BULETINUL INSTITUTULUI POLITEHNIC DIN IAŞI
Publicat de
Universitatea Tehnică „Gheorghe Asachi” din Iaşi
Tomul LX (LXIV), Fasc. 2, 2014
Secţia
ELECTROTEHNICĂ. ENERGETICĂ. ELECTRONICĂ

A WIND ENERGY PROFILING METHOD BASED ON UNSUPERVISED LEARNING TECHNIQUES

BY

FLORINA SCARLATACHE and GHEORGHE GRIGORAŞ[*]

“Gheorghe Asachi” Technical University of Iaşi
Faculty of Electrical Engineering

Received: January 6, 2014
Accepted for publication: March 3, 2014

Abstract. Due to the increasing integration of renewable energy resources into electrical networks, it is necessary to know a lot of information’s regarding their operation characteristics. Considering that some renewable energy sources, like small hydropower plants or cogeneration power plants, can be clearly characterized by representative generation profiles, the wind power sources represent one of the most unpredictable energy sources. Therefore, this paper presents a clustering based method used in obtaining of representative generation profiles corresponding wind farms.

Key words: clustering techniques; wind sources; representative generation profiles.

1. Introduction

In order to fulfill its 2020 climate and energy targets, (Communication From The Commission, 2014), the EU has to accelerate its ambition to create a single European power market, based on renewable electricity, an EU Super Grid as well as a smart grid in order to facilitate an intelligently and efficiently interconnected electric energy system of both centralized and decentralized renewable energy sources. Especially in the period leading up to 2020, Europe has to invest in new energy production capacity to replace ageing plants while meeting future demand. Integration of renewable energy resources can present

[*]Corresponding author : *e-mail*: ggrigor@ee.tuiasi.ro

challenges for the controllability and dispatchability of these sources and for operation of the electric energy system and has the potential to respond to global sustainability and environmental, safety, social, and economic goals.

The changes brought by the renewable energy sources are certain to have a significant impact on system performance and efficiency and necessitate advances in the planning and operation of electric system.

Integration of renewable sources into an electric power system is the bighest challenge for specialists in Europe and for the Romanian power system. The development of renewable energy sector in Romania is based on the existence of a important and diversified potentially regarding the renewable energy sources: hydropower which provides more than 35.7% of the electric energy production in the entire country, wind power with a installed capacity of 17%, solar energy with 12%, biomass (agricultural and forest origin) that representing 7% of primary energy demand and 50% of potential resources. Romania is ranked on 10 in the top of most attractive countries in the world on investment in wind power and on 13th in total renewable energy sources, into a classification compiled by “Ernst & Young”, which comprises 40 countries, including United States, Germany, China and United Kingdom, (Renewable Energy, 2014).

The wind energy and the other renewable energy sources are conditioned by the restrictions on the safe operation of power systems and also a major restriction in cases of lack of consumption, lower consumption or discontinuous. It is known that the electric energy generated from wind power can be highly variable at several different timescales: hourly, daily, or seasonally. However, wind is always in constant supply somewhere, making it a dependable source of energy because it will never expire or become extinct.

Based on the unpredictability of the wind energy, a new approach is proposed in this paper based on similarities that exist between daily generation patterns of the wind farm and their grouping into representative clusters. By knowing the generation profile, the companies responsible for the wind farm management can estimate in function of the wind speed, the representative generation profile that can be associated for a characteristic day, into a specific zone.

In the literature different techniques have been used in profiling process, but most of them were implemented to solve the problems from power systems. A review of the literature revealed two types of methods: statistical methods (Lubosny, 2003; Carter-Brown, 1999) and methods based on artificial intelligence techniques: fuzzy logic (Lo *et al*., 2004; Gerbec *et al*., 2004), neural networks (Sarlak *et al*., 2012), data mining (Nizar, 2006), clustering, (Chicco, 2012; Mahmoudi-Kohan, 2009; Grigoraş *et al*., 2013).

In this paper, an efficient profiling method based on unsupervised learning techniques is proposed. The method uses the *k*-means clustering algorithm for obtaining the representative generation profiles of a wind farm from southeastern of Romania.

2. Clustering Techniques

2.2. *k*-Means Methods

Clustering is a division of data into groups or clusters of similar objects. Each group, called *cluster*, consists of objects that are similar between themselves and dissimilar to objects of other groups. Data modeling puts clustering in a historical perspective encountered in statistics, mathematics, and numerical analysis. The clusters correspond to hidden patterns, the search for clusters is unsupervised learning, and the resulting system represents a data concept. Therefore, clustering is unsupervised learning of a hidden data concept.

