A Predictive Modeling and Parametric Optimization for Wind Energy Power Plant

R. Hassan[1], M. Mahboob[2], M. Ali[2], F.Ahmad[3]
[1]Department of Electrical Engineering and Engineering Management, College of Engineering and Sciences, Institute of Business Management Sindh, Karachi, Pakistan
rabia.hassan@iobm.edu.pk
[2]Department of Mechanical Engineering, University of Engineering and Technology, Taxila, Pakistan
Muhammad.Mahboob14140@gmail.com
[3]Industrial Management Department, Government College University Faisalabad
fayyazahmad@gcuf.edu.pk

ABSTRACT

Energy demand forecasting and modelling are rationalized and supported by the substantial amount of energy usage in the various sectors. Although, there are many methods to develop energy models, but statistical modeling is the easiest and accurate approach to get the future information. Simple Linear regression analysis is the most simple and precious technique among the other statistical techniques. In this study the output power of a wind turbine is determined and compare with the real data gathered from HAWA Power plant located in Sindh, Pakistan. The most important regressor which effects the seasonal power generation of the power plant is modeled. For this purpose, the independent variables are Nacelle temperature (X1), Gear box temperature (X2) and Wind speed (X3) similarly, the Power output is the dependent variable (Y). The model reliability of power output forecasting is a major concern. The wind generator is a center of instability from a system perspective, and the system must prepare for it. The output power of wind power plant is highly influenced direction, wind speed and climatic condition. The results indicate that for the season of summer the R^2 value is 0.983 whereas for the season of winter the R^2 value is 0.883. The Skewness and Kurtosis distribution also verify for the independent variable normality and the values lie between -1.0 to +1.0 for both seasonal models. This study's structure and material have been developed into a comprehensive asset that may serve as the basis for more research in this captivating field of study.

Keywords: Power output, Regression, Nacelle Temperature, Prediction

1. INTRODUCTION

Renewable energy as wind and solar power is gaining significant attention because of the variations in the prices of coal and oil in the global market, and perhaps even the injurious environmental issues generating from the fossil fuel power generation process [1]. The demand for energy is rising significantly [2]. This growing requirement may be met through renewable and energy alternatives that are assumed as a feasible green source of energy for clean operations and healthy living [3]. Wind power is indeed one among the fastest rising zero-emission clean energy source [4]. Wind turbines generate renewable energy and does not generate emissions like other energy sources [5]. Modern wind turbines are powerful, reliable and generate energy at an affordable rate.

Highest potential of wind in Pakistan is in Sindh and some areas of Balochistan as shown in figure 1. According to Ministry of Water and Power, Pakistan has a 1,046 km coastline in the South (Sindh and

Balochistan), nonetheless, the vast majority of wind energy installations are now ongoing at Hyderabad and Gharo-Keti Bander wind corridor. Pakistan's total wind power capacity is projected to be 349,315 MW [6]. The Coastal zone of Sindh is enriched with a wind path which is 180 km long and 60 km wide [7]. Consequently, this corridor has the capability to produce electricity of 50,000 MW utilizing wind energy [8]. There are supplemental wind locations situated in Gilgit Baltistan and Khyber Pakhtunkhwa. When installed, small wind turbines would be capable of electrifying rural areas that cannot be linked to a grid system. Wind energy is projected to be able to power around 5,000 villages in Gilgit Baltistan, Baluchistan and Sindh [9]. Wind speeds of approximately 5-7 m/s maintain at a 50 m of height in the Sindh coastal regions and several other Pakistan's North-West areas. Various locations in Pakistan have a Wind Power Generation capacity approximately more than 25 percent, which is appropriate for the insertion of higher capacity commercial wind farm in Global standards [10]. The overall supply of wind power to the national grid only 6 MW during 2011o to 2012, which improved dramatically to 786 MW in 2015 to 2016 [11]. The wind power capacity of 280 MW was added to the national grid system in 2017 to 2018 which include 50 MW Artistic Wind Power, 50 MW Three Gorges Second Wind Farm, 50 MW Jhampir Power, 50 MW Hawa Energy, 50 MW Three Gorges Third Wind Farm and 30 MW Tapal Wind Energy [12].

