

# A survey of applications of simple and multiple linear regression in wind power generation

**Abstract.** This paper presents the results of a survey on the application of simple and multiple linear regression in wind power generation research. Relevant publications were searched for, found, reviewed, and summarised. An increasing trend of publications on this topic was found. The main categories of publications forecasting of wind output power, forecasting of wind speed, and wind turbine generator temperature monitoring. The paper presents concise summaries of publications and details the references identified, all of this in one repository.

**Streszczenie.** W artykule przedstawiono wyniki badań ankietowych dotyczących zastosowania prostej i wielokrotnej regresji liniowej w badaniach energetyki wiatrowej. Odpowiednie publikacje zostały wyszukane, znalezione, zrecenzowane i podsumowane. Stwierdzono rosnący trend liczby publikacji na ten temat. Główne kategorie publikacji: prognozowanie mocy wiatru, prognozowanie prędkości wiatru oraz monitorowanie temperatury generatorów turbin wiatrowych. W pracy przedstawiono zwięzłe streszczenia publikacji i wyszczególnienie zidentyfikowanych pozycji literaturowych, a wszystko to w jednym repozytorium. (Przegląd zastosowań prostej i wielokrotnej regresji liniowej w energetyce wiatrowej)

**Keywords:** Forecasting; Simple and multiple linear regression; Temperature; Wind output power; Wind speed  
**Słowa kluczowe:** prognozowanie, regresja liniowa, generatory wiatrowe.

## Introduction

The present-day energy demands are largely met by using fossil fuels [1] which are non-renewable and generate large amounts of greenhouse gasses, which leads to global warming that result in climate change, often threatening the natural and man-made systems. This has catalyzed the search for alternative energy sources, such as renewable energy sources.

Among the renewable energy sources, wind power generation [2] has received significant interest, partly because it can be easily captured by wind generators of higher power capacity in comparison to other renewable energy sources. This has led to a phenomenal growth in globally installed wind power generation capacity, from 48 GW [3] in 2004 to around 747 GW in 2020.

In the research related to wind power generation, there may arise situations in which it may be beneficial to establish whether or not a relationship exists between some predictor variables and some response variable [4-5]. If this relationship is linear, then simple and multiple linear regression may be used to study such a relationship [6].

Some key assumptions that have to be made [7-8] about data for which the simple and linear regression models are being developed and such assumptions include the following.

- The assumption of linearity is made, i.e., a linear relationship between each  $x_i$  and  $y_i$  exists and, thus, the behaviour of data is adequately described by the model.
- All the observations, i.e.,  $y_i$ s, are independent of each other.
- All the observations,  $y_i$ s, follow a normal distribution.
- The mean of the distribution of each  $y_i$  is a linear function of each  $x_{ik}$ .
- The distribution of all the observations,  $y_i$ s, are the same, and the variances of the observations are also the same.
- The error component of the model is an independent variable, which is normally distributed variable, with a constant variance and mean of zero.

The approach to building simple and linear regression models may be described as follows.

Suppose there are  $N$  observations,  $1, \dots, N$ , in a sample. For any individual observation  $i$ , one can write

$$(1) \quad y_i = \mathbf{x}_i \boldsymbol{\beta} + u_i$$

where  $\boldsymbol{\beta}$  is a  $K + 1$  vector of parameters,  $\mathbf{x}_i$  is a  $(K + 1)$  row vector, and  $u_i$  is the error term. For the sample of  $N$  observations, the following expression, in matrix form, can be written

$$(2) \quad \mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{u}$$

where  $\mathbf{y}$  is a  $N$  dimensional column vector,  $\mathbf{X}$  is a  $N \times (K + 1)$  matrix, and  $\mathbf{u}$  is a  $N$  dimensional vector of error terms, i.e.,

$$(3) \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1K} \\ 1 & x_{21} & x_{22} & \dots & x_{2K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & x_{N2} & \dots & x_{NK} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} \text{ and } \mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}$$

The ordinary least squares (OLS) estimate for the model parameters is estimated using

$$(4) \quad \hat{\boldsymbol{\beta}} = (\hat{\mathbf{X}} \mathbf{X})^{-1} \hat{\mathbf{X}} \mathbf{y}$$

and the values of the dependent variable can be predicted by

$$(5) \quad \hat{y}_i = \mathbf{x}_i \hat{\boldsymbol{\beta}}$$

and the error term, also called the residual, can be estimated by

$$(6) \quad \hat{u}_i = y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}$$

Regarding the application of simple and multiple linear regression technique in renewable energy research, a structured survey was conducted. The survey observed is that the body of literature is awash with the use of the technique is renewable energy. However, no publications that provided consolidated reviews related to this subject could be found.