Traditionally clustering techniques are divided in hierarchical and partitioning methods, (Cârțină *et al*., 2005; Grigoraş *et al*., 2011; Yatskiv *et al*., 2004; Jain *et al*., 1999).

Hierarchical clustering is subdivided into agglomerative and divisive. While hierarchical algorithms build clusters gradually, partitioning algorithms learn clusters directly. So, the partitioning methods discover the clusters by iteratively relocating points between subsets, or identify the clusters as zone highly populated with data. These methods are categorized into probabilistic clustering and *k*-means methods. Such methods tend to build clusters of proper convex shapes. In *k*-means case a cluster is represented by its centroid, which is a mean, usually weighted average, of points within a cluster.

The *k*-means algorithm is the most popular clustering tool used in scientific and industrial applications, (Berkhin www). The name comes from representing each of K clusters C_j by the mean (or weighted average), c_j, of its points, the so-called *centroid*.

The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid:

$$\min(E) = \min\left(\sum_{i=1}^{K}\sum_{x\in C_i} d\left(x, z_i\right)\right), \tag{1}$$

where z_i is the center of cluster C_i, while $d(x, z_\mathrm{i})$ is the Euclidean distance between a point, x, and z_i.

The criterion function, E, attempts to minimize the distance of each point from the center of the cluster to which the point belongs. More specifically, the algorithm begins by initializing a set of K cluster centers. Then, it assigns each object of the dataset to the cluster whose center is the nearest, and recomputed of the centers. The process continues until the centers of the clusters stop changing.

Evaluating and assessing the results of the *k*-means algorithm represents the main subject of cluster validity. Internal cluster validation tests are more popular in practice of cluster analysis. From these, the test based on the

Silhouette Global Index calculation is one of the most used (Yatskiv *et al.*, 2004; Halkidi *et al.*, 2001; Grigoraş *et al.*, 2011). This calculates the silhouette width for each sample, average silhouette width for each cluster and overall average silhouette width for a total data set. Using this approach each cluster could be represented by so-called silhouette, which is based on the comparison of its tightness and separation.

3. Determination of Representative Generation Profiles

An approach based on *k*-means clustering algorithm to determinate the generation profiles for a wind farm is proposed. The shape of operational characteristics is influenced by the type of day or season of year. Because a large number of operational characteristics create problems in handling them, they could be grouped into coherent groups, seeing that some similarities between these. For each coherent group a representative generation profile is determined, (Garcia *et al.*, 2005; Scarlatache *et al.*, 2012).

Each representative generation profile is represented by a vector x_i = {x_{ih}, h = 1,…, T} for i = 1, …, K, and the comprehensive set of representative generation profiles is contained in the set P = {x_i, i = 1, …, K}. The time scale along the day is partitioned into T time intervals of duration, Δt_h , for h = 1,…, T. The variables used in the calculations are assumed to be represented as constant (average) values within each time interval.

The algorithm is based on the profiling process. The flow chart of the algorithm is presented in Fig. 1. The major steps are:

1. *Measurements*: In this step a sample of the set of operational characteristics is identified. Finally the collected data is gathered in a large database.

2. *Data cleaning and pre-processing*: Considering a large number of measurements from a large geographic area during a considerable period of time, it can detect different kind of problems that will affect the quality of the database. The result will be a very large database with problems like noise, missing values and outliers. These data (after being cleaned, pre-processed and reduced) are used in clustering process.

3. *Classification*: To realize this classification, the *k*-means algorithm is used. Each generation profile is determined the normalized characteristics using a suitable normalizing factor (average power, energy over the surveyed period, etc.):

$$z_h^{(j)} = \frac{x_h^{(j)}}{X^{(j)}}, \ (j = 1,\dots, N,\ h = 1,\dots, T), \tag{4}$$

where: $z_h^{(j)}$ is the normalized value, $x_h^{(j)}$ – the measured value, $X^{(j)}$ – the normalizing factor over the surveyed period and N – the number of hourly generation patterns.

4. *Determination of representative generation profile*: Using a hierarchical clustering method, the normalized wind characteristics are refined so as to desist at the unrepresentative characteristics. The representative generation profile for each cluster is obtained by averaging the values for each hour. These coefficients lead us to the normalized representative generation profiles.