Forecasting and modeling of wind power generation is an important and relevant concern. The wind generator is a center of instability from a system perspective, and the system must prepare for it. The system operator requires a unit guarantee for one day advance, and these wind power plants are required to provide reverse supply. So, enhancing the Efficiency of modeling and forecasting methods would have an overall impact on the system. Many related works are conducted to predict the power like to recommend a regression tree for predicting the power output of a 1.5 MW wind turbine [13].

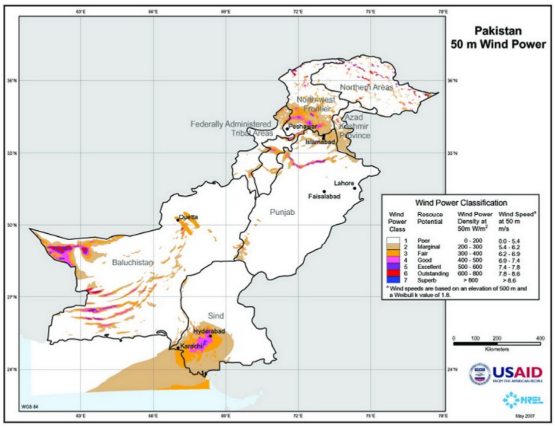


Figure 1 Wind Power Map of Pakistan

They use the datasets gathered from industrial simulations of wind and turbine models to train the algorithm. In linear regression, k-nearest neisghbor, and support vector regression are used for evaluating forecasts for turbines and whole wind farms with a period of thirty minutes. Their simulations use data from previous power wind interpretations to get forecasts [14]. The forecasting of power is reliant either on historic wind time series or numerical weather prediction (NWP) values. Physical models, traditional predictive techniques, and, more recently, data-driven or learning techniques are used to model power production forecasting.

Regression analysis is often used in many fields of study including biology, social sciences, medicines, econometric and engineering etc. [15]. If there is a correlation between two or several variables that cannot be clarified by any law, it is usually considered statistical. In either situation, one or several independent variables can influence a variable retort. Multiple linear regression analysis (MLRA) is a technique which is statistical for evaluating the correlation among a particular dependent variable with two or several independent variables [16]. In engineering, now a days majority of energy related regression analysis surveys are conducted to get the prediction of wind turbine power generation [17]-[22]. In addition, regression analysis is frequently utilized to forecast properties of wind including its direction and speed [23], [24]. The significant proportion of research in the field emphasis on regression analysis. However, neither of them is related to the estimate of seasonal time series plant data emphasizes on the nacelle temperature impact of power generation wind turbine.

This paper emphasizes on the estimation of wind turbine power generation plant based on literature review, incorporates the MLRA of an operational 50 MW wind plant in Jhampir Sindh, Pakistan, utilizing the software SPSS. The validated power plant efficiency model was used to conduct a power generation prediction of wind turbine with variables of wind speed, nacelle temperature, and gearbox temperature. Ambient temperature neglected, due to its highest correlation with nacelle temperature.

2. Dataset and Methodology

All the data employed in this study came from the SCADA system of wind farm located in Jhampir, Sindh, Pakistan. It is a 50MW wind farm generating sustainable and green energy for Pakistan and gets operational in 2018. The data of wind speed (m/s), power output (kW), ambient temperature (℃), nacelle Temperature (℃), and gearbox temperature (℃) was collected from SCADA covering the periods June-July 2019 and December-2019 to January 2020. The goal was to use regression analysis to evaluate the properties of the independent variables on power generation, and the ambient temperature was neglected as it has a correlation with nacelle temperature. As a result, just the impact of wind speed, nacelle temperature, and gearbox temperature for net power have been investigated as the ambient temperature have greatest correlation with nacelle temperature.

