Nhlanhla Mbuli[1], Mthobisi Diya[2]

University of South Africa, 28 Pioneer Ave, Florida Park, Roodepoort, 1709, South Africa (1)
Eskom Rotek Industries, Lower Germiston Road, Rosherville, 2022, South Africa (2)
ORCID: 1 0000-0002-5901-4971

doi:10.15199/48.2023.04.24

A survey of applications of the Kruskal Wallis test in wind power generation

***Abstract**. The authors present the results of a survey on the use of the Kruskal Wallis test in wind power generation research. An overall assessment of the qualifying publications suggested that they could be categorized into 4 logical application areas. The time series of the annual number of publications indicated a steady trend in the numbers produced annually and most publications were in the category of environmental issues. The survey contributes to the body of knowledge on wind power generation and also creates a depository of references in one source.*

Streszczenie. Autorzy przedstawiają wyniki ankiety dotyczącej wykorzystania testu Kruskala Wallisa w badaniach energetyki wiatrowej. Ogólna ocena *kwalifikujących się publikacji sugeruje, że można je podzielić na 4 logiczne obszary zastosowań. Szeregi czasowe rocznej liczby publikacji wskazywały na stały trend w liczbach wydawanych rocznie, a większość publikacji dotyczyła kwestii środowiskowych. Ankieta wzbogaca wiedzę o energetyce wiatrowej, a także tworzy depozyt referencji w jednym źródle.* (***Przegląd zastosowań testu Kruskala Wallisa w energetyce wiatrowej***)

Keywords: Analysis of variance (ANOVA); Kruskal Wallis test; nonparametric test; renewable energy sources; wind.
Słowa kluczowe: test Kruskal Wallis, wariancja, .elektryka wiatrowa.

Introduction

Although fossil fuels are still the major source of energy globally [1], concerns about the environmental problems they cause have forced mankind to look for alternative sources of energy, such as renewable energy sources. There is increasing interest in renewable energy [2], wind in particular is one of the renewable energy sources [3] , it is clean and abundant, and has been used to produce power for the utility, home owner and remote consumer. Wind can be easily captured by wind turbines [4] with higher power capacity then other renewable energy sources and this has influenced the phenomenal growth in

its utilization. The globally installed capacity of wind generation has grown tremendously from [5] 48 GW in 2004 to 370 GW in 2014. This capacity then reached 372 GW in 2015, can reach 3293 GW in 2030 and 3154 GW in 2050 [6].

In The life cycle of assets includes all the processes that are necessary for [7] establishing, operating and maintaining, and divesting a physical asset, while considering constraints brought about by economics, ergonomics, technical integrity and performance of the business. In research related any of these process in the life-cycle of a wind power plants, situations may arise where several populations of data may be studied, and a particular interest may be on whether or not there are differences among these populations.

To perform the aforementioned assessment, a technique called analysis of variance (ANOVA) may be used [8]. The ANOVA requires that the data being assessed meet a number of underlying assumptions, i.e., the data from each of the samples be normally distributed, with the estimated mean and estimated variance of the distribution determined. Hence, this is referred to as a parametric method as the underlying distribution must be known. Moreover, the variances of the groups must be the same and the data in the groups must be independently drawn.

When the data in the samples cannot meet the assumption of normality, ANOVA cannot be used, and the alternative is to use the Kruskal Wallis test which does not impose any distribution-related requirement, and is thus called a nonparametric method. The test can be used when data are ordinal or interval. The following steps [9] summarise how the test works.

- The test assesses whether the locations of the populations are the same. The null hypothesis, H_0 is that the locations of all populations are the same, and the alternative hypothesis, H_1, is that at least two locations are different.
- The test statistic is calculated using

$$H = \left[\frac{12}{n(n+1)}\sum_{j=1}^{k}\frac{T_i^2}{n_j}\right] - 3(n-1) \quad (1)$$

Firstly, the observations in all samples are ranked, based on their magnitudes, from 1 to n. The number of samples is, T_i is the sum of ranks for sample j , and n_j is the number of observations for sample j . If there are tied ranks, the average of the ranks is assigned to the tied observations. Furthermore, the H statistic is adjusted by diving equation (1) by

$$\boldsymbol{C} = 1 - \sum_{i=1}^{g}\left(\frac{t_i^2-t_i}{n^3-n}\right) \quad (2)$$

where g is the number of groups of tied ranks, and t_i is the number of tied observations in a group i.

