TITLE: Wind Speed Clustering Using Linkage-Ward Method: A Case Study of Khaaf, Iran

AUTHORS: Elnaz AZIZI,Hamed KHARRATISHISHAVAN,Behnam MOHAMMADIIVATLOO,Amin

MOHAMMADPOUR SHOTORBANI

PAGES: 945-954

ORIGINAL PDF URL: https://dergipark.org.tr/tr/download/article-file/792550

Wind Speed Clustering Using Linkage-Ward Method: A Case Study of Khaaf, Iran

Elnaz AZIZI[1] , Hamed KHARRATI-SHISHAVAN[2] , Amin MOHAMMADPOUR SHOTORBANI[3] ,

Behnam MOHAMMADI-IVATLOO[2]

[1] Department of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran

[2] Department of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran

[3] Faculty of Applied Science, University of British Columbia, Kelowna, BC, Canada

Highlights
• Linkage-ward used to analyze the wind clustering approach is speed pattern in Khaaf, Iran.
• Through clustering, the big dataset is compressed into few meaningful representatives.
• The probability of occurrence of each cluster is calculated.
• Analyzing the representatives, instead of the whole dataset, is simple and time-saving.

Abstract

The use of renewable energy for providing electricity is growing rapidly. Among others, wind power is one of the most appealing energy sources. The wind speed has direct impact on the generated wind power and this causes the necessity of wind speed forecasting. For better power system planning and operation, we need to forecast the available wind power. Wind power is volatile and intermittent over the year. For getting better insight and a tractable optimization problem for different decision making problems in presence of wind power generation, it is required to cluster the possible wind power generation scenarios. This article presents probabilistic wind speed clustering prototype for wind speed data of Khaaf, Iran. This region is known as one of the high potential wind sites in Iran and several wind farm projects is planned in this area. The average speed of wind for a ten-minute period measured at height of 40m over a year (2008) is used for clustering. From the result of this research, the most appropriate probabilistic model for the wind speed can be obtained.

NOMENCLATURE

Indices

i, j, k	Index of cluster

Parameters

$N_{Cluster}$	Number of clusters
n_i, n_j, n_k	Number of members withen cluster i, j and k, respectively
d_{ik}, d_{jk}, d_{ij}	The pairwise distances between the clusters i and k, j and k, and i and j, respectively.

Variables

x	Each observation in dataset
$error$	Euclidian error
P_i	Probability of the occurrence of cluster i
c_i	Centroid of cluster i

*Corresponding author, e-mail: a.m.shotorbani@ubc.ca

1. INTRODUCTION

Using renewable sources for providing electricity has different advantages, such as reducing air pollution, cutting global warming emissions, diversifying world power supply, decreasing dependence on coal and other fossil fuels and etc. [1,2]. Renewable energy provides clean and reliable electricity and nowadays they are significant sources of electricity all around the world. For instance, Denmark currently seeks to provide all its energy need just by using renewable energy sources by 2050 [3]. Among others, wind plants generated about 32% of electricity in 2012, and researchers aimed at providing more than 50% of overall demand in 2020 by wind farms [4]. In different countries, researchers try to use different strategies in order to use alternative renewable sources like wind energy and degraded the use of fossil fuels. Wind power electricity has been generated in Iran since 1994.

Among other renewable sources, use of wind energy has the maximum increasing rate. Due to the intermittent nature of wind, it is a fluctuating source of electricity production. One of the main reasons of wind's fluctuation is the wind speed. It is varying in different sites and of course in different hours of the day in one site. Therefore, accurate forecasting and modeling of wind speed are essential to minimize the impact of its intermittency and uncertainty. Different algorithms such as neural networks [5,6] and statistical approaches [7,8] have been applied on measured wind speed data (MWSD) to extract different models. One of the appropriate methods of analyzing the MWSD is clustering which is a well-known method in data mining. By clustering the data, we try to extract the pattern and the informative characteristics, and then approach a probabilistic model of the MWSD [9].

The evolution of wind speed has been studied in [10] based on cluster analysis. This study helps to detect zones with practical surface wind. Similarly, using an evolutionary algorithm [11] looks for the best clustering of synoptic pressure situation and the proper objective function considering the wind speed.

In [12], prediction of wind speed in Kudat, Malaysia, was verified using different methods, including K-means clustering. It was shown that K-means clustering, which segmented the MWSD into given number of clusters, has given more accurate prediction outcomes. It was also figured out that the wind speed in this area is sufficient for operating the wind turbines.

