

## Pattern Aggregation of Wind Energy Conversion Technologies using Clustering Analysis

Paula Odete Fernandes

Polytechnic Institute of Bragança  
Campus de Santa Apolónia, Apartado 1134  
5301-857 Bragança, Portugal

UNIAG, Applied Management Research Unit & NECE,  
Research Unit in Business Sciences  
e-mail: pof@ipb.pt

Ângela Paula Ferreira

Polytechnic Institute of Bragança  
Campus de Santa Apolónia, Apartado 1134  
5301-857 Bragança, Portugal

CISE, Electromechatronic Systems Research Centre,  
University of Beira Interior, Covilhã, Portugal  
e-mail: apf@ipb.pt

**Abstract**— The main objective of this research is the identification of homogeneous groups within a set of wind farms of a major wind energy promoter in Portugal, based on two multivariate analyses: Hierarchical Cluster Analysis and K-means Clustering, using two independent variables, capacity factor and net production, both per year. K-means Clustering output provides the same results as the Hierarchical Cluster Analysis. Outputs allowed the identification of three homogenous groups of wind farms: (1) medium installed capacity and asynchronous generator based technologies, (2) high installed capacity and direct driven synchronous generator based technology and (3) low installed capacity with no differentiation on the technology concept, but including the wind farms with the higher capacity factors.

**Keywords**- Wind farms; Wind turbine generators; Hierarchical Cluster Analysis; K-means Clustering.

### I. INTRODUCTION

In order to cope with the worldwide climate change and the expected increase of the electricity demand, low-carbon and energy efficient technologies are required, mainly based on renewable energy sources. In this scenario, wind energy plays an important role, capable of decarbonizing the power sector, by reducing the usage of fossil fuels. Wind energy is a clean and environmentally friendly technology. Its renewable character and the fact it does not pollute during the operational phase makes it one of the most promising energy source in reducing environmental problems at both global and local levels.

According to Global Wind Energy Council (GWEC), installed wind power capacity has grown to cumulative worldwide installation level of 318 GW, with 35,5 GW alone installed in 2013. Europe continues to lead the world in total installed capacity, despite the fact that other markets (USA, India, China, e.g.) have also launched in recent years. Portugal accounts for about five percent of the wind energy installed capacity of the European Union, with approximately 4,8 GW of accumulated installed capacity in 2013 which is capable to generate about 15% of the electrical energy consumption [1].

Despite the advantages inherent to a renewable energy source, wind energy has also some drawbacks. The small

power density of the wind leads to wide and material extensive turbines, thereby hindering the on-site assembly and the electrical infrastructure. Concerning the prime source, wind is stochastic in nature and essentially ruled by random meteorological changes. Due to its intermittent and unpredictable behavior, wind energy systems do not have the ability to produce electrical energy following load requirements which implies structural changes in power systems as, for instance, the usage of storage systems and/or coupling hydro and wind systems to smooth the output pattern [2]. The inherent variability of wind power is also raising concerns regarding the reliability and cost-effectiveness of the transmission and distribution power systems while supporting large wind farms [3].

Wind energy conversion systems are coming-of-age: after the oil crisis of the 1970s, when wind turbine technology was still in its infancy, technology has changed with time and it is expected soon to be directly competitive with conventional energy sources. The life span of the first wind farms is coming to an end, which implies repowering processes aiming an augmented efficiency and reliability of the wind turbines. From the available technologies of the conversion systems, it is far from clear which of them is the optimal.

Clustering wind farms allocate different units into a group which contains some common characteristics, which may be used to reduce the size and the order of mathematical models and also to perform pattern classification into extensive multidimensional data set [4].

The motivation for this work is supported by the absence of a deterministic certainty in allocating outputs of wind farms, regarding the technological conversion system trends [5, 6]. Therefore, this study aims at giving some insights into technological approaches for wind turbines, using probabilistic clustering to identify homogeneous groups.

A previous study was performed using Hierarchical Cluster Analysis and Discriminant Analysis [7]. This ongoing work uses a different approach based on Hierarchical Cluster Analysis and K-Means Clustering. In order to identify the clusters characteristics, it is also performed an exploratory descriptive analysis and also an inferential analysis.

The paper is organized as follows: next section presents technological trends on wind energy conversion systems, Section III overviews the clustering and validation methodologies, Section IV applies the proposed approach to a case study, presents the main results and discussion and, finally, Section V rounds up the paper with the main conclusions.

## II. WIND ENERGY CONVERSION SYSTEM TECHNOLOGIES

Despite the fact first development of commercial wind energy technology began in the late 1930s, only after the oil crises of the 1970s, there had begun economical incentives to develop the technology further [8]. Since the 1980s, there has been a significant consolidation of the design of wind turbines. This section describes main design styles in wind energy conversions systems and points out the technology trends.