- The rejection region for the test is

$$H \geq \chi^2_{\alpha,k-1} \quad (3)$$

since large values of H are associated with different populations. Thus, if equation (3) holds or, alternatively, the $p-\text{value}$ of the H statistic is less than the level of significance α , the null hypothesis is rejected.

- The conclusion is drawn that there is no statistically significant difference between the experimental groups.

The authors investigated the use of the Kruskal Wallis test in power systems, and found that there were many publications that had used the test, but not structured surveys could be found. Thus, in this paper the aim of the authors is to present the results of a survey conducted on the use of the test in wind power generation research. After this Introduction, Section II follows describing the methodology of the survey. Thereafter, Sections III and IV follow, which present the summaries of relevant publications and an overall assessment of publications, respectively. Finally, the conclusions of the survey are drawn in Section V.

Research methodology

The authors conducted a literature search in Google Scholar, utilizing the advance search option, using “Kruskal Wallis” “Power” “Systems” as search strings, and search options selected including “no specific timeframe”, “sort by relevance”, “exclude citations”, and “exclude patents”. Publications in the first 20 pages of the results were assessed for possible inclusion in the survey.

Results of the review

The overall review of the qualifying publications suggested grouping them into four categories, viz., (i) wind turbine selection, efficiency, and output, (ii) forecasting and resource assessment (iii) environmental impacts, (iv) societal issues, and (v) others. The summaries of the publications are presented here.

Wind turbine selection, efficiency, and energy output

In a study [10] on optimal selection of a wind turbine for a plant in Burfel, Iceland by Haghi $et\ al.$(2013), several wind speed time series (WSTS), namely, copula based model, Markov chain models, and autoregressive moving average (ARIMA) models were assessed for suitability by using the test to evaluate whether or not there was statistically significant difference between the actual data and model data. In [11] a study for optimal selection of a wind turbine for a plant in Burfel, Iceland, several algorithms (namely, (i) simple genetic algorithm, (ii) random restarts hill climber, (iii) Cross-generational elitist selection, heterogeneous recombination and cataclysmic mutation (CHC), and random tabu search) were used by Perkin $et\ al.$(2015) to optimise the levelised cost of energy (LCoE) and then the test was applied to assess whether or not, based on 30 trials for each algorithm, there was a statistically significant differences among the algorithms with respect to the results obtained.

Ertek $et\ al.$(2012), in [12], to assess the efficiencies of on-shore wind turbines (74) of top 10 wind turbine manufacturers in the world, considering efficiency score built by DEA models that incorporated three attributes (namely diameter, nominal wind speed, and nominal output), and using the test to determine whether or not there were significant differences in efficiencies score (i) among manufacturers, (ii) among cut-in wind speeds (iii) among cut-out wind speeds, and (iv) among the blade materials. In a study [13] to assess the efficiencies of the 39 state’s wind power plants in the USA by Saglam (2017), ten (10) input-output and output-input oriented Charnes, Cooper and Rhodes (CCR) models were used to calculate the efficiencies and the test is used to assess whether or not there were statistically significant differences in the efficiencies calculated using these models.

Bilbao and Alba (2010) in [14] a study to determine the distribution of wind turbines in a wind farm to maximize the total output energy and minimize the number of wind turbine, considering wind conditions and the terrain, two metaheuristic algorithms (the CHC and Simulated Annealing algorithms) are utilized, and the optimisation results, based on 30 trials, are assessed using the test to determine if there is a statistically significant difference in the objective function values and execution times for the two algorithms.

Wind forecasting and resource assessment

As part of a study on sequential reliability forecasting for wind energy, in [15] Callaway (2010) use the test to assess whether or not there is a statistically significant difference between the forecast effective load carrying capability (ELCC) and actual ELCC distributions for two locations (Breckenridge and Chandler in North Dakota, USA).

In a study [16] on one-hour ahead wind power forecasting, Mbuvha $et\ al.$(2017) used the test to assess whether or not there are differences in the forecasting accuracies of various models (viz., multi-layer peceptron (MLP), Bayesian neural networn (BNN), and BNN with automatic relevance determination (ARD RMSE), on the basis of the root mean square error (RMSE) values, obtained from running 30 trials of forecasting.