The figure 2 illustrates the methodology in the first phase real time data was collected from the operational wind power plant after that the data arranged and sort into a meaningful sequence to make it easy for better understanding, analyzing and visualization. The dependent and independent variables were defined and the data import into the SPSS software, the multiple linear regression analysis was performed to identify the VIFs values, and significance level. If the VIFs value is greater than five and there is more than one value exists in the model, then removed the variable with a highest VIF value. If the VIFs value less than 5 run best subsets to obtain best models for the prediction of power generation.

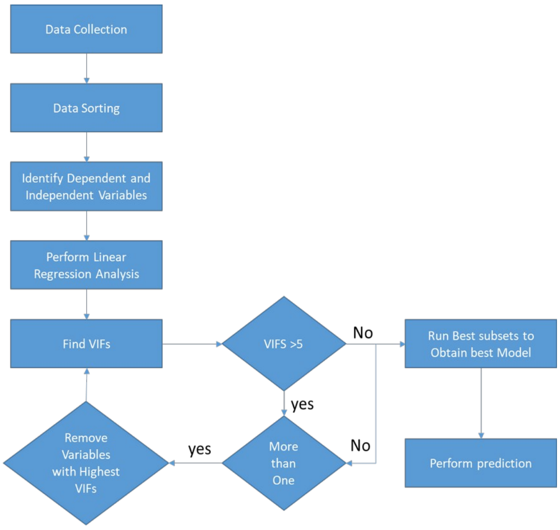


Figure 2 Regression Analysis Pathway

3. Dataset and Methodology

The wind turbine power output is the dependent variable in this analysis, whereas the wind speed, nacelle temperature, and gearbox temperature are the independent variables. The following is a method for a multiple linear regression model of power generation.

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & X_{12} & X_{13} & X_{14} & \cdots & X_{1N} \\ 1 & X_{22} & X_{23} & X_{24} & \cdots & X_{2N} \\ 1 & X_{32} & X_{33} & X_{34} & \cdots & X_{3N} \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & X_{N2} & X_{N3} & X_{N4} & \cdots & X_{NK} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_N \end{bmatrix} \quad \textbf{(1)}$$

$$Y_i = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_N X_N + \varepsilon_i \quad \textbf{(2)}$$

In the above model the dependent variable is Y, the independent variables are X's, the regression constant is β_o, the partial regression coefficients are β_1, β_2, $\beta_3 \ldots \beta_N$ whereas, the error term is ε_i. In the regression analyses, the least squares approach is used to measure $\beta' s$. This approach efficiently evaluates error term and ß coefficients, which are complicated and variable for each Y prediction. As a result, the least square estimation techniques of ß's which are b's are determined rather than real ß coefficients. The error can be calculated by equation 3.

$$\varepsilon = Y - X\beta \quad \textbf{(3)}$$

The square of a matrix is obtained by multiplying it by its own transpose, and the error sum of squares can be determined using equation 4.

$$S_\beta = \varepsilon \varepsilon' = (Y - X\beta)(Y - X\beta)' = Y' Y - 2\beta' X' Y + \beta' X' X\beta \quad \textbf{(4)}$$

The least square estimators of ß's (b's) are determined by equation 5. while the derivation of error total of squares is collected according to ß.

$$b = (X' X)^{-1} X' Y \quad \textbf{(5)}$$

The basic presumptions of traditional linear regression should be tested when constructing a prediction model in MLRA. Multiple linear regression analysis (MLRA) is based on six presumptions which include zero mean, linearity, normality, multicollinearity, heteroscedasticity, regular distribution of errors, and autocorrelation [25].

Normality is the first assumption; this suggests that each of the independent variables should be distributed normally. Skewness and Kurtosis distributions can be used to verify the independent variables normality. The distribution results must be within -1.0 to +1.0. If the variable's distribution is not normal, transforming of the independent variable may be needed [15]. Errors are distributed anomaly and have a zero mean in the second assumption. The third presumption "linearity," depends on the independent and dependent variables' on linear relation. To determine if these factors have a linear relationship, hypothesis tests can be used. A statistical test is used to verify the coefficients of the regulated variables, and irrelevant variables can be omitted from the formula. The F-test is then used to determine whether the model is effective.