### A. Design styles

The mainstream commercial market uses horizontal axis wind turbines, meaning the rotating axis is parallel to the ground. This option is inherently more efficient than vertical axis. Concerning the number of blades, the aerodynamic efficiency and reduced acoustic noise emission establish three-bladed rotor design.

Other important issue related with the design of a wind energy conversion system is the mean of limiting rotor power in high operational wind speeds. There are two main approaches: stall and pitch control.

In stall regulated machines, speed regulation is intrinsic to the aerodynamic design, without any change of the rotor geometry. Under this control approach, wind turbine runs at approximately constant speed even when the wind speed is high, without producing excessive power. The constant speed is achieved through the connection of the electric generator to the grid. Regarding this aspect, the grid behaves like a large flywheel, holding the speed of the turbine nearly constant irrespective of changes in wind speed.

Pitch control involves pitching the blades (*i.e.*, turning the wind blades about their main axis) in order to regulate the power the rotor extracts from wind. This control involves an active control system, which should sense the blade position and defines appropriate changes of blade pitch, according to the measured output power.

Another important and decisive design issue of the wind turbines is the use of variable rotational speed versus fixed speed, with consequences on the overall performance of the system [9, 10].

The constant speed turbine designs consist on generators operating at fixed speed when producing power, directly connected to the utility grid which, through the generator, holds the speed constant. This concept makes use of Squirrel Cage Induction Generators (SCIG) with a geared drive train to adapt the rotational speed to the frequency of the grid. The wind energy capture and also the power quality in the utility grid are reduced.

Variable speed wind energy systems in operation below rated power can enable increased energy capture, and above

rated power, even over quite a small speed range, can substantially ease pitch system duty and reduce output power variability, which in turn improve the power quality when compared with constant speed systems. This concept may be implemented using synchronous or asynchronous generators, allowing wider or narrower wind speed ranges.

Solutions based on asynchronous generators, the so called Doubly Fed Induction Generators (DFIG), with the stator windings directly connected to the grid and a partial scaled electronic converter between the rotor and the grid, allow a low to moderate variation of the rotor speed. Since the power converter is partially scaled, typically one third of the rated power of the system [11], this solution is somewhat cost effective but, on the other hand, there are limitations to control effectively the grid variables, which translates in a deficient quality power system [12]. It should be pointed out that this concept uses a geared drive train to match the low rotational speed promoted by wind velocities to the higher efficient rotational speed of this generator type.

Solutions based on Synchronous Generators (SG) use full scaled electronic converters. The electrical energy is generated at variable frequency (strictly related to the rotational speed of the rotor) and then converted to the frequency of the grid. This concept takes advantage of the wide speed range operation allowed by the full scale converter between the generator and the grid, which also allows boosting the grid stability and performance. Additionally, this type of generators requires lower ratio gearboxes (or even its omission) than DFIG, which translates in higher reliability and lower maintenance costs [12].

### B. Technology Trends

Due to the high wind speed variability and intermittency, the actual demand on power quality issues rises for generators featuring variable speed, which is the dominant trend in the actual market.

Comparing partial speed range systems, promoted by DFIG, and full-range variable speed drives based on SG, the later bring some attractions, specially on operational flexibility and power quality issues, but also have some drawbacks related with the higher power of the electronic converter, with the same rating of the generator [13]. In fact, there was never a clear case for full variable speed range on economic grounds, with small energy gains being offset by extra costs and also additional losses in the power converter.

Another technological trend is related with direct driven generators, *i.e.*, gearless systems. The direct drive systems of Enercon [14] are long established, and gearless systems or with low ratio gearboxes, using Synchronous Permanent Magnet Generator (SPMG) technology have emerged in recent years [15]. In fact, some manufacturers that in past had based their technology on asynchronous generators are now moving to SPMG with full scaled converter [16].

Permanent magnet technology allows a higher power-to-volume-ratio, and fully rated power converter based systems can be applied without design hardware modifications in both 50 Hz or 60 Hz power systems, which increases flexibility for international developers operating in multiple wind markets [8].

Concerning the power control in high operational wind speeds, the design issues of pitch versus stall and degree of rotor speed variation are evidently connected. The stall-regulated design remains viable, but pitch control offers potentially better output power quality, while overall costs of both systems remain similar [8].

