International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:12, No:5, 2018

Wind Farm Power Performance Verification Using Non-Parametric Statistical Inference

M. Celeska, K. Najdenkoski, V. Dimchev, V. Stoilkov

Abstract**—**Accurate determination of wind turbine performance is necessary for economic operation of a wind farm. At present, the procedure to carry out the power performance verification of wind turbines is based on a standard of the International Electrotechnical Commission (IEC). In this paper, nonparametric statistical inference is applied to designing a simple, inexpensive method of verifying the power performance of a wind turbine. A statistical test is explained, examined, and the adequacy is tested over real data. The methods use the information that is collected by the SCADA system (Supervisory Control and Data Acquisition) from the sensors embedded in the wind turbines in order to carry out the power performance verification of a wind farm. The study has used data on the monthly output of wind farm in the Republic of Macedonia, and the time measuring interval was from January 1, 2016, to December 31, 2016. At the end, it is concluded whether the power performance of a wind turbine differed significantly from what would be expected. The results of the implementation of the proposed methods showed that the power performance of the specific wind farm under assessment was acceptable.

Keywords**—**Canonical correlation analysis, power curve, power performance, wind energy.

I. Introduction

In the process of designing a wind farm (WF), the estimation of the total production of electrical energy is the most essential phase. The estimation is based on the power curve of the selected wind turbine (WT) that would be installed in the WF. Power curves are catalogue data that are proclaimed and guaranteed by the manufacturer of the WTs. After a WF is built, the power performance of each WT must be verified in accordance with the international standard IEC61400-12-1 [1]. Accurate determination of WT performance is necessary for economic operation of a WF. Generally, the application of the method elaborated in the international standard is not inexpensive. For this reason, it is important to devise alternative methods, with wide and simple applicability.

Neural network approaches have been used to try and characterise the power performance of WTs in a WF. Stochastic techniques are often dismissed as being inferior to those which use neural networks, but those stochastic techniques proposed are often overly simplistic [2], [3]. The simplicity of the stochastic methods is due to deduction of data variability to only one parameter- the power output, [4], [5]. That means, the variation of power output of each WT in the WF, is analyzed upon the wind speed measurements gained from the measuring mast positioned within the farm, [1]. Power curve measurements based on met masts according to the IEC standard, not always can be carried out at WTs located: (i) in the middle of WFs, concerning large WF on flat terrain, neither (ii) small WFs on semi-complex terrain, [6], [7]. In the first case, due to the wake immersion of neighboring WTs, no position can be found for the met mast where the wind measured at the mast would be representative for the inflow of the WT. The same applies to the second case, where due to local vortices caused by the terrain orography, the wind speed mast measurements cannot be completely confidential for further analysis, [8]-[11]. It is clear that, the main factors affecting array efficiency are WF layout, wind regime and the type of terrain, [2].

In this paper, the power performance verification is carried out by using statistical method, [12]. Canonical correlation analysis (CCA) is the one of the oldest and best-known methods for discovering and exploring dimensions that are correlated across sets. The canonical correlation is a multivariate analysis, which allows to correlate the measured wind speed and power outputs variations, to those provided in the power curve from the WT manufacturer. The method elaborated in the paper estimates the similarity among guarantied power curve (GCp) and estimated power curve from each WT and says whether the power performance of the specific WT under assessment differs significantly from what would be expected.

II. General Data for WP Bogdanci

The WF analyzed for this report is located in southeast part of Rep. of Macedonia. It consists of 16 turbines from the same type – SIEMENS SWT-2.3-93 (80-m hub height). The GCp is represented by data pairs, where the wind speed interval between cut-in (4 m/s) and cut-out wind speed (25 m/s) sampled at a distance of 1 m/s.