The 60-minutes MWSD were collected during four years and were verified in [13] in order to evaluate the wind energy potential for electricity production in Binalood, Iran. The wind speed in different heights, the soil of the region and the distance to residential area were studied. Based on specifications of this region, Binalood is a proper zone for utilizing wind power and generating electricity during the entire year.

The result of wind speed analyzing plays an important role in wind farms [14-16]. The application of these results can be listed as:

- Selecting the proper site for installing wind turbines,
- Predicting the optimum size of the wind machine that is used for a special site,
- Scheduling the connection or disconnection of wind turbines or conventional generators,
- Obtaining low spinning reserve and optimal operating cost,
- The optimal planning of wind-distributed generators and switched capacitor banks.

Following the previous discussion, this research aims to analyze the MWSD, and extract the optimal number of clusters and the most appropriate representative for each cluster. At the end, the probability of occurrence for each cluster is computed.

The reminder of this research is organized as follows: Section 2 presents motivation and methodology. Section 3 depicts the results and finally Section 4 concludes the paper.

2. MOTIVATION AND METHODOLOGY

Due to the increasing electricity demand and on the other hand devastating effects of fossil fuels on the environment, limited natural supplies, harnessing the renewable sources and using them in different applications is getting interest all around the world. Given that Iran has practical zones for installation wind turbines, verifying the features and potential of wind at different sites and choosing the most proper one for wind farms is important and effective gets great importance. One of the main features that effects the wind power generation is wind speed. Analyzing and forecasting wind speed play a principle role in wind generation and power consumption demands. Wind speed is noisy, irregular, and easily affected by weather and geographical factors that drive the fluctuations of wind speed and increase the difficulties of precise forecasting. In order to obtain an precise estimated wind speed for a specific time, the most appropriate cluster must be considered for an optimal decision making in different regions.

Due to the fact that the installation of wind turbines and the restructuring of power plants in Iran are in progress, this is urgent in the operational planning and cost control in the power plants.

2.1. The Geographical Location of Khaaf

The studied zone, Khaaf, is located in the southeast of the province of Razavi Khorasan which is located in north-east of Iran (Figure 1). This province ranks fourth among the large provinces of Iran. Its area is approximately 117,769 km2. In terms of climate categorization, it has semi-arid climate and located in a moderate rainfall area. Khaaf is located in Razavi Khorasan province in the geographical coordination 34º34'N 60º08'E, in the east border strip of Iran. Its population is over 100,000 and covers an area of 9797 km². Iranian Renewable Organization (SUNA) has analyzed the MWSD of different areas in Iran. Figure 2 displays the wind speed status at the height of 50 meters [17].

In this paper, a 10-minute period MWSD of Khaaf at the heights of 10 m, 30 m and 40 m which were collected at 2008 is considered. By analyzing the MWSD in different heights, the following characteristics are captured, given in Table 1.

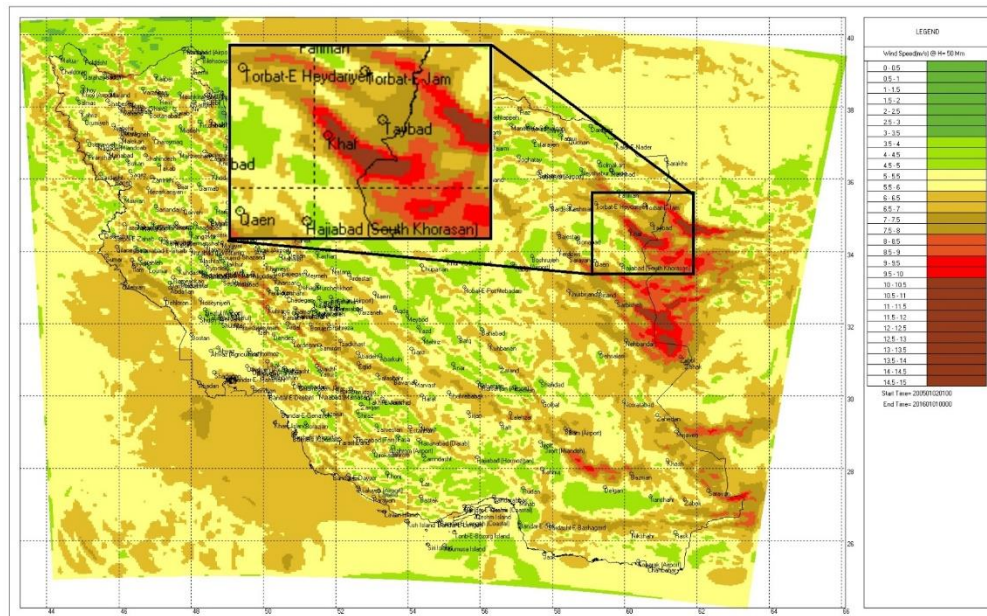**Figure 1.** *Map of Khaaf in Razavi Khorasan province of Iran [17]*