### III. CLUSTERING AND VALIDATION METHODOLOGY

Cluster models are typically used to find groups (or clusters) of similar records based on a data set of variables, where the similarity between members of the same group is high and the similarity between members of different groups is low and the results can be used to identify associations that would not otherwise be apparent [17]. Clustering techniques are useful in a wide variety of situations for example [18]: (i) Market Segmentation, to identify distinct groups among a customer base, allowing precise targeting of sales efforts; (ii) Product Bundling, to identify groups of products that tend to appeal to specific customer types; (iii) Formal Classification to classify groups, such as plants or animals into formal taxonomies; (iv) Medical Diagnosis use biological patterns to uncover rules for identifying or diagnosing medical disorders; (v) Web mining, in which clustering is used to discover significant groups of documents on the Web huge collection of semi-structured documents.

Therefore, the empirical component of this study follows all the steps suggested in the literature. In this sense, the first step is to find the optimum number of clusters, given that initially this is unknown. A hierarchical cluster analysis using the method proposed by Ward [19], is the most common for the problem to be analysed, being the one who stands a more consistent solution and also recommended for quantitative variables measured on a ratio scale. In this methodology, an objective function, defined as the sum of squares of deviations of the individual observations compared with the average of the group, is minimized, aiming at creating groups which have maximum internal cohesion and maximum separate external distance [20]. This method uses the variance to evaluate distances between clusters, which results in an efficient approach when compared with other hierarchical methods (for instance, nearest neighbour, furthest neighbour and median clustering). The Ward's distance,  $D_w$ , between clusters  $C_i$  and  $C_j$  is the difference between the total within cluster sum of squares for the two clusters separately, and within cluster sum of squares, which results from merging the two clusters in cluster  $C_{ij}$  [19]:

$$D_w(C_i, C_j) = \sum_{x \in C_i} (x - r_i)^2 + \sum_{x \in C_j} (x - r_j)^2 - \sum_{x \in C_{ij}} (x - r_{ij})^2 \quad (1)$$

where  $r_i$  is the centroid of  $C_i$ ,  $r_j$  is the centroid of  $C_j$  and  $r_{ij}$  is the centroid of  $C_{ij}$ .

To implement a dissimilarity measure between subjects, it is selected the Euclidean Distance Squared. The distance is defined as the square root of the sum of the squared differences between the values of  $i$  and  $j$  for all the selected variables ( $k = 1, 2, \dots, p$ ), [21]:

$$D_{ij} = \sqrt{\sum_{k=1}^p |x_{ik} - x_{jk}|^2} \quad (2)$$

where  $x_{ik}$  is the value of the variable  $k$  for cases  $i$  and  $x_{jk}$  is the value of the variable  $k$  for cases  $j$ .

To assess the adequacy of classification produced with the hierarchical cluster analysis, a nonhierarchical cluster analysis with K-means is also applied. The K-means cluster analysis aims the partition of  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest mean. Given a set of data  $(x_1, x_2, \dots, x_n)$ , where each data is a  $d$ -dimensional real vector, K-means cluster analysis consists in the partition of the  $n$  data into  $k$  sets ( $k \leq n$ ),  $S = \{S_1, S_2, \dots, S_k\}$  in order to minimize the within cluster sum of squares [22]:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^j - c_j\|^2 \quad (3)$$

where  $\|x_i^j - c_j\|^2$  is a chosen distance measure between a data point,  $x_i^j$ , and  $c_j$  is an indicator of the distance of the  $n$  data points from their respective cluster centroids.

The algorithm is composed by the following steps [23]:

- Step 1: Place  $k$  points into the space represented by the objects that are being clustered. These points represent the initial group centroid.
- Step 2: Assign each object to the group which has the closest centroid.
- Step 3: When all objects have been assigned, recalculate the positions of the  $k$  centroids.
- Step 4: Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

The K-means cluster analysis achieves this by subdividing the data into the required number of clusters by grouping records. Thus, the Euclidean distance between the record's dimensions and the clusters centroid are as small as possible [22]. In order to normalize the variables measures in different scales,  $z$  scores standardization method is used.

### IV. WIND FARMS CLUSTERING

Wind farms clustering is addressed here by applying the previous described methodology to a case study in order to identify possible clusters and their main attributes.

### A. Case Study

Data set comprises 35 wind farms of a major promoter acting in the wind energy sector in Portugal, with a market share of 13,6% [1]. The information was collected from institutional and technical Annual Reports available from [24], including a time span of 3 years (from 2010 till 2012).

The chosen support variables used to cluster wind farms are the capacity factor,  $C_F$ , and net production,  $E$ , both per year.

Capacity factor of a wind farm is the ratio of actual productivity to its theoretical maximum, given by:

$$C_F = \frac{E}{P_N 8760} \quad (4)$$

where  $P_N$  is the installed capacity in the wind farm, *i.e.*, the power available from the wind turbines of the farm under given test wind speed conditions.

