables such as air density, temperature, and blade degradation over time are excluded from this specific analysis, which may leave some variance in the Capacity Factor unexplained. Although, more variables may be included in the research later.

References

Akdağ, S. A., & Dinler, A. (2009). A new method to estimate Weibull parameters for wind energy applications. *Energy Conversion and Management*, *50*(7), 1761–1766. https://doi.org/10.1016/j.enconman.2009.03.020

Bouabdallaoui, D., Haidi, T., Derri, M., Hbiak, I., & El Jaadi, M. (2025). Multi-temporal forecasting of wind energy production using artificial intelligence models. *International Journal of Renewable Energy Development*, *14*(3), 505–517. https://doi.org/10.61435/ijred.2025.61086

Carta, J. A., Velázquez, S., & Cabrera, P. (2013). A review of measure-correlate-predict (Mcp) methods used to estimate long-term wind characteristics at a target site. *Renewable and Sustainable Energy Reviews*, *27*, 362–400. https://doi.org/10.1016/j.rser.2013.07.004

Kusiak, A., Zheng, H., & Song, Z. (2009). Wind farm power prediction: A data-mining approach. *Wind Energy*, *12*(3), 275–293. https://doi.org/10.1002/we.295

Plumley, C., & Takeuchi, R. (2025). *Kelmarsh wind farm data* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.16807551