Figure 2. *The wind speed map at the height of 50 meters [18]*

Table 1. *Wind characteristics*

Data property	Height (m)		
	10	30	40
Mean Speed (km/h)	9.37	10.68	11.13
Max Speed (km/h)	34.30	35.90	36.30
Min Speed (km/h)	0.10	0.10	0.10
Median Speed (km/h)	9.60	11.00	11.70

2.2. Clustering

Nowadays, traditional algorithms are not efficient for analyzing ever-increasing volume of data in smart grid. Therefore, new approaches of data mining, like "clustering" can be utilized for analyzing these data. Because of the amount of data, traditional approaches are difficult to use for analyzing them. Therefore, [19]. Clustering partitions objects into optimal number of groups on the basis of measures of similarity among those objects. Clustering algorithms can be divided into two main groups, "exclusive clustering" and "overlapping clustering". In exclusive clustering, each datapoint is assigned to just one cluster. However, in overlapping clustering, a datapoint belongs to all clusters with different percentage of membership. Exclusive clustering can be further categorized into two main algorithms:

I.   Partitioned clustering: In this approach without defining any hierarchical structure, data points are divided into pre-determined number of clusters. The most common algorithm of portioned clustering is K-means.
II. Hierarchical clustering: This algorithm builds a hierarchy of clusters with sequence of nested partitions. This structure can be bottom-up (agglomerative) or up-bottom (divisive).

There are different ways to cluster *N* data points in to optimal number of groups. Linkage-Ward clustering as the most common and accurate method of agglomerative clustering is used in this research. The accuracy of this method in comparison with K-means clustering is high but the computation of this method is more than K-means clustering, which will be discussed in section 3.4. (R2) The basic algorithm of the Linkage-Ward clustering method is shown in Table 2. In this method, the dissimilarity between the

clusters is calculated based on the distances between the clusters' centroids, given as (1). A larger distance magnitude represents a higher dissimilarity value between the clusters [20],

$$d_{i+j,k} = ad_{ik} + ad_{jk} + bd_{ij} + c|d_{ik} - d_{jk}|, \tag{1}$$

where a and b are defined as (2), and $c=0$ in the Linkage-Ward clustering method,

$$a = \frac{n_i + n_k}{n_i + n_j + n_k}, b = -\frac{n_j + n_k}{n_i + n_j + n_k}. \tag{2}$$

Then, the clusters which result in the lowest increase in the cost function of (1), are combined together. In this method, the objective function is the sum of squares from the points to the centroids of the clusters. Figure 3 depicts its flowchart. Bu clustering a big amount of data and computing the centroids of clusters, we compress a big amount of data to few centroids with decreases the computation time [20].