Here, Nascimento and de Souza (2017)[17]] studied the potential for wind-solar hybrid power generation in nine municipalities within the state in the state of Minas Gerais, in Brazil, by first establishing which of those states had minimum wind and solar resources for the viability of plants to be achieved, and then using the test to evaluate whether or not there was a statistically significant variation in the amounts of solar and wind resources among those municipalities

Environmental impact

As part of a study [18] to assess the potential impact of wind turbines on the petrels at San Cristobal Island in Galapagos, passage rates of these birds were monitored and standardized to number of contacts per hour per sample point, for dawn and dusk. The test was then one among those utilized by Cruz-Delgado $et\ al.$(2010) to evaluate whether or not there were variations in frequencies, passage rates, and flight altitudes among different locations, as well as whether or not there different weather conditions had influence on passage rates.

In [19] a study by Pescadora $et\ al.$(2019), in Central Eastern Spain, on assessing the effectiveness of mitigation measures implemented to avoid and minimize collisions of lesser kestrel with wind turbines and reduce mortalities, i.e., making the area around the turbines less attractive tokestrels by tilling and reducing the amount of vegetation, and, consequently, cutting down the abundance of potential prey. The test was one among those used here to study the mortalities considering variables such as temporal variables (e.g., year and month), installation, biotic variables, and effectiveness of mitigation measures.

Perrow $et\ al.$(2011) in [20], to assessed the impact of an offshore Scroby Sands wind farm, located 2.5 km off the cost of Great Yarmouth (Norfolk) in Eastern England, UK and commissioned in 2004, on the predator-prey effects, the mean rates of fish capture per minute by little terns during foraging watches of focal birds recorded over the years 2002 to 2006 were assessed using the test to see whether or not there was a statistically significant variation in the rate over the period. In a study [21] to estimate the potential mortality of griffon vultures due to wind energy development on the island of Crete (Greece) by Jung and Schindler (2019), the test was used to test whether or not the mean numbers of fatalities differed significantly among the island’s prefectures and among the different colony sizes.

In a study to estimate bird fatalities caused by wind turbines in Turkey [22] by Arikan and Turan (2017), number of carcasses were recorded from 2010 and 2014 in the vicinity of 30 wind turbines on 3 wind farms and the test was used to assess whether or not there was a significant differences by size of the species (i.e., large, medium, and small). The effect of the noise from the Manjil Wind farm, in Northern Iran, on the sleep disorders of workers was assessed by Abbasi $et\ al.$(2015) in [23] , with the test utilized to evaluate whether or not the sleep disorder significantly varied among groups of workers (i.e., repairing, security, and official).

Table 1. Classification of reviewed publications on a Kruskal Wallis test application survey in wind power generation.

Item No.	Area	Year	Ref. No.	Author	Assessment made
1	Turbine selection, efficiency and energy output	2013	[9]	Haghi $et\ al.$	Do actual and modeled data for wind speed time series differ
2		2015	[10]	Perkin $et\ al.$	Do algorithms for optimal selection of a turbine yield different results
3		2012	[11]	Ertek $et\ al.$	Do efficiencies of wind turbines differ among for various attributes
4		2017	[12]	Saglam	Do wind efficiencies obtained using different calculation models differ
5		2010	[13]	Bilbao and Alba	Do different maximum outputs result from turbine arrangement algorithms
6	Forecasting and assessment	2010	[14]	Callaway	Do wind energy effective load carrying capability (ELCC) and actual ELCC distributions differ for two locations
7		2017	[15]	Mbuvha $et\ al.$	Do accuracies of various wind forecasting models differ
8		2017	[16]	Nascimento and de Souza	Do amounts of solar and wind resources vary among nine municipalities.
9	Environmental impact	2010	[17]	Cruz-Delgado $et\ al.$	Do frequencies, passage rates, and flight altitudes of petrels differ among locations? Do passage rates vary with different weather conditions
10		2019	[18]	Pescadora $et\ al.$	Do collisions of lesser kestrel vary with levels of variables of interest
11		2011	[19]	Perrow $et\ al.$	Do mean rates of fish capture per minute by little terns vary over years
12		2019	[20]	Xirouchakus $et\ al.$	Do mortalities of griffon vultures vary among the island's prefectures and among the different colony sizes
13		2017	[21]	Arikan and Turan	Do bird mortalities caused by wind turbines vary by species size
14		2015	[22]	Abbasi $et\ al.$	Does sleep disorder vary among different worker categories
15	Societal issues	2018	[23]	Bauwensa and Devine-Wright	Do attitudes of groups towards renewable energy projects vary with involvement in projects
16		2018	[24]	Thomson and Kempton	Do attitudes of, visual impacts and auditory impacts on residents of nearby plants varied significantly across the demographic groups
17		2016	[25]	Alves de Sena $et\ al.$	Evaluation of community knowledge level about wind and solar power, their social acceptance, and perceptions towards cost, local development and environmental impacts.
18		2020	[26]	Cronin $et\ al.$	Do perception of public of offshore wind farms vary with education level
19	Other	2015	[27]	Haghi and Lotfifard	Do vine-copula based model and multivariate autoregressive (MAR) model adequately represent the distribution of wind speeds
20		2014	[28]	Li $et\ al.$	Do results of different algorithms for solving economic dispatch with wind power differ
21		2018	[29]	Jung and Schindler	Do the moments of the original wind speed distribution (ORI) and modified distribution (MOD) differ