Thus, this paper aims to present the results of a survey on the application of simple and multiple linear regression methods in research related to wind power generation. The remainder of the manuscript comprises the following. The next section addresses the methodology of the survey.

Thereafter, salient points of the relevant publications are presented. The overall discussion of the results is then made. Finally, the conclusions of the survey drawn are drawn.

### Research methodology

A structured search of the publications was done in Google Scholar's search engine, using the advanced search option. The string used in the search was 'simple and multiple linear regression "survey" wind power generation'. There was no range of dates specified. The chosen options were "sort by relevance", "do not include patents" and "do not include citations". All relevant publications identified in the first 60 pages of the Google Scholar were included.

### Results of the review

A holistic review of the relevant papers found that they could be categorised into those on wind output power, wind speed forecasting and turbine generator temperature monitoring. In this section, the summaries of the relevant publications are presented under these categories. The key information from these summaries is extracted and succinctly recorded in Table 1.

### Wind output power

Some of the publications studied various issues related to output of wind power plants. For three areas (Szczecin, Lodz, Rzeszów) in Poland, Piotrowski *et al.* (2019) [9] forecasted wind power using air temperature, atmospheric pressure, wind azimuth and day length as explanatory variables. For Benaiciai Wind Farm in Lithuania, Deksnys and Stankevicius (2017) [10] predicted wind farm gross power output using atmosphere pressure, temperature, wind speed, and wind direction as independent variables.

The piecewise affine functions has slopes, and these functions can be used to estimate the power spectral density (PSD) which can be utilised to analyse the variability of wind power output, and in this paper Lee and Baldick (2012) [11] developed a multiple linear regression model to estimate the third slope with candidate explanatory variables including the capacity, number of wind farms, mean, standard deviation, coefficient of variation, variability, and their transformed variables. Ren *et al.* (2013) [12] [predicted wind power (MW) considering wind speed, wind direction, temperature, humidity and pressure as independent variables.

Table 1. Publications reviewed in the survey of applications of simple and multiple linear regression in wind power generation.

Item No.	Ref. No.	Area	Year	Authors	Dependent variable	Independent variables
1	[9]	Wind output power	2019	Piotrowski <i>et al.</i>	Wind power generation	air temperature, atmospheric pressure, wind azimuth & day length
2	[10]		2017	Deksnys & Stankevicius	Wind farm gross power output	atmosphere pressure, temperature, wind speed, & wind direction
3	[11]		2012	Lee & Baldick	Wind power output (slope)	capacity, number of wind farms, mean, standard deviation, coefficient of variation, variability, & transformed variables
4	[12]		2013	Ren <i>et al.</i>	Wind power prediction	wind speed & direction, temperature, humidity, & pressure
5	[13]		2016	Liu <i>et al.</i>	Wind power forecasting	wind speed & direction
6	[14]		2011	Jones <i>et al.</i>	Onshore wind-power estimates	perceived knowledge of wind power, community attachment, general attitude, visual attractiveness of wind turbines, environmental values, & perceived fairness & equity
7	[15]		2015	Soukissian & Papadopoulos	Offshore wind speed measurements from buoy	wind speed data from numerical weather prediction (NWP) simulations, & another data from satellite
8	[16]		2017	Fischer <i>et al.</i>	Wind power output	wind speed, wind direction & temperature
9	[17]	Wind speed	2017	Um & Kim	Monthly wind speed	variables based on geographical information and temperature
10	[18]		2018	Arzu <i>et al.</i>	Wind speed forecasted in Kiribati	humidity, power generation hours, mean temperature, wind gust, wind direction, & barometric pressure
11	[19]		2020	Arzu <i>et al.</i>	Wind speed forecasted in Suva	humidity, power generation hours, mean temperature, wind gust, wind direction, & barometric pressure
12	[20]		2019	Wu & Xiao	Wind speed forecasting	temperature, dew point temperature, relative humidity, station pressure, visibility and wind speed
13	[21]	Temperature	2013	Abdusamad <i>et al.</i>	Wind turbine generator stator winding temperature	cooling air temperature of the generator, nacelle temperature, ambient temperature and generator power

In [13], Liu *et al.* (2016) forecasted wind power in China using wind speed and direction as independent variables. In a study on capacity estimates for onshore wind-power development in a region of the UK, Jones *et al.* (2011) [14] identified prominent predictors to be perceived knowledge of wind power, community attachment, general attitude, visual attractiveness of wind turbines, environmental values, and issues relating to perceived fairness and equity. In a study to assess the effects of

different sources of data on offshore wind power assessment [15], Soukissian and Papadopoulos (2015) developed a linear relationship between buoy wind measurements, on one side, and, on the other, gridded wind speed data from numerical weather prediction (NWP) simulations and another data from satellite. Wind power output prediction, in the North and East of France, by Fischer *et al.* (2017) [16] relied on wind speed, wind direction and temperature as predictor variables.