The wind park area covers elevations of between 280 and 500 m above sea level. The long-term average wind speed at an elevation of 500 m has been established as 6.9 m/s at 50 m above ground. The turbines are generally lined up in rows which are perpendicular to the prevailing wind direction (northwest) at the site. Average distance between turbines in the same row is roughly 2,2 rotor diameters. Also, there are two MET towers which measure wind speed (m/s), direction

M. Celeska is with Electrical Machines, Transformers and Apparatuses Department, Ss. Cyril and Methodius University in Skopje, Republic of Macedonia (phone: 0038975286876; e-mail: celeska@feit.ukim.edu.mk).

K. Najdenkoski and V. Stoilkov are with the Electrical Machines, Transformers and Apparatuses Department, Ss. Cyril and Methodius University in Skopje, Republic of Macedonia.

V. Dimchev is with the Electrical Measurements and Materials Department, Ss. Cyril and Methodius University in Skopje, Republic of Macedonia.

(0-359°) at 50 m, barometric pressure (mBar) and temperature (°C). The data sampling from the SCADA system is averaged over 10-minute periods, as superposed in IEC 61400–12–1.

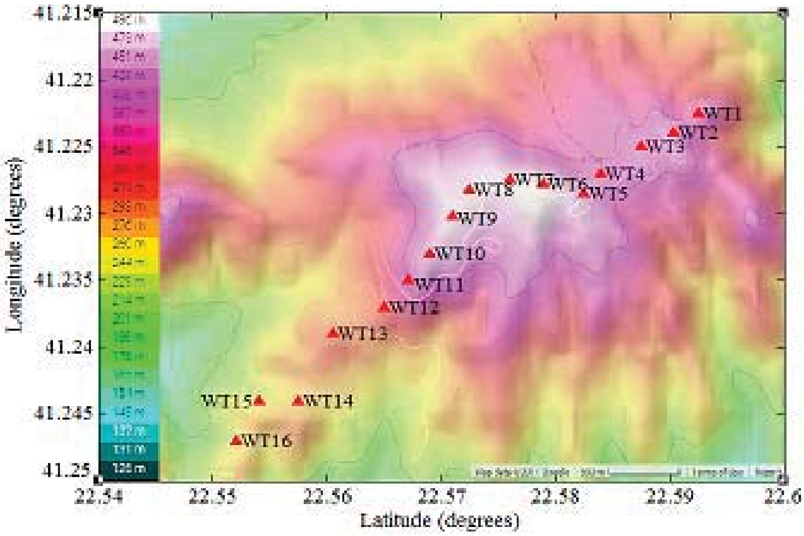


Fig. 1 Orographic map of the WF with WT positions in geographic coordinates system

III. Proposed Method

A. Identification Criteria for Outliers

Before implementation of the main procedure for power curve verification, equally important is the methodology of filtering the data from outliers. Outliers are defined as the data points that lie outside the probabilistic WF power curve and are caused by multiple non-meteorological factors: communication error, WT outage or curtailment action, [13]. Yet, there are cases when some coarse data should not be classified as outliers because they contain relevant information for the power performance of some WT. So, the entire process of filtering the data must be performed very warily.

Fig. 2 shows the scatter plot of no filtered data of the wind speed versus the power output of the WTs, which gives us rough but initial assessment for the work of every WT. WTs that are adjacent to one another have correlated power outputs, [14]. It can be noticed that there are outliers and extreme observations at each WT. To avoid the consideration of any wrong data during calculations and final comparisons, a set of right data have been obtained for each turbine.

In this paper, three identification criteria for outliers are proposed:

i) All recorded numerical values for power outputs that have values equal or below zero when the wind speed>cut-in speed, are rejected because these values represent faults in the performance of the WT during its operation, or because the turbine status was 0 (not in operation). It must be noticed that, at some wind speed values, there were many negative active power outputs, generally between 2-7 m/s for all turbines without exception.

ii) Determination of outliers on a criterion of greater than three times standard deviations (3σ) from the WT power curve in order to eliminate the power outputs outliers, registered around each wind speed. Here, it is proposed to use modified z-score method for detecting that type of outliers, [15].