Societal issues

In [24], a study on community participation and attitudes towards renewable energy in Belgium, the test was utilized by Bauwensa and Devine-Wright (2018) to determine whether or not there were significant differences in the attitudes of groups towards renewable energy depending on their involvement in such projects. In [25], a study by Thomson and Kempton (2018) to assess the attitudes of, the visual impacts and auditory impacts on residents of nearby coal and generation plants, the test was used to evaluate if the responses varied significantly across the demographic groups tested, i.e., gender, age, home ownership, education, and income.

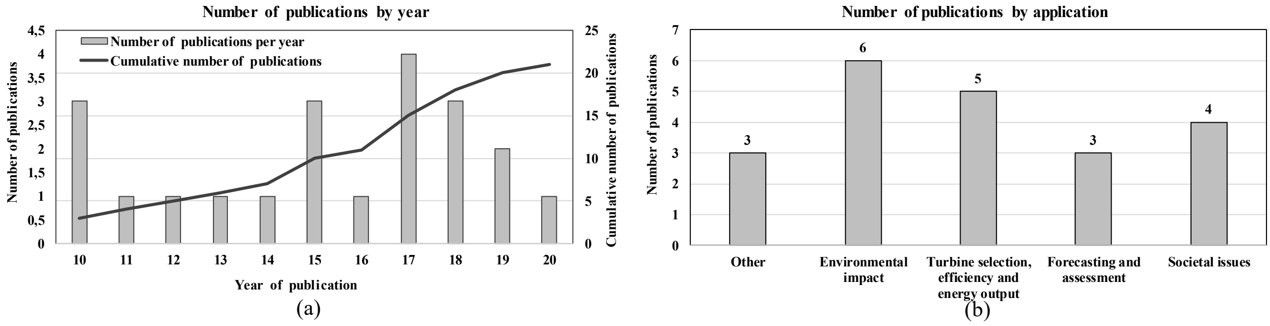
In a study by Alves de Sena $et\ al.$(2016) [26], in the Brazilian electricity system, to evaluate level of knowledge of wind and solar power, their social acceptance, and perceptions towards cost, local development and environmental impacts, the test was used to determine whether responses to these aspects significantly differed for various types of groupings of the population surveyed. In [27] by Cronin $et\ al.$(2020) assess the perception of the Irish public of offshore wind farms, the test was used to evaluate whether the attitude was significantly affected by the level of education.

Other applications

In a study [28] on spatiotemporal modelling of wind generation with a view to optimal sizing of energy storage, the test is used by Haghi and Lotfifard (2015) to test whether or not the vine-copula based model and multivariate autoregressive (MAR) model can adequately represent the distribution of wind speeds.

In [29] the solution of the economic dispatch problem, with wind power incorporated, several algorithm (multiple-group search optimiser with multiple producers (MGSOMP), GSOMP and non-dominated sorting genetic algorithm-II (NSGA-II)) are used by Li $et\ al.$(2014)to solve the problem and the test is used to evaluate whether or not, based on 30 independent stochastic searches of the optimal solutions, there is a statistically significant difference in the optimal solutions obtained using these methods.

Jung and Schindler (2018) in a study [30] to assess the robustness of fitted wind distribution to measurement errors, missing data, and low temporal resolution, the test was applied to assess whether or not modifications of data caused significant differences between the moments of the original wind speed distribution (ORI) and modified distribution (MOD), using data for a number of wind speed time series.