## Wind speed forecasting

Some of the identified publications dealt with issues related to wind speed. In [17], for Jeju Island in South Korea, Um and Kim (2017) constructed a multiple linear regression model to predict the wind speeds, using four predictor variables based on geographical information and temperature, and then filled the gaps in observational data collected at the stations.

Arzu *et al.* (2018) and (2020), in [18] and [19], forecasted wind speed using explanatory variables including relative humidity, power generation hours, mean temperature, wind gust, wind direction and barometric pressure. To forecast wind speeds for the Ontario Province, in Canada, Wu and Xiao (2019) [20] used meteorological factors including temperature, dew point temperature, relative humidity, station pressure, visibility and wind speed as explanatory variables.

## Wind turbine generator temperature

One paper was dedicated to the monitoring of the wind turbine generator temperature. In that paper, Abdusamad *et al.* (2013) [21] predicted the generator stator winding temperature using cooling air temperature of the generator, nacelle temperature, ambient temperature and generator power as independent variables.

## Discussion of results

Fig. 1 (a) shows that although the volume of publications identified by the survey is quite low, with thirteen (13) relevant publications found, the number of publications showed an increasing trend of publications in this topic. This correlates well with the increasing interest in renewable energy, and wind power generation specifically, globally.

Furthermore, Fig. 1 (b) shows that most of the interest has been related to what wind plants can produce, directly focussing on the power outputs of the plants and indirectly via studying the wind potential in various locations.

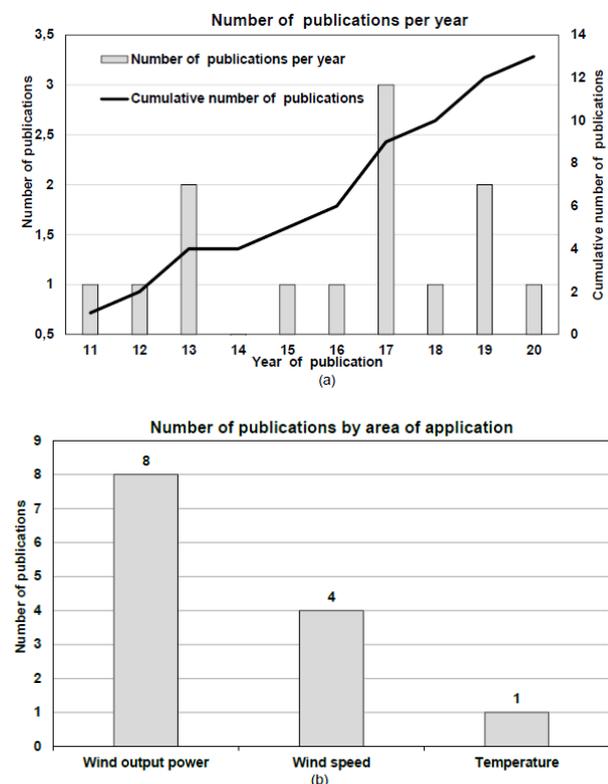


Fig. 1. Number of publications (a) by year of publication and (b) by area of application.

## Conclusion

A survey on application of simple and multiple linear regression methods in wind power generation was presented in this paper. Thirteen (13) relevant papers were identified, reviewed and summarised. An overall assessment of the publications determined that these publications could be categorised into those on wind output power (comprising the majority of publications), on wind speed forecasting and those on wind turbine generator temperature monitoring.

The paper makes a contribution to the discourse on renewable energy, in general, and wind power generation, in particular. Furthermore, the survey provides a useful resource to researchers interested in this technique by presenting concise summaries of publications and detailing all the references identified in one repository.

## Acknowledgements

The authors wish to express their appreciation to the University of South Africa and Eskom Holdings SoC, both in South Africa, for support provided to the project.

**Authors.** prof. Nhlanhla Mbuli, University of South Africa, 28 Pioneer Ave, Florida Park, Roodepoort, 1709, South Africa. E-mail: mbulin@eskom.co.za.