The fourth presumption is that all records must be heteroscedasticity free. Every error has same (constant) variance, and the regression is homoscedastic. In this case, analysis of residual must be examined. For time series records, the heteroscedasticity test is usually not needed.

Multicollinearity is the fifth presumption that should be tested in a study. When the regulated variables have a strong correlation degree of multicollinearity happens. In a regression model, the Variance Inflation Factor (VIF) is being employed to calculate the independent variables and multicollinearity degree [26].

$$VIF(b_i) = \frac{1}{1-R^2} \tag{6}$$

If the VIF value is less than 5, multicollinearity is not really an issue. Multicollinearity is significant if VIF is greater than 5. When the VIF value exceeds than10, it becomes more serious. Eventually, the autocorrelation can be tested using the Durbin-Watson statistic (d). Durbin-Watson is a statistical test that is intended to identify the existence of autocorrelation between regression errors [27].

$$d = \frac{\sum_{u=2}^{N} (\varepsilon_u - \varepsilon_{u-1})^2}{\sum_{u=2}^{N} \varepsilon_u^{\,2}} \tag{7}$$

The d value is typically in the range of 0 to 4. There is no autocorrelation when d = 2. There is proof of a strong serial correlation if the d<2. If d > 2, consecutive error terms have significantly different values, implying that they are negatively connected. If the model has an autocorrelation, the Prais-Winsten method and Orcutt-Cochran procedure can be used to eradicate it [28].

4. Result and Discussion

In this analysis by utilizing all independent variables, linear models were developed, evaluated, through a statistical package software (SPSS) to analyze the results. Skewness and Kurtosis tests were used to find out the normality of every variable that is independent in relation to the dependent variable which described in the table 1 and 2, whereas the figure 3 and 4 shows probability plots and indicates that the data is normally distributed as required for regression analysis.

Table 1 Independent variables Skewness and kurtosis values (June-July)

Independent Variables	Skewness	Kurtosis
Wind Speed	0.741	-0.404
Nacelle Temperature	-0.068	0.998
Gearbox Temperature	-0.636	-0.081

Table 2 Independent variables Skewness and kurtosis values (December-January)

Independent Variables	Skewness	Kurtosis
Wind Speed	0.472	-0.658
Nacelle Temperature	-0.109	-0.790
Gearbox Temperature	0.077	-0.728

Table 1 and 2 indicates that all independent variables are satisfying the normality leading to Kurtosis and Skewness tests except nacelle temperature Kurtosis test result for the month of June-July exceeds a little from the defined values. It should be in between the define limits -1.0 to +1.0.

Table 3 represents the R and Adjusted R2 value for the month of June and July that the Model is 99.2% is statistically. Significance and fit. The p value is less than 0.05 that indicates that null hypothesis is rejected.

The Durbin Watson value of the model is 1.244 that shows there is non-autocorrelation among the variables.

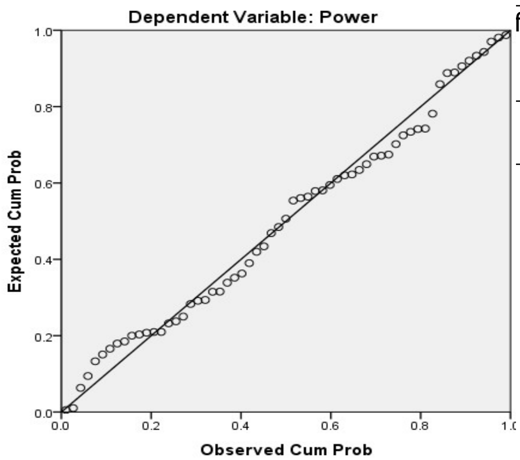
Table 4 represents the R and Adjusted R^2 value for the month of June and July that the Model is 94% is statistically significant and fit. The p value is less than 0.05 that indicates that we reject the null hypothesis. The Durbin Watson value of the model is 1.244 that shows there is non-autocorrelation among the variables.