(a) (b)

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

Discussion of results

The time series of the number of publications per annum is shown in Fig. 1(a), and it shows a constant pretty much constant trend, with variation in the numbers within a very narrow range. Moreover, Fig. 1(b), the number of publications by area of application, indicates that the environmental impact category has the marginally highest number of publications. Leaving the “Other” category aside, the remainder of the categories do not significantly different numbers among themselves and when compared to the environmental impact category.

Conclusion

In this paper, the authors present the results of a survey conducted on the use of the Kruskal Wallis test in wind power generation research. After conducting a structured literature in Google Scholar, the authors did an overall assessment of the qualifying publications and found that these could be categorized into 4 areas, namely, (a) wind turbine selection, efficiency, and energy output, (b) wind forecasting and resource assessment, (c) environmental impacts, (d) societal issues, and (e) others. It was found that the annual number of publications on this subject has remained stable of the period of interest. Furthermore, the number of publications is spread relatively evenly among the four categories, leaving “Other” side, although the environmental impact category has marginally the highest number of publications. The survey generally contributes to the body of knowledge on wind power generation. Also, it provides succinct overviews of relevant publications, and creates a depository of references for researchers.

Acknowledgements

The authors are grateful to University of South Africa and Eskom, both in South Africa, for supporting this project.

Authors. Prof. Nhlanhla Mbuli, University of South Africa, 28 Pioneer Ave, Florida Park, Roodepoort, 1709, South Africa. E-mail: mbulin@eskom.co.za.
Mthobisi Diya, Eskom Rotek Industries, Lower Germiston Road, Rosherville, 2022, South Africa. E-mail: diyam@eskom.co.za.

REFERENCES

1 M. Cheng, J. Zhu, N.A. Rahim and K.H. Solangi. (2014) “The state of the art of wind energy conversion systems and technologies: A review.” *Energy Conversion and Management* 88 (2014): 332-347.
2 N. Mbuli, B. Mendu, “A survey of applications of simple and multiple linear regression in wind power generation.” PRZEGLĄD ELEKTROTECHNICZNY, ISSN 0033-2097, R. 98 NR 3/2022, (2022) 1-7.
3 X. Sun, D. Huang and G. Wu. (2012) “The current state of offshore wind energy technology development.” *Energy* 41 (2012): 298-312.
4 R. Saidur, M.R. Islam, N.A. Rahim and K.H. Solangi. (2010) “A review on global wind energy policy, Renewable and Sustainable Energy Reviews.” 14 (2010): 1744-1762.
5 World Energy Council. (2016) “Variable Renewables Integration in Electricity Systems: How to get it Right. *World Energy Perspectives: Renewables Integration* 2016.
6 M. Ram, D. Bogdanov, A. Aghahosseini, A. S. Oyewo, A. Gulagi, M. Child, H.-J. Fell and C. Breyer. (2017) “Global Energy Systems based on 100% Renewables-Power Sector.” *Study by Lappeenranta University of Technology and Energy Watch Group*, Lappeenranta, Berlin, November 2017.
7 M. Haffejee and A. C. Brent. (2008) “Evaluation of an Integrated Asset Life-Cycle Management (ACLM) Model and Assessment of Practices in the Water Utility Sector.” *Water SA* 34.2 (2008) 285-290.
8 P. Sprent and N. C. Smeeton. (2001) “Applied Nonparametric Statistical Methods.” *3rd Edition, Chapman and Hall/CRC*, Florida, USA 2001.
9 G. Keller. (2014) “Statistics for Management and Economics*,” 10th Edition, CENGAGE:* Connecticut, USA 2014
10 H. V. Haghi, M. T. Bina, and M. A. Golkar. (2013) “Nonlinear Modeling of Temporal Wind Power Variations.” *IEEE Transactions on Sustainable Energy* 4.4 (2013): 838-847.
11 S. Perkin, D. Garrett, P. Jensson. (2015) “Optimal wind turbine selection methodology: A case-study for Búrfell, Iceland.” *Renewable Energy* 75 (2015) 165-172.
12 G. Ertek, M. M. Tunç, E. Kurtaraner and D. Kebude. (2012) “Insights into the Efficiencies of On-Shore Wind Turbines: A Data-Centric Analysis.” *2012 International Symposium on Innovations in Intelligent Systems and Applications*, Trabzon, Turkey 2-4 July 2012.
13 Ü. Saglam.(2017) “A two-stage data envelopment analysis model for efficiency assessments of 39 state’s wind power in the United States.” *Energy Conversion and Management* 146 (2017): 52–67.
14 M.Bilbao and E. Alba. (2010) “CHC and SA Applied to the Distribution of Wind Turbines on Irregular Fields.”*16th Argentine Congress Of Computer Sciences,* Cacic (2010): 72-81.
15 D. S. Callaway. (2020) “Sequential Reliability Forecasting for Wind Energy: Temperature Dependence and Probability Distributions.” *IEEE transactions on energy conversion* 25.2 (2010): 577-585
16 R. Mbuvha, M. Jonsson, N. Ehn and P. Herman. (2017) “Bayesian Neural Networks for One-Hour Ahead Wind Power Forecasting.” *6th International Conference on Renewable Energy Research and Application*, San Deigo,CA.USA 5-8 Nov (2017): 591-596
17 E. M. Nascimento, and J. D. de Souza. (2017) “Hybrid Power Plants: Viability for Cities in Minas Gerais.” *Engineering Journal* 21.5(2017): 37-52.
18 G. F. Cruz-Delgado1, D. A. Wiedenfeld, and J. A. González. (2010) “Assessing the potential impact of wind turbines on the endangered Galapagos Petrel Pterodroma phaeopygia at San Cristóbal Island.” *Biodivers Conserv* 19 (2010): 679–694 https://doi.org/10.1007/s10531-009-9727-y.
19 M. Pescadora, J. I. G. Ramírezb and S. J. Perisc. (2019) “Effectiveness of a mitigation measure for the lesser kestrel (Falco naumanni) in wind farms in Spain.” *Journal of Environmental Management* 231 (2019) 919–925.