Bongumsa Mendu, Eskom Holdings SoC Limited, Power Road Industrial Area, De Aar, 7000, South Africa. E-mail: mendubw@eskom.co.za.

## REFERENCES

1. J. Wang, S. Qin, Q. Zhou, and H. Jiang, "Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China." *Renewable*, 76 (2015), 91-101.
2. M. Cheng, and Y. Zhu, "The state of the art of wind energy conversion systems and technologies: A review." *Energy Conversion and Management*, 88 (2014), 332-347.
3. World Energy Council, "Variable renewables integration in electricity systems: How to get it right." *World Energy Perspectives: Renewables Integration*, 2016.
4. A. Giloni, and M. Padberg, "Alternative methods of linear regression." *Mathematical and Computer Modelling*, 35 (2002), 361-374.
5. M. Łatka and M. Nowak, "Statistical analysis of the amount of the power generated by the wind power plant, according to weather conditions." *Przegląd Elektrotechniczny*, ISSN 0033-2097, R. 93 NR 10/2017, (2017) 170-174.
6. K. Schmidheiny, "The multiple linear regression model." *Short Guides to Microeconometrics*, Universtat Basel, Fall, (2013), 1-15.
7. O. Fedotova, L. Teixeira, and H. Alvelos, "Software effort estimation with multiple linear regression: Review and practical application." *Journal of Information Science and Engineering*, 29 (2013), 925-945.
8. W. T. Ambrosius, "Methods in molecular biology: topics in biostatistics." *Humana Press*, Totowa, New Jersey, USA, (2007).
9. P. Piotrowski, D. Baczyński, M. Kopyt, K. Szafranek, P. Helt, and T. Gulczyński, "Analysis of forecasted meteorological data (NWP) for efficient spatial forecasting of wind power generation." *Electric Power Systems Research*, 175 (2019), 1-9.
10. R. P. Deksnys, and A. Stankevicius, "Methodology and precision research of wind farm power prediction." *Elektronika IR Elektrotechnika*, 23 (2017), No.1, 49-56.
11. D. Lee, and R. Baldick, "Analysing the variability of wind power output through the power spectral density." *2012 IEEE Power and Energy Society General Meeting*, San Diego, USA, (2012), 1-8.
12. Z. Ren, C. Huang and M. Li, "Research on wind power prediction." *Preprints of the 3rd IEEE conf. on Energy Internet and Energy System Integration*, Changsha, China, (2019), 1504-1507.
13. Y. Liu, Y. Wang, L. Li, S. Han, and D. Infield, "Numerical weather prediction wind correction methods and its impact on

- computational fluid dynamics based wind power forecasting." *Journal of Renewable and Sustainable Energy*, 8 (2016), No. 3, 1-14.
14. C.R. Jones, B.J. Orr and J.R. Eiser, "When is enough, enough? Identifying predictors of capacity estimates for onshore wind-power development in a region of the UK." *Energy Policy*, 39 (2011), No. 8, 4563-4577.
  15. T. H. Soukissian, and A. Papadopoulos, "Effects of different wind data sources in offshore wind power assessment." *Renewable Energy*, 77 (2015), 101-114.
  16. A. Fischer, L. Montuelle, M. Mougeot, and D. Picard, "Statistical learning for wind power: a modeling and stability study towards forecasting." *Wind Energy*, 20 (2017), No. 12, 2037 - 2047.
  17. M. J. Um, and Y. Kim, "Estimating potential wind energy from sparsely located stations in a mountainous coastal region." *Meteorological Applications*, 24 (2017), 279–289.
  18. A. Arzu, S. S. Kutty, M.R. Ahmed, and M.G.M. Khan, "Wind speed forecasting using regression, time series and neural network models: a case study of Kiribati." *21st Australasian Fluid Mechanics Conference*, Adelaide, Australia (2018), 1-4.
  19. A. Arzu, M.R. Ahmed, and M.G.M. Khan, "Wind speed forecasting using regression, time series and neural network models: a casestudy of Suva." *22nd Australasian Fluid Mechanics Conference AFMC2020*, Brisbane, Australia (2020), 1-4.
  20. C.R. Jones, B.J. Orr, and J.R. Eiser. (2011) "When is enough, enough? Identifying predictors of capacity estimates for onshore wind-power development in a region of the UK." *Energy Policy*, 39 (2011), No. 8, 4563-4577.
  21. K. B. Abdusamad, D. W. Gao, and E. Muljadi, "A condition monitoring system for wind turbine generator temperature by applying multiple linear regression model." *2013 North American Power Symposium (NAPS)*, Manhattan, KS, USA (2013), 1-8.