Each wind speed value paired with power output value defined in the guaranteed power curve by the manufacturer, represents a single sample of interest. Intervals of each sample are $-0.5+v_i<v_i<0.5+v_i$. We denote the order statistics for each wind speed value as: $v_{(1)},\ldots,v_{(N)}$, where N is the size of the sample $P_1,\ldots,P_N$ of power output. We have a sample of N observations with median $\widetilde{P}$. The modified z-score of an observation is defined as:

$$M_i = \frac{0.6745(P_i - \widetilde{P})}{MAD} \tag{1}$$

with MAD denoting the median absolute deviation. Calculated modified z-scores for each recorded power output-P_i from the defined sample, with an absolute value of greater than 3MAD are labeled as potential outliers. Outliers detected at each wind speed by the modified z-score, are around 0.3% for each WT.

iii) Third type is consecutive outliers, which are detected below the cut-in wind speed and above 23 m/s. Since, the production of electricity at intake wind speeds in the interval 0-3.5 m/s is insignificant, these data are eliminated. On the other hand, that range of data is not included in the GCp. Similar, there were not any significant power output points for the wind speed interval from 23 m/s and along. In the range 23-25 m/s, only 0.08% of data was registered and the output power was not as expected.

Discarded data in percent from the whole data base for each WT, according to the three criteria, are given in Table I. It can be noted that the percent of outlier data is relatively big,

International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:12, No:5, 2018

starting with 28.99% at WT8, to 39.23% at WT15. In the other words, we reject irregular data in duration of 103-143 days/year, depending on the WT. It must be noted that the purpose of the paper is verification of the similarity between GCp and power curves from each WT gained form valid data, not a correction of outliers.

TABLE I
PERCENT OF OUTLIERS FOR EACH WT

No. of WT	Criteria $P_{out} \leq 0$ kW (%)	Criteria $\lvert M_i \rvert > 3$MAD (%)	Criteria $v_i<3.5$ m/s (%)	$v_i>23$ m/s (%)	Total outliers (%)
WT1	17.43	3.48	13.94	0.06	34.86
WT2	16.10	3.22	12.88	0.18	32.20
WT3	16.17	3.23	12.94	0.01	32.34
WT4	16.29	3.25	13.03	0.03	32.58
WT5	15.55	3.11	12.44	0.07	31.10
WT6	15.12	3.02	12.10	0.39	30.24
WT7	14.79	2.95	11.84	0.14	29.59
WT8	14.49	2.89	11.60	0.15	28.99
WT9	14.66	2.93	11.73	0.42	29.32
WT10	14.58	2.91	11.67	0.55	29.17
WT11	15.96	3.19	12.77	0.03	31.90
WT12	17.52	3.50	14.02	0.07	35.04
WT13	19.57	3.91	15.66	0.04	39.14
WT14	17.73	3.57	14.19	0.05	35.47
WT15	19.61	3.92	15.69	0.16	39.23
WT16	18.60	3.72	14.88	0.24	37.20

B. Finding the Most Representative Points of the Power Output of Each WT

The data required for modeling a power curve are the wind speed and power output recorded at periodic intervals over a long time, [2]. In this paper, the time measuring interval was from January 1, 2016 to December 31, 2016.

In order to construct power curve for each WT that will represent the most adequate, i.e. most significant pair points (wind speed, power output) from the time series data, we propose choosing the most frequently recorded data pairs. The methodology used for this purpose is bivariate normal distribution. From the time series data, subordination by frequency of occurrence is done for every interval $-0.5+v_k<v_k<0.5+v_k$, where v_k denotes for wind speeds in the range 4-23 m/s (defined by GCp), i.e. k=1,…20. For each WT and each wind speed interval, the bivariate normal distribution is applied.