20 M. R. Perrow, J. J. Gilroy, E. R. Skeate and M. L. Tomlinson. (2011) “Effects of the construction of Scroby Sands offshore wind farm on the prey base of Little tern Sternula albifrons at its most important UK colony.” *Marine Pollution Bulletin* 62 (2011) 1661–1670
21 S. M. Xirouchakus, E. Armeni, S. Nikdopoulou and J. Halley. (2019) “Estimating the potential motality of Griffon Valtures (Gyps fulvus) Due to wind energy development on the Island of Crete (Greece).” *Springer Nature Switzerland AG 2019 R. Bispo et al. (eds.), Wind Energy and Wildlife Impacts* (2019): 205-222 https://doi.org/10.1007/978-3-030-05520-2_13 .
22 K. Arikan and S. L. Turan. (2017) “Estimation Of Bird Fatalities Caused By Wind Turbines In Turkey.” *Fresenius Environmental Bulletin* 26.11(2017): 6543-6550.
23 M. Abbasi, M. R. Monnazzam, S. A. Zakerian and A. Yousefzadeh.(2015) “Effect of Wind Turbine Noise on Workers’ Sleep Disorder: A Case Study of ManjilWind Farm in Northern Iran.” *World Scientific Publishing Company, Fluctuation and Noise Letters* 14. 2 (2015) 1550020 (15 pages).
24 T. Bauwensa, P. Devine-Wright. (2018) “Positive energies? An empirical study of community energy participation and attitudes to renewable energy.” *Energy Policy* 118 (2018) 612-625
25 H. Thomson, W. Kempton. (2018) “Perceptions and attitudes of residents living near a wind turbine compared with those living near a coal power plant.” *Renewable Energy* 123 (2018) 301-311.
26 L. A. de Sena, P. Ferreira and A. C. Braga. (2016) “Social acceptance of wind and solar power in the Brazilian electricity system.” *Environ Dev Sustain, Springer Science and Business Media*, Dordrecht 2016.
27 Y. Cronin, E. Wolsztynski and V. Cummins. (2020) “Public Perception of Offshore Wind Farms Report Part 2.” *EirWind Project Deliverable D4.7 Report*, MaREI Centre, ERI, University College Cork, Ireland.(2020) DOI: http://doi.org/10.5281/zenodo.3948450
28 H. V. Haghi, and S. Lotfifard. (2015) “Spatiotemporal Modelling of Wind Generation for Optimal Energy Storage Sizing.” *IEEE Transactions on Sustainable Energy* 6.1 (2015): 113-121.
29 Y.Z. Li, Q.H. Wu, M.S. Li, J.P. Zhan. (2014) “Mean-variance Model for Power System Economic Dispatch with Wind Power Integrated.” *Energy* 72 (2014) 510-520.
30 C. Jung and D. Schindler. (2018) “Sensitivity analysis of the system of wind speed distributions.” *Energy Conversion and Management* 177 (2018) 376–384.