Table 2. *Pseudo code of Linkage-Ward clustering Algorithm*

Step	Tasks
I	Give a dataset with n objects, consider each object as a cluster.
II	Compute the distance matrix based on the cost function.
III	Merge clusters with minimum cost function.
IV	If number of clusters=$N_{cluster}$ stop. Otherwise, go back to step 3.
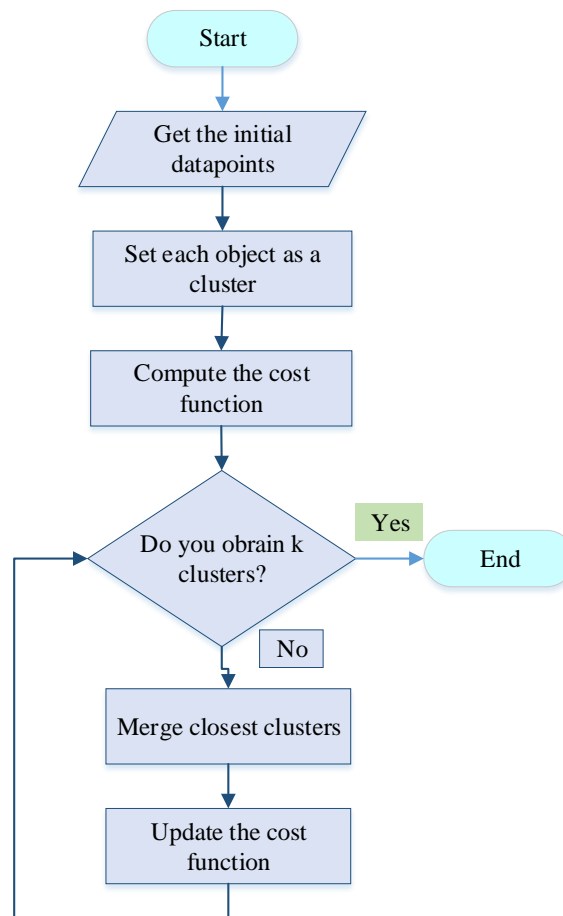
**Figure 3**. *Linkage-Ward clustering step-by-step algorithm*

2.3. Probability of Occurrence

The probability of occurrence of each cluster is calculated as:

$$P_i = \frac{n_i}{\sum_{j=1}^{N_{cluster}} n_j}.$$

(3)

3. SIMULATION RESULTS

The MWSD of Khaaf, Iran, which was collected at 2008 is used for analysis purpose, which is measured each 10 minutes. Clustering can be applied on different features of wind speed like its average value, and minimum and maximum values of wind speed in different heights. Here, we focus on the average value of wind speed in 40 m height over 10-minute intervals. Figure 4 displays the average values of wind speed for 50 days of a year. These days are chosen randomly.
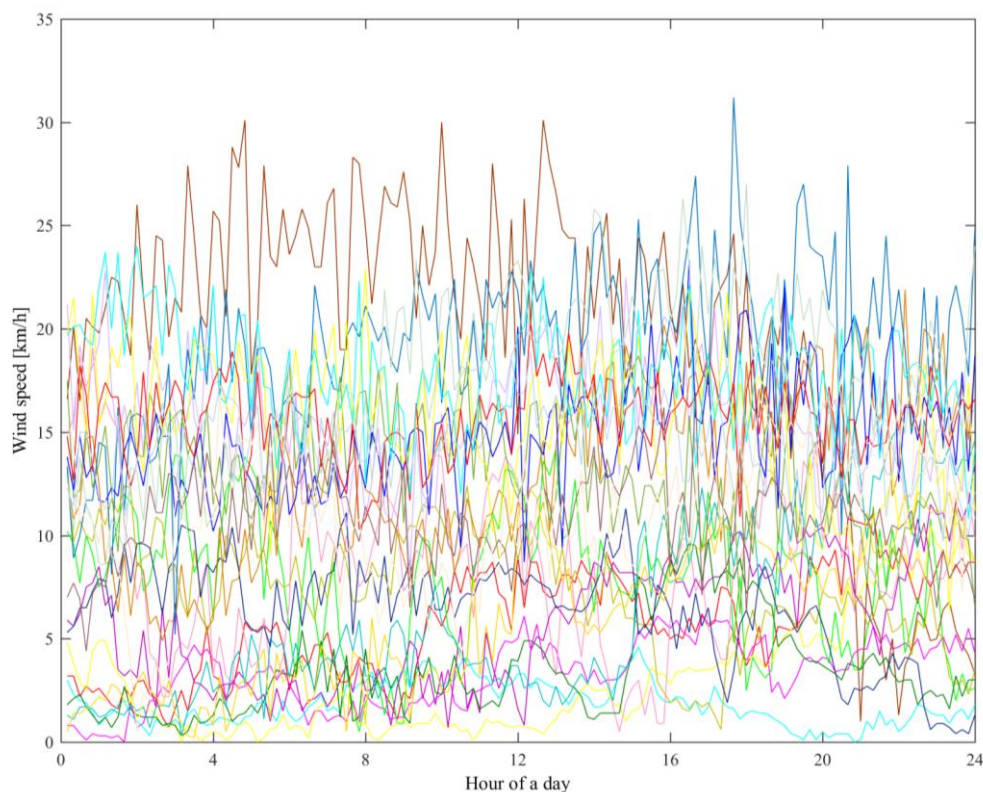