An $r \times c$ contingency table with cell probabilities p_{ij} specifies the bivariate normal distribution of two discrete v_i and P_j, where $1 \leq i \leq r$ and $1 \leq j \leq c$. The bivariate normal distribution is the statistical distribution with probability density function, [16]:

$$p(v_i, P_j) = \frac{1}{2\pi\sigma_v\sigma_P\sqrt{1-\rho^2}} \exp\left[-\frac{z_{ij}}{2(1-\rho^2)}\right] \tag{2}$$

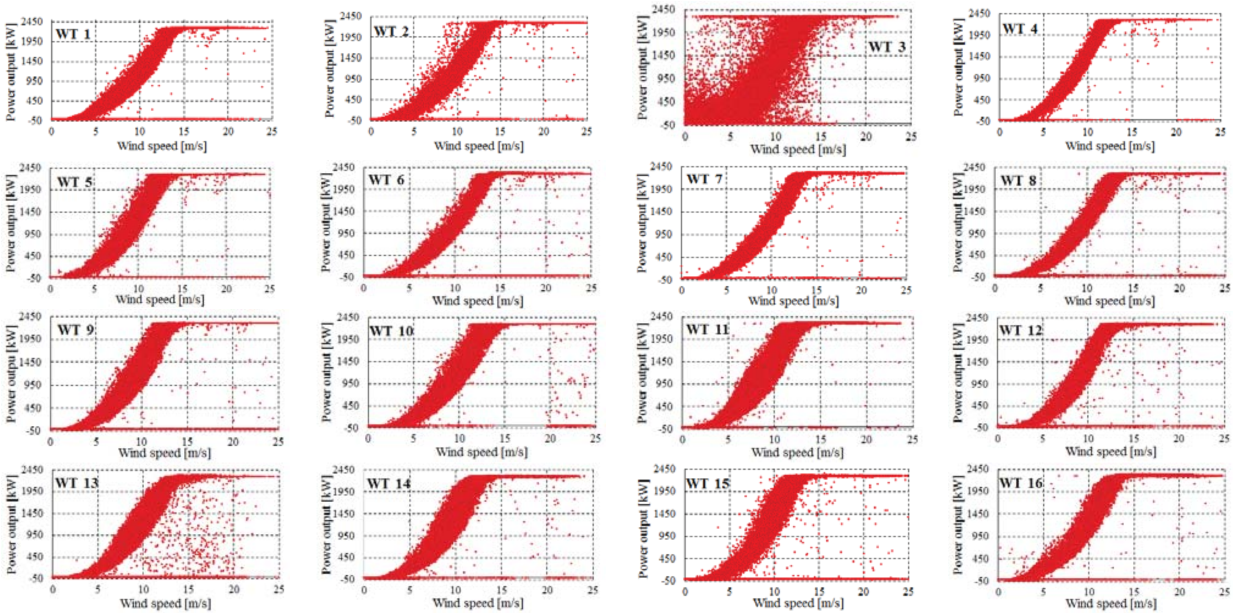
where

$$z_{ij} = \frac{(v_i-\bar{v})^2}{\sigma_v^2} - \frac{2\rho(v_i-\bar{v})(P_j-\bar{P})}{\sigma_v\sigma_P} + \frac{(P_j-\bar{P})^2}{\sigma_P^2} \tag{3}$$

and

$$\rho = cor(v,P) = \frac{V_{12}}{\sigma_v\sigma_P} \tag{4}$$

is the correlation of v and P, and V_{12} is the covariance. Symbols σ_v, σ_P, $\bar{v}$ and $\bar{P}$ in (2)-(4), stand for standard distribution and mean value of wind speed and power output, respectively.



International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:12, No:5, 2018

Fig. 2 Scatter plot of the wind speed vs the power output of the 16 WTs for 2016

The variance of the wind speed is analyzed in each interval $0.5+v_k<v_k<0.5+v_k$, because of the big difference of so-called "response" of the WT, i.e. power output. Fig. 3 represents the GCp and all 16 power curves obtained for the 16 WTs. The variance in the wind speed among different WTs in the interval of 5-11 m/s can be noticed. Generally, a performance improvement is recommended for all WTs in this interval. Before achieving nominal wind speed (13 m/s) and above it, at each WT, the performance was found within the expected limitations.