5. *Assignation*: Finally, to the each cluster is made the assignation of a representative generation profile.

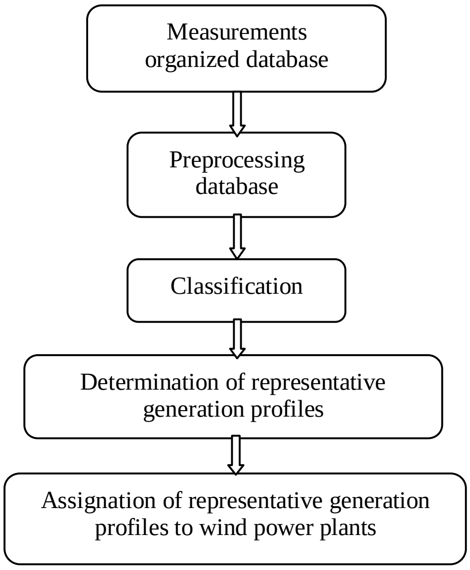

Fig. 1 – Flow-chart of the profiling process.

4. Case Study

In order to show the features of clustering based method, a database that contains the generation profiles a wind farm is considered. These generation profiles were recorded in autumn season (N = 90 days) of 2013 year. The time interval is defined by taking hourly steps within a day, T = 24 and Δt_h = 1 h. The normalization of these profiles was made in relation with daily wind energy production. The normalized values of generation profiles are presented in Fig. 2. Further, the optimal number of clusters was determined applying the algorithm proposed by Halkidi *et al.*, (2001). First of all, the maximum number of clusters $K_{\max}$ was calculated ($K_{\max} = \sqrt{N} = 9$). Then, for the set of normalized generation profiles from the data base, the k-means clustering method with given K ($2 \leq K \leq K_{\max}$) is used. Finally, the silhouette global coefficient is determined to assess the partition quality. Because the silhouette global coefficient registered the highest value for K = 3, this represents the optimal solution for clustering process (Fig. 3). For this solution, the silhouette plot is presented in Fig. 4.

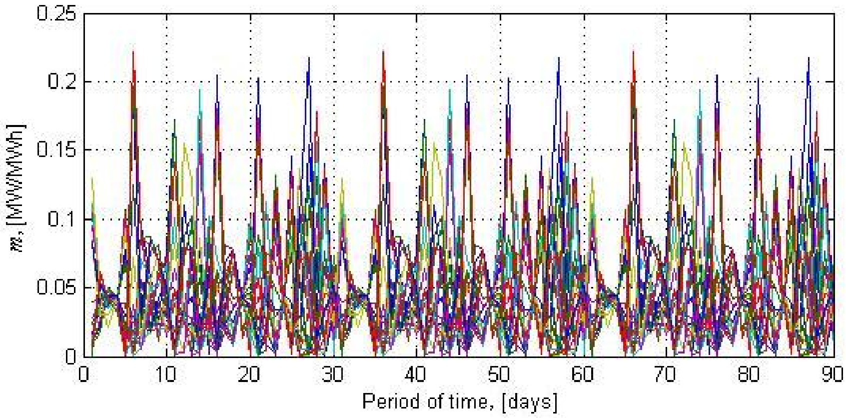

Fig. 2 – The normalized generation profiles of the wind farm, autumn season.

Fig. 3 – The silhouette global coefficient for different values of number of clusters.

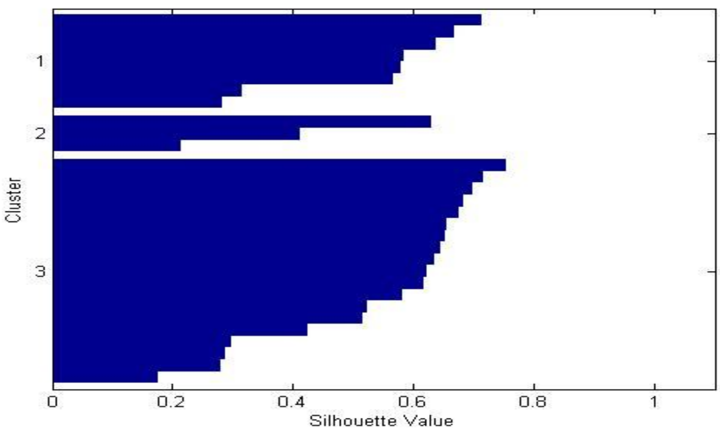

Fig. 4 – The silhouette plot for K = 3.