Figure 3 Normal P-P plot of Regression Standardized Residual for

Figure 4 Normal P-P plot of Regression Standardized Residual for month December-January

Table 3 Model summary for the month of June-July

Model Summar		
Mc	f	n- n
	R^2 Change	
	.983	4



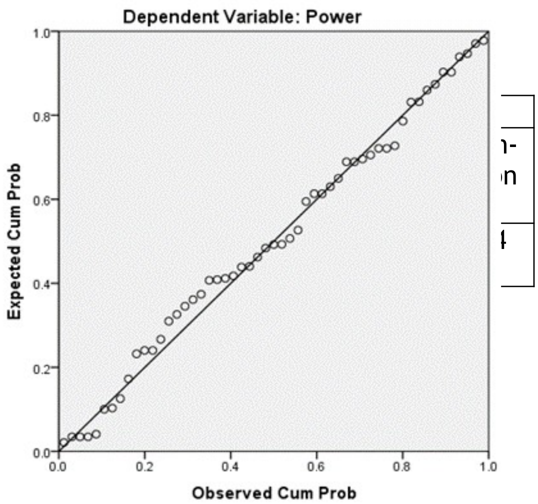


Table 4 Model summary for the month of December-January

Model Summary										
Model	R	R^2	Adjusted R^2	Std. Error of the Estimate	Change Statistics					Durbin-Watson
					R^2 Change	F Change	df1	df2	Sig. F Change	
1	.940	.883	.876	1824.11614	.883	123.687	3	49	.000	1.799

The significance level α=0.05, so there is sufficient proof to determine that predictors are efficient for estimating output power. Thus, the models are valuable. The VIF values for the two models are less than 5 which shows that among the independent variables there

are no multicollinearity exits. The below table 5 shows the relevant model's equations. The coefficients of determination for both models are 0.992 and 0.940, respectively.

Table 5 Model Equations

Model	Equation
June-July	Y= $-138660.91+1571.96\,W_s-28$
December -January	Y= $5395.94-159.27\,W_s-1263.60$

The figure 5 indicates that our finding error range between (-8% to +8%) hold nicely for these randomly chosen points

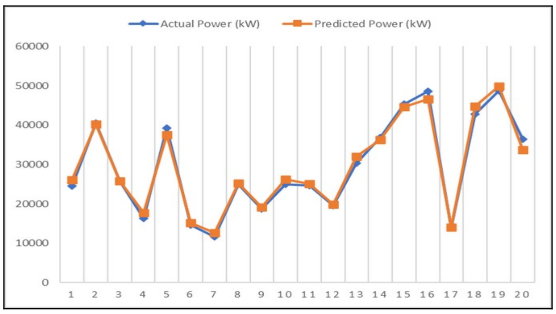


Figure 5 Actual and Predicted Power output at random 20 data points for month of Jun-Jul

The figure 6 indicates that our finding error range between (-18% to +18%) hold nicely for these randomly chosen points.

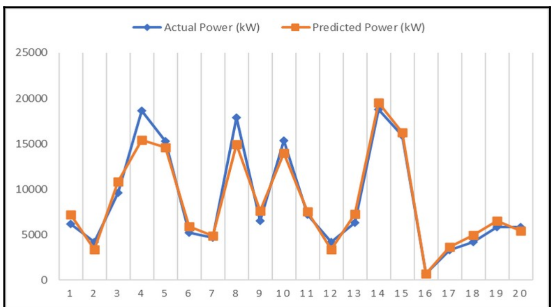
5. Conclusion

This study is about the forecasting of power output of wind turbines. Multiple regression analysis is used to analyze the seasonal power output of plant which located in Jhampir, Sindh, Pakistan. A Seasonal model for summers (June-July) and winters (Dec-Jan) was developed by using real time data of operational plant using three independent variables.