C. Canonical Correlation

After establishing the power curves for each WT, the procedure to carry out the power performance verification is next to be done.

CCA is part of multivariate analysis of variance (MANOVA), which represents a method for exploring the relationships between two multivariate sets of variables, all measured on the same individual, [12], [17]. For CCA, $\boldsymbol{X}$ has to be an $n \times p$ and $\boldsymbol{Y}$ $n \times q$ matrix, with p and q at least 2, where p is the number of variables contained in the set $\boldsymbol{X}$ and q is the number of variables in the set $\boldsymbol{Y}$.

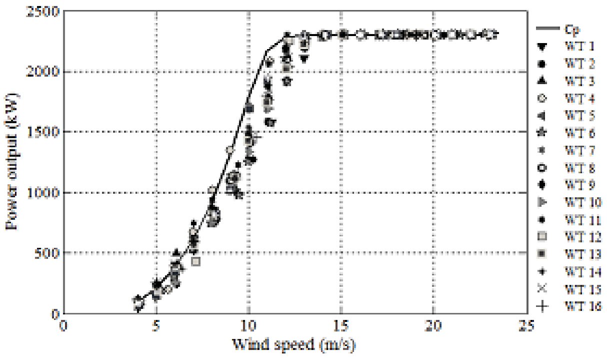


Fig. 3 Power curves of the WTs of WF Bogdanci in 2016 year and the GCp

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_n \end{pmatrix} \tag{5}$$

$$\boldsymbol{Y} = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1q} \\ y_{21} & y_{22} & \cdots & y_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nq} \end{pmatrix} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \boldsymbol{Y}_2 \\ \vdots \\ \boldsymbol{Y}_n \end{pmatrix} \tag{6}$$

There are different canonical variates within each set. If there are p variables in $\boldsymbol{X}$ and q variables in $\boldsymbol{Y}$, then there are at most k=min(p,q) canonical variates in either set. These are $U_i=a'_iX$ and $V_i=b'_iY$, with i ranging from 1 to k. Across each set, U_i and V_i are uncorrelated. The correlation between corresponding canonical variates U_i and V_i is the ith canonical correlation, as shown in (7). In the other words, the first canonical variate is the linear combination of variables in one set that has the highest possible multiple correlation with the variables in the other set, [18].

$$Var(U_i) = a' \sum_{1p} a;\ Var(V_i) = b' \sum_{nq} b;\ Cov(U_i,V_i) = a' \sum_{pq} b \tag{7}$$

Although the correlation measure is:

$$Corr(U_i,V_i) = \frac{Cov(U_i,V_i)}{\sqrt{Var(U_i)Var(V_i)}} \tag{8}$$

The first pair of canonical variables, or first canonical coefficient, is the pair of linear combinations U_1, V_1 having unit variances, which maximize the correlation. The kth canonical coefficient is the pair of linear combinations U_k, V_k having unit variances, which maximize the correlation among all choices uncorrelated with the previous k-1 canonical variable pairs, [15]. The existence of overall relationships between two sets of variables is tested by the canonical correlation coefficients and the significance measures the size of relationships. As the correlation coefficients have greater value, the similarity in the set is higher.

Lastly Wilk's lambda, is used as a test of significance of the canonical correlation coefficient. The closer to zero the statistic is, the more the variable in question contributes to the model. The null hypothesis is rejected, when Wilk's lambda is close to zero, although this should be done in combination with a small p-value [15]. The null hypothesis is that the data among one set are strongly correlated and the significance level is 5 %. Wilk's lambda is given as:

$$\Lambda = \prod_{p=1}^{k} \frac{1}{1+\lambda_p} \tag{9}$$

where $\lambda_{1,\dots,}$ λ_p are eigenvalues of from the matrix term produced from the submatrices of the covariance matrix, [15].