Figure 4*. The MWSD for 50 days at Khaaf, Iran*

In this research, the studied data is verified considering linkage-ward clustering. There are three main topics about clustering:

3.1. Number of Clusters

Choosing a large number of clusters does not cause dimension reduction. Contrarily, choosing small number of clusters brings out groups of dissimilar objects. Therefore, choosing an optimal number of clusters is important. In this study, the optimal number of clusters is chosen by calculating the error between each cluster's centroid and its members, using (4)

$$error = \sum_{i=1}^{N_{Cluster}} \sum_{j=1}^{n_j} |x_j - c_i|.$$

(4)

Figure 5 displays this error for different number of clusters. It is figured out that the minimum error occurred at 4 clusters which is the most proper number of clusters for this dataset.
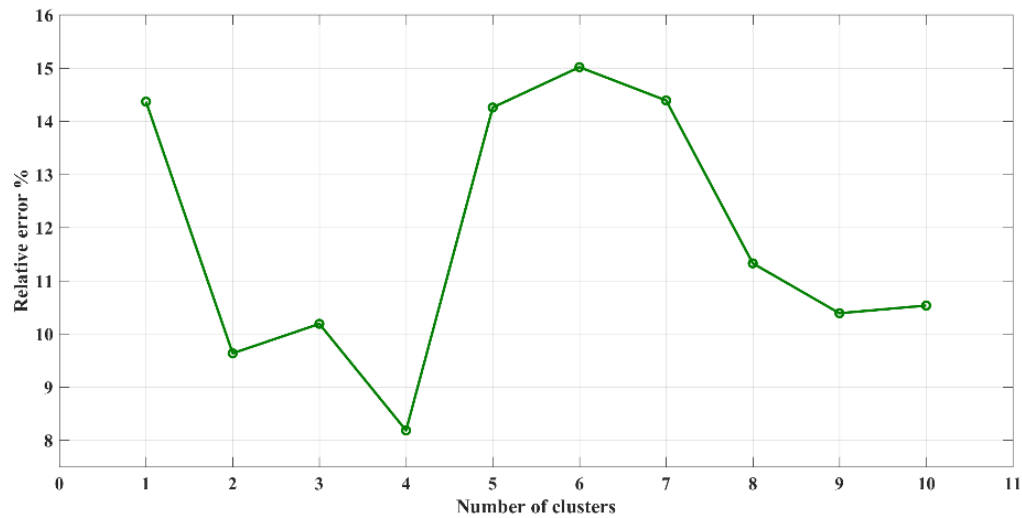

Figure 5. *The error for different numbers of clusters*

3.2. Probability of Occurrence of each Cluster

Table 3 and Figure 6 illustrate the probability of occurrence of each cluster. Cluster 2 has the highest probability of occurrence, which means that it must be used for appropriate planning of wind farms.

Table 3. *Probability of occurrence of each cluster*

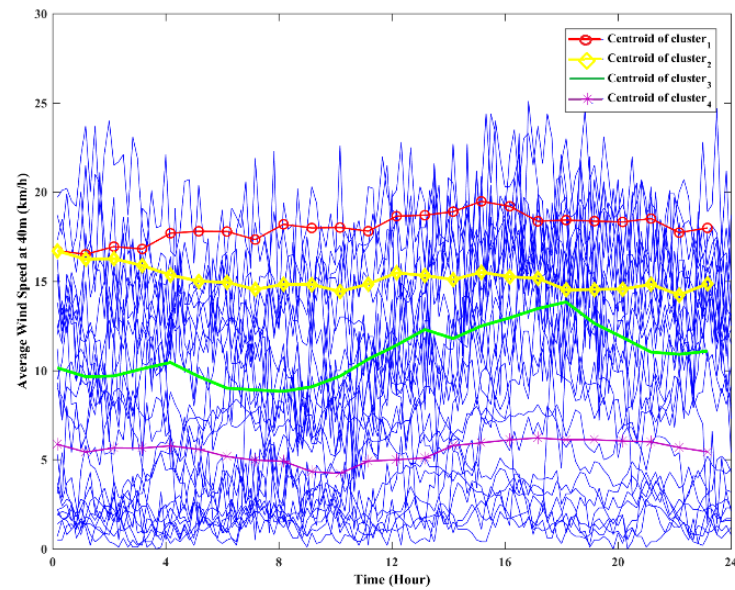|	*No. 1*	*No. 2*	*No. 3*	*No. 4*
Probability	0.1974	0.3776	0.1931	0.2317
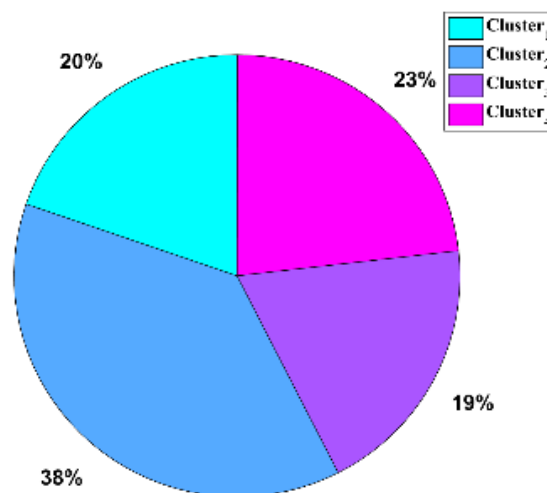
Figure 6. *Probability of occurrence of each cluster*