The results of data showed that errors and data are normally distributed. Furthermore, both models have no multi-collinearity between the independent variables. Moreover, the F-test at significance level α=0.05 result indicates that models are useful. However, the coefficients of determination for both models are 0.992 and 0.940, respectively. The graphical comparison is also identifying that the error and actual power generation are acceptable.

6. References

[1] **T. Mäkitie, H. E. Normann, T. M. Thune, and J. Sraml Gonzalez,** "The green flings: Norwegian oil



and gas industry's engagement in offshore wind power," Energy

[2] **A. K. Azad, M. M. K. Khan, T. Ahasan, and S. F. Ahmed**, "Energy Scenario: Production, Consumption and Prospect of Renewable Energy in Australia," *J. Power Energy Eng.*, vol. 02, no. 04, pp. 19–25, 2014, doi: 10.4236/jpee.2014.24004.

[3] **J. Li, P. Liu, and Z. Li**, "Optimal design and techno-economic analysis of a solar-wind-biomass off-grid hybrid power system for remote rural electrification: A case study of west China," *Energy*, vol. 208, p. 118387, 2020, doi: 10.1016/j.energy.2020.118387.

[4] **D. Gielen, F. Boshell, D. Saygin, M. D. Bazilian, N. Wagner, and R. Gorini**, "The role of renewable energy in the global energy transformation," *Energy Strateg. Rev.*, vol. 24, no. January, pp. 38–50, 2019, doi: 10.1016/j.esr.2019.01.006.

[5] **R. Saidur, N. A. Rahim, M. R. Islam, and K. H. Solangi**, "Environmental impact of wind energy," *Renew. Sustain. Energy Rev.*, vol. 15, no. 5, pp. 2423–2430, 2011, doi: 10.1016/j.rser.2011.02.024.

[6] **W. Iqbal *et al.***, "Assessment of wind energy potential for the production of renewable hydrogen in Sindh Province of Pakistan," *Processes*, vol. 7, no. 4, 2019, doi: 10.3390/pr7040196.

[7] **Y. A. Solangi, Q. Tan, M. W. A. Khan, N. H. Mirjat, and I. Ahmed**, "The selection of wind power project location in the Southeastern Corridor of Pakistan: A factor analysis, AHP, and fuzzy-TOPSIS application," *Energies*, vol. 11, no. 8, 2018, doi: 10.3390/en11081940.

[8] **P. Of *et al.***, *Artistic wind power (private) limited generation license application*, vol. 92, no. 4. 2006.

[9] **M. S. Khalil, "Renewable Energy in Pakistan**: Status and Trends," *Altern. Energy Dev. Board*, 2004.

[10] **Z. Ullah, S. M. Ali, I. Khan, F. Wahab, M. Ellahi, and B. Khan**, "Major Prospects of Wind Energy in Pakistan," in *2020 International Conference on Engineering and Emerging Technologies, ICEET 2020*, 2020, no. April, pp. 1–6, doi: 10.1109/ICEET48479.2020.9048201.

[11] **S. Bilal, Hazrat, Siwar Chamhuri, Mokhtar B.M, Ahmad**, "Recent development and sustainability of the wind power sector in Pakistan," *Int. J. Biomass Renewables*, vol. 7, no. 1, pp. 24–34, 2018.

[12] "State of Industry Report 2018," *Natl. Electr. Power Regul. Auth.*, p. 231, 2018, [Online]. Available: http://library1.nida.ac.th/termpaper6/sd/2554/19755.pdf.

[13] **A. Clifton, L. Kilcher, J. K. Lundquist, and P. Fleming**, "Using machine learning to predict wind turbine power output," *Environ. Res. Lett.*, vol. 8, no. 2, 2013, doi: 10.1088/1748-9326/8/2/024009.