In our case, every WT is a single individual, the set $\boldsymbol{X}$ contains the wind speed and power output vector data from GCp, and $\boldsymbol{Y}$ contains the wind speed and power output vector data from the calculated WT curve. Namely, p=q=2 (two vector columns for wind speed and power output) and the number of sets is n=1,…,20 because of the wind speeds range, 4-23 m/s. Since, we are testing bivariate set of variables, the number of correlation coefficients is two (F_1, F_2). Each WT's power curve is correlatively tested to the GCa. Table II shows the calculations obtained with the CCA. Taking into consideration all test score values listed in the table, it can be considered without any exception, that there is a positive correlation among each WT with the GCp. "*Sig*" or significance (p-value) is to quantify the importance of the canonical coefficients. If the significance is small, (i.e. under 0.05) the null hypothesis will be reject. The statistical significance is additional measurement when the values of the canonical coefficients F_1 and F_2 are questionable, or have values with limited importance for the canonical correlation, [17]. From Table II, it can be noted that all of these tests, for each WT, are significant with $p<0.05$. The correlation output shows overall model fitting.

IV. Conclusion

In the paper, a subsequent methodology for obtaining and verification of power curve of a WT in operation has been presented. The usefulness of the CCA is demonstrated by applying over actual WF data. For a more detailed and more reliable verification of the power curve, the method seeks for similarity among the most frequent data pairs, not only comparison of the WTs' power output and the one provided by the WT manufacturer. The following experience has been gained by the application of the methodology:

(i) it is well situated to analyze changes of the power curve of WTs- both separate for each interval of the GPc, and overall assessment;

TABLE II
Results of the CCA at Significance Level of 5 %

No. of WT	F_1	F_2	*Wilk's* λ_1	*Sig. (1)*	*Wilk's* λ_2	*Sig (2)*
WT1	0.931	0.020	0.133	0.7	0.999	0.936
WT2	0.920	0.090	0.152	0.139	0.992	0.174
WT3	0.912	0.297	0.154	0.164	0.912	0.217
WT4	0.847	0.077	0.282	0.101	0.994	0.754
WT5	0.920	0.067	0.153	0.76	0.996	0.786
WT6	0.936	0.010	0.123	0.2	0.999	0.968
WT7	0.911	0.257	0.159	0.206	0.934	0.287
WT8	0.912	0.095	0.167	0.153	0.991	0.7
WT9	0.902	0.156	0.183	0.423	0.976	0.524
WT10	0.918	0.291	0.143	0.569	0.915	0.227
WT11	0.871	0.403	0.202	0.329	0.838	0.087
WT12	0.881	0.172	0.217	0.516	0.971	0.482
WT13	0.921	0.097	0.151	0.161	0.991	0.693
WT14	0.905	0.141	0.177	0.345	0.98	0.565
WT15	0.904	0.027	0.182	0.212	0.999	0.914
WT16	0.909	0.175	0.168	0.54	0.969	0.473

(ii) the procedure is not subject to any costs, as only SCADA data is needed;

(iii) there is no necessity of air density data normalization, as in IEC 61400–12–1;

(iv) the methodology is less sensitive to site effects than measurements with masts;

(v) the WT power curve verification method is not a significant problem. This however is very dependent on data purification form outliers. As noted in the identification criteria of outliers, the whole process of filtering the data must be accomplished deliberately in order not to remove relevant data from the power performance of the WTs.

Acknowledgment

The authors would like to thank ELEM-Macedonian Power Plants Company for the information provided.

References

[1] IEC 61400. Part 12–1: Power Performance Measurements of Electricity Producing Wind Turbines; IEC 61400-12-1, International Standard; 2005.

[2] M. Lydiaa, S. S. Kumarb, A. I. Selvakumara, G. E. Prem Kumar, "A comprehensive review on wind turbine power curvemodeling techniques", *ELSEVIER Renewable and Sustainable Energy* vol. 30, pp. 452–460, 2014.