3.3. Defining Centroids of Clusters

In this paper, the mean of the members within a cluster is calculated and defined as a representative for each cluster using (3). Figure 7 and Table 4 show the centroids of each cluster.

Figure 7. *Cluster centers versus the dataset*

Table 4. *Centroids of clusters*

Time	Centroids of clusters			
	No. 1	No. 2	No. 3	No. 4
0	16.71739	16.70227	10.14666	5.85925
1	16.50869	16.23522	9.651111	5.42592
2	16.94130	16.26477	9.691111	5.66111
3	16.80869	15.87954	10.09333	5.65000
4	17.69130	15.35000	10.44888	5.77222
5	17.80217	15.00000	9.662222	5.58148
6	17.78913	14.93977	9.006666	5.17222
7	17.33478	14.54318	8.908888	4.97592
8	18.19130	14.82272	8.828888	4.91296
9	17.98695	14.81818	9.086666	4.32407
10	18.01086	14.42159	9.682222	4.22962
11	17.80652	14.84545	10.63555	4.92037
12	18.64782	15.46590	11.41555	4.99444
13	18.70217	15.32386	12.29333	5.10000
14	18.90000	15.09318	11.79111	5.78888
15	19.47826	15.50113	12.49333	5.94629
16	19.20217	15.25454	12.94666	6.11296
17	18.36086	15.16590	13.47111	6.21851
18	18.43043	14.51818	13.83777	6.11481
19	18.36739	14.53977	12.63777	6.11851
20	18.32391	14.55681	11.85333	6.04814
21	18.513043	14.82613	11.03333	6.00185
22	17.723913	14.23295	10.90444	5.68148
23	17.989130	14.85227	11.08444	5.44074

3.4. Comparison with K-means Clustering

In this part we considered the most common method of partitioned clustering, k-means. The main drawback of this method is its dependency on initial points. Different initial datapoints for each cluster can result in different final clusters' centers. The accuracy and time consuming of k-means and Linkage-Ward clustering are illustrated in Tables 5 and 6, respectively.

Table 5. *Time of clustering in different methods*

Method	K-means	Linkage-ward
Time (s)	0.37	0.52

Table 6. *Relative error between each cluster members and their centers in different methods*

Method	K-means	Linkage-ward
Relative Error	10.3%	8.2%

There is a trade-off between the time of clustering and its accuracy. Even though the computation time of K-means clustering is less than Linkage-ward method, but linkage-ward clustering resulted in more accurate results. In our study, obtaining exact behavior of wind speed is more important than time of calculation.

4. CONCLUSIONS

Wind energy, as a natural and clean source, can be converted to mechanical or electrical energy. Therefore, identifying the most proper sites for deployment of wind farms is essential. One of the critical features affecting the output wind energy potential is the wind speed profile. The following specifications make Khaaf a proper region for wind power generation:

1) High average wind speed in different heights of this region,
2) Existence of practical zone for building wind farms.