After aggregation of normalized generation profiles of each cluster, the representative profiles were determined. The representative generation profiles of the wind farm corresponding to three obtained clusters (C1, C2, and C3), were obtained by averaging the values for each hour. The graphical representation of these profiles is given in Figs. 5,…,7.

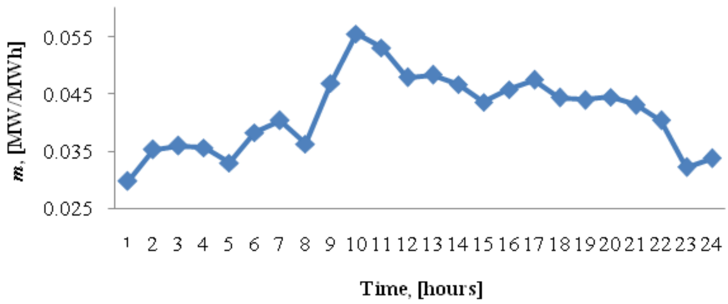

Fig. 5 – Representative generation profile for cluster C1.

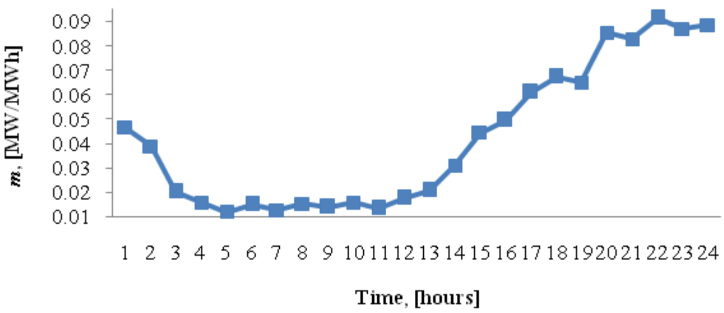

Fig. 6 – Representative generation profile for cluster C2.

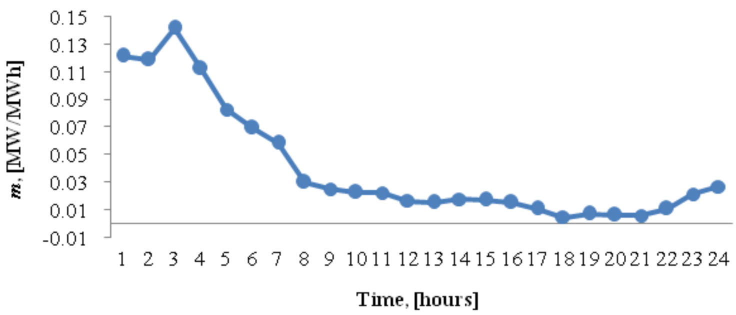

Fig. 7 – Representative generation profile for cluster C3.

The results after applying *k*-means clustering method, allow to analyze and to interpret the structure of each representative generation profiles obtained. Therefore, the representative profile corresponding to cluster C1 accounted 33%

of the total generation profiles registered, with a number of 30 profiles, the cluster C2, 23%, including a number of 21 profiles, and cluster C3 accounted 44% with 39 profiles.

Analysing these results, it can see that a classification of the wind operation characteristics is useful to view the optimal operation and planning of a power system. The representative profile for cluster C1 corresponds to the best operation, taking into consideration that has a full generation operation state all the day and the other two representative generation profiles for clusters C2 and C3 correspond to partial generation operation state.

5. Conclusion

In this paper, *k*-means clustering method was proposed to determine the representative generation profiles for a wind farm. The database contains the generation profiles registered for three months. The obtained results describe very well the operation states of a wind farm and can be used with success to estimate, in function of the wind speed, the representative generation profile that can be associated for a certain day, into a specific zone in order to make a short-term forecast.

REFERENCES

Berkhin P., *Survey of Clustering Data Mining Techniques. Technical Report, Accrue Software*. San Jose, California, [Online]. Available: http://citeseer.nj.nec.com/berkhin02survey.html.

Carter-Brown C.G., *Load Profile Modeling for Integrated Energy Planning*. Proc. of Domestic Use of Electrical Energy Conf., Cape Town, 1999, 13-18.