[3] A. Llombart, S. J. Watson, D. Llombart, J. M. Fandos, "Power Curve Characterization I: Improving the Bin Method", *Proceedings of International Conference on Renewable Energies and Power Quality*, Zaragoza, Spain, March 2005.

[4] W. Hernandez, J. L. López-Presa, J. L. Maldonado-Correa, "Power Performance Verification of a Wind Farm Using the Friedman's Test", *Sensors* 2016, vol. 16, 816; doi:10.3390/s16060816.

[5] W. Hernandez, J. L. Maldonado-Correa, "Power Performance Verification of a Wind Turbine by using the Wilcoxon Signed-Rank Test", *IEEE Transactions on Energy Conversion*, vol. 32, no. 1, March 2017.

[6] A. Albers, H. Klung, D. Westermann, "Power Performance Verification", *Proceedings of European Wind Energy Conference*, 1-5 March, 1999, Nice, France, pp. 657–660.

[7] H. Oh, B. Kim, "Comparison and verification of the deviation between guaranteed and measured wind turbine power performance in complex terrain", *Energy Journal,* vol. 85, 2015, pp. 23-29.

[8] F. Pedersen, S. Gjerding, P. Enevoldsen, J.K. Hansen, H.K. Jørgensen, "Wind turbine power performance verification in complex terrain and wind farms", Denmark. Forskningscenter Risoe. Risoe-R; No.1330, 2002.

[9] M. Lee, S. Hur, N. Choi, "A numerical simulation of flow field in a wind farm on complex terrain", *Proceedings of the Seventh Asia-Pacific Conference on Wind Engineering*, Taipei, Taiwan, 8–12 November 2009; pp. 1–8.

[10] G. Polanco, V. M. Shakeel, "Role of advanced CAE tools in the optimization of wind resource assessment of complex terrains", *Proceedings of the 4th IEEE International Conference on Cognitive Info communications*, Budapest, Hungary, 2–5 December 2013; pp. 687–691.

[11] C. Xu, J. Yang, C. Li, W. Shen, Y Zheng, D Liu, "A research on wind farm micro-sitting optimization in complex terrain", *Proceedings of the 2013 International Conference on Aerodynamics of Offshore Wind Energy Systems and Wakes*, Lyngby, Denmark, 17–19 June 2013; pp. 669–679.

[12] Brian S. Everitt, David Howell, "Encyclopedia of Statistics in Behavioral Science", ISBN: 978-0-470-86080-9, Hoboken, New Jersey, John Wiley & Sons, April 2005.

[13] X. Ye, Z. Lu, Y. Qiao, Y. Min, M. O'Malley, "Identification and Correction of Outliers in Wind Farm Time Series Power Data", IEEE Transactions on power systems, vol. 31, no. 6, November 2016, pp. 4197-4205.

[14] Y. Wan, M. Milligan, and B. Parsons, "Output power correlation between adjacent wind power plants," J. Sol. Energy Eng., vol. 125, no. 4, November 2003, pp. 551-555.

[15] H. Pham, *Handbook of Engineering Statistics*, Springer-Verlang, London, 2006, pp. 117-118.

[16] Weisstein, Eric W. "Bivariate Normal Distribution." From MathWorld--A Wolfram Web Resource, last visited 07.11.2017 at 11:21-hour http://mathworld.wolfram.com/BivariateNormalDistribution.html

[17] J. E. Borovsky, "Canonical correlation analysis of the combined solar wind and geomagnetic index data sets", *Journal of Geophysical Research: Space Physics*, vol. 119, July 2014, pp. 5364–5381.

[18] J. H. Steiger, A. R. Hakstian, "The asymptotic distribution of elements of a correlation matrix: Theory and application", *British Journal of Mathematical and Statistical Psychology*, vol. 35, Issue 2, November 1982, pp. 208–215.