Analyzing the MWSD plays a critical role in modeling wind power. In this study the Linkage-Ward clustering method is used for analyzing the MWSD in Khaaf and grouping this data. Based on the Euclidian distance, it is figured out that the most proper number of clusters is 4. The mean of the cluster members is considered as centroid of them. For analyzing the whole dataset, researchers can analyze the centroids of clusters which increases the speed of analyzing. The results can be used for precise planning of the expanding wind farms, which will provide a drastic reduction in fossil fuel consumption for energy production.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Richardson, D.B., "Electric vehicles and the electric grid: A review of modeling approaches, Impacts, and renewable energy integration", Renewable and Sustainable Energy Reviews, 19: 247-254, (2013).

[2] Twidell, J., Weir, T., Renewable energy resources, 3rd Ed., Routledge, New York, USA, (2015).

[3] Meibom, P., Hilger, Madsen, K.B., H., Vinther, D., "Energy comes together in Denmark: The key to a future fossil-free Danish power system", IEEE Power and Energy Magazine, 11: 46-55, (2013).

[4] Marinelli, M., Sossan, F., Costanzo, G.T., Bindner, H.W., "Testing of a predictive control strategy for balancing renewable sources in a microgrid", IEEE Transactions on Sustainable Energy, 5: 1426-1433, (2014).

[5] Fadare, D., "The application of artificial neural networks to mapping of wind speed profile for energy application in Nigeria", Applied Energy, 87: 934-942, (2010).

[6] Mellit, A., Kalogirou, S.A., "Artificial intelligence techniques for photovoltaic applications: A review", Progress in Energy and Combustion Science, 34: 574-632, (2008).

[7] Erdem, E., Shi, J., "ARMA Based Approaches for Forecasting the Tuple of Wind Speed and Direction", Applied Energy, 88: 1405-1414, (2011).

[8] Morales, J., Minguez, M.R., Conejo, A.J., "A methodology to generate statistically dependent wind speed scenarios", Applied Energy, 87: 843-855, (2010).

[9] Chicco, G., "Overview and performance assessment of the clustering methods for electrical load pattern grouping", Energy, 42: 68-80, (2012).

[10] Gómez, G., Cabos, W.D., Liguori, G., Sein D., Lozano Galeana, S., Fita, L., Montávez, J.P., Domínguez, M., "Characterization of the wind speed variability and future change in the Iberian Peninsula and the Balearic Islands", Wind Energy, (2015).

[11] Carro-Calvo, L., Salcedo-Sanz, S., Prieto, L., Kirchner-Bossi, N., Portilla-Figueras, A., Jiménez-Fernández, S., "Wind speed reconstruction from synoptic pressure patterns using an evolutionary algorithm", Applied Energy, 89: 347-354, (2012).

[12] Goh, H., Lee, S., Chua, Q., Goh, K., Teo, K., "Wind energy assessment considering wind speed correlation in Malaysia", Renewable and Sustainable Energy Reviews, 54: 1389-1400, (2016).

[13] Mostafaeipour, A., Sedaghat, A., Ghalishooyan, M., Dinpashoh, Y., Mirhosseini, M., Sefid, M., Pour-Rezaei, M., "Evaluation of wind energy potential as a power generation source for electricity production in Binalood, Iran", Renewable Energy, 52: 222-229, (2013).

[14] Mehmood, K.K., Kim, C., Khan, S.U., Haider, Z.M., "Unified Planning of Wind Generators and Switched Capacitor Banks: A Multiagent Clustering-Based Distributed Approach." IEEE Transactions on Power Systems, 33: 6978 - 6988, (2018).

[15] Khodayar, M., Wang, J., Manthouri, M., "Interval Deep Generative Neural Network for Wind Speed Forecasting", IEEE Transactions on Smart Grid , 1-1, (2018).

[16] Lydia, M., Kumar, S.S., Selvakumar, A.I., Kumar, G.E.P., "Wind Farm Power Prediction Based on Wind Speed and Power Curve Models", Intelligent and Efficient Electrical Systems. Springer, Singapore, 446: 15-24, (2018).

[17] Khorasan.ir, Razavi Khorasan Province Portal, http://www.khorasan.ir.

[18] http://www.satba.gov.ir/fa/regions/windatlas.

[19] Azizi, E., Kharrati, H., Mohammadi-Ivatloo, B., Shotorbani, A.M., "Clustering Electricity Big Data for Consumption Modeling Using Comparative Strainer Method for High Accuracy Attainment and Dimensionality Reduction", Journal of Energy Management and Technology, 3.1, (2019).

[20] Han, J., Kamber, M., Pei, J., Data mining: concepts and techniques, 3rd Ed., Elsevier, Waltham, USA, (2011).