Cârţină G., Grigoraş G., Bobric E.C., *Tehnici de clustering in modelarea fuzzy. Aplicații în electroenergetică*. Casa de Editură Venus, Iaşi, 2005.

Chicco G., *Overview and Performance Assessment of the Clustering Methods for Electrical Load Pattern Grouping*. Energy, **42**, *1*, 68-80 (2012).

Garcia-Escudero L.R., Gordaliza A., *A Proposal for Robust Curve Clustering*. J. of Classification, **22**, 185-201 (2005).

Gerbec D., Gašperič S., Šmon I., Gubina F., *Determining the Load Profiles of Consumers Based on Fuzzy Logic and Probability Neural Networks*. IEE Proc. Gener., Transmiss. a. Distrib., **151**, 95–400 (2004).

Grigoraş G., Cârţină G., *The Impact of the Fuzzy Modeling in Electric Distribution Systems*, Lambert Academic Publ., Saarbrűcken, Deutschland, 2011.

Grigoraş G., Istrate M., Scarlatache F., *Electrical Energy Consumption Estimation in Water Distribution Systems Using a Clustering Based Method*. 5th Internat. Conf. Electron., Comp. a. Artificial Intell., Piteşti, 2013.

Halkidi M., Batistakis Y., Vazirgiannis M., *On Clustering Validation Techniques*. J. of Intell. Inform. Syst., **17**, *2/3*, 107-145 (2001).

Jain A.K., Murty M.N., Flynn P.J., *Data Clustering: A Review*. ACM Comp. Surveys, **31**, *3*, 264-323 (1999).

Lo K.L., Zakaria Z., *Electricity Consumer Classification Using Artificial Intelligence*. Proc. of the 39th Internat. Univ. Power Engng. Conf., Bristol, United Kingdom, 2004.

Lubosny Z., *Wind Turbine Operation in Electric Power Systems*. Springer-Verlag, Berlin, 2003.
Mahmoudi-Kohan N., Moghaddam M.P., Bidaki S.M., *Evaluating Performance of WFA K-means and Modified Follow the Leader Methods for Clustering Load Curves*. Proc. IEEE/PES Power Syst. Conf. and Exposition, Seattle, 2009.
Nizar A.H., *Load Profiling and Data Mining Techniques in Electricity Deregulated Market*. Power Engng. Soc. General Meeting, IEEE/PES, 2006.
Sarlak M., Ebrahimi T., Karimi Madahi S.S, *Enhancement the Accuracy of Daily and Hourly Short Time Load Forecasting Using Neural Network*. J. of Basic a. Appl. Sci. Res., **2**, *1*, 247-255 (2012).
Scarlatache F., Grigoraş G., Chicco G., Cârțină G., *Using K-Means Clustering Method in Determination of the Optimal Placement of Distributed Generation Sources in Electrical Distribution Systems*. Internat. Conf. on Optim. of Electrical a. Electr. Equip., 953-958, 2012.
Yatskiv I., Gusarova L., *The Methods of Cluster Analysis Results Validation*. Proc. of Internat. Conf. RelStat'04, Riga, Latvia, 2004, 75-80.
* * * *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions* (A policy framework for climate and energy in the period from 2020 to 2030), Brussels, 22.01.2014. Available: http://ec.europa.eu/energy/doc/2030/com_2014_15_en.pdf.
* * * *Renewable Energy Country Attractiveness Indices*. Ernst & Young; Available: http://www.ey.com/Publication/vwLUAssets/RECAI_40_February_2014/$FILE/EY_RECAI%2040_Feb%202014.pdf, 2014.

METODĂ DE PROFILARE A ENERGIEI EOLINE BAZATĂ PE TEHNICI DE ÎNVATARE NESUPRAVEGHEATĂ

(Rezumat)

Integrarea într-un număr tot mai mare a surselor de energie regenerabilă în rețelele electrice presupune cunoaşterea unor informații suplimentare referitoare la regimul lor de încărcare. Având în vedere că unele surse de energie regenerabilă, cum ar fi micro-hidrocentralele sau centralele electrice de cogenerare, sunt caracterizate în mod clar de profile reprezentative de generare, sursele de energie eoliană reprezintă una din sursele de energie cele mai imprevizibile. Prin urmare, în această lucrare se propune o metodă bazată pe tehnici de „clustering” în scopul obținerii unor profile reprezentative de generare corespunzătoare parcurilor eoliene.