[14] **Shivani, K. S. Sandhu, and A. R. Nair**, "Machine Learning approach for Short Term Wind Speed Forecasting," 2019, doi: 10.1109/ICACCE46606.2019.9079959.

[15] I. M. M. Ghani and S. Ahmad, "Stepwise multiple regression method to forecast fish landing," in *Procedia - Social and Behavioral Sciences*, 2010, vol. 8, pp. 549–554, doi: 10.1016/j.sbspro.2010.12.076.

[16] **J. B. Franklin, T. Sathish, N. V. Vinithkumar, and R. Kirubagaran**, "A novel approach to predict chlorophyll-a in coastal-marine ecosystems using multiple linear regression and principal component scores," *Mar. Pollut. Bull.*, vol. 152, no. September 2019, p. 110902, 2020, doi: 10.1016/j.marpolbul.2020.110902.

[17] **M. Marčiukaitis, I. Žutautaitė, L. Martišauskas, B. Jokšas, G.** Gecevičius, and A. Sfetsos, "Non-linear regression model for wind turbine power curve," *Renew. Energy*, vol. 113, pp. 732–741, 2017, doi: 10.1016/j.renene.2017.06.039.

[18] **N. A. Treiber, J. Heinermann, and O. Kramer**, "Wind power prediction with machine learning," *Stud. Comput. Intell.*, vol. 645, pp. 13–29, 2016, doi: 10.1007/978-3-319-31858-5_2.

[19] **T. Ouyang, X. Zha, and L. Qin**, "A combined multivariate model for wind power prediction," *Energy Convers. Manag.*, vol. 144, pp. 361–373, 2017, doi: 10.1016/j.enconman.2017.04.077.

[20] **N. Amjady, F. Keynia, and H. Zareipour**, "Short-term wind power forecasting using ridgelet neural network," *Electr. Power Syst. Res.*, vol. 81, no. 12, pp. 2099–2107, 2011, doi: 10.1016/j.epsr.2011.08.007.

[21] **J. A. Carta, S. Velázquez, and J. M. Matías**, "Use of Bayesian networks classifiers for long-term mean wind turbine energy output estimation at a potential wind energy conversion site," *Energy Convers. Manag.*, vol. 52, no. 2, pp. 1137–1149, 2011, doi: 10.1016/j.enconman.2010.09.008.

[22] **H. Liu, J. Shi, and X. Qu**, "Empirical investigation on using wind speed volatility to estimate the operation probability and power output of wind turbines," *Energy Convers. Manag.*, vol. 67, pp. 8–17, 2013, doi: 10.1016/j.enconman.2012.10.016.

[23] **S. Salcedo-Sanz, E. G. Ortiz-García**, Á. M. Pérez-Bellido, A. Portilla-Figueras, and L. Prieto, "Short term wind speed prediction based on evolutionary support vector regression algorithms," *Expert Syst. Appl.*, vol. 38, no. 4, pp. 4052–4057, 2011, doi: 10.1016/j.eswa.2010.09.067.

[24] **H. Utsunomiya, F. Nagao, and I. Urakami,** "Regression analysis of local wind properties with local topographic factors," *J. Wind Eng. Ind. Aerodyn.*, vol. 74–76, pp. 175–187, 1998, doi: 10.1016/S0167-6105(98)00015-4.

[25] **S. G. Damodar N. Gujarati**, Dawn C. Porter, *Basic Econometrics*. 2009.

[26] **R. M. O'Brien**, "A caution regarding rules of thumb for variance inflation factors," *Qual. Quant.*, vol. 41, no. 5, pp. 673–690, 2007, doi: 10.1007/s11135-006-9018-6.

[27] **N. Drapper**, *Applied regression analysis*. 1980.

[28] **D. Cochrane and G. H. Orcutt**, "Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms," *J. Am. Stat. Assoc.*, vol. 44, no. 245, pp. 32–61, 1949, doi: 10.1080/01621459.1949.10483290.