Higher capacity factors indicate a better utilization of the installed capacity, which helps to reduce investment costs. In fact, capacity factors are particularly important in evaluating the overall economics of wind farms. Typically, capacity factors need to be elevated to values about 50% (or better) to make a modern wind farm commercially viable [12].

The installed capacity is given by the sum of the rated power of all turbines into the wind farm. If the rated power of a wind turbine is small it could lead to an higher capacity factor, but the turbine may not be able to produce energy at higher wind speeds, which translates in less profit. On the other hand, if the rated power of units is high, the turbine may stall at low wind speeds and the extra energy at high wind speeds may not compensate the higher costs.

Regarding the other support variable, the annual net production,  $E$ , is the energy actually produced by the wind farm in a year, considering programmed and random unavailability (failures) of wind turbines. At present, this output is not constrained by load demand or wholesale markets, as currently regulated.

### B. Results and Discussion

A common way to visualize the cluster analysis progress is through the draw of a dendrogram, displaying the distance level at which there is a combination of wind farms and clusters (Fig. 1). Therefore, through the graphical visualization of Fig. 1 it is possible to anticipate the hypothetical optimal number of wind farms' clusters as well as their composition. At the rescaled distance of 10, it is straightforward the definition of three notable groups.

Nevertheless, in order to identify the optimal number of clusters, it is used the test R-Squared (R-Sq). The results of the relativized distance between clusters is shown in Fig. 2. A solution of tree clusters was chosen, explaining 45% of the total variance.

The following step consists in application of the K-means method to validate the Hierarchical Cluster Analysis. As previously pointed out, it is necessary to specify the number of clusters when conducting K-means clustering, that in this study is 3.

Table 1 outlines groups of wind farms with similar features, allocated in the 3 identified clusters. The Cluster A comprises 21 wind farms, Cluster B includes 9 wind farms and, finally, Cluster C has 5 wind farms.

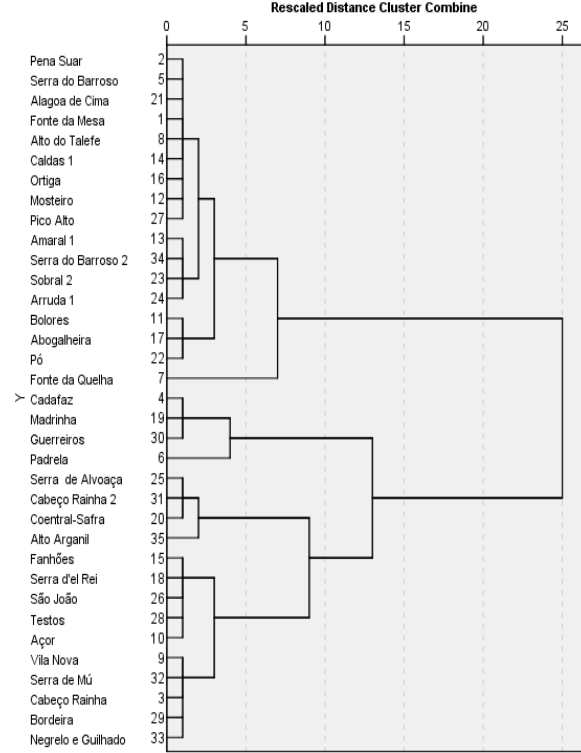


Figure 1: Dendrogram using the Ward linkage method.

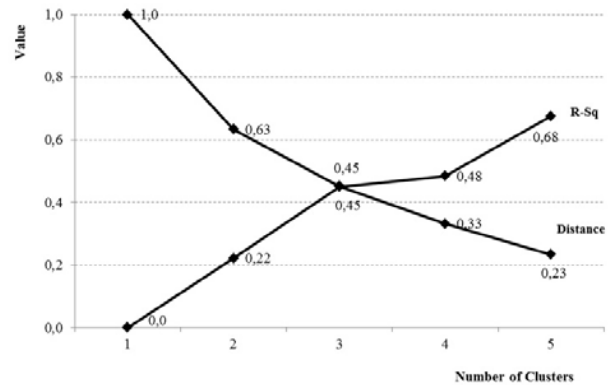


Figure 2: Optimal number of clusters.

TABLE I. CLUSTERING COMPONENTS BASED ON VARIABLES CAPACITY FACTOR AND NET PRODUCTION PER YEAR.

	Cluster A	Cluster B	Cluster C
Wind Farms	1, 2, 5, 7, 8, 10 11 - 18, 21 - 23 26 - 28, 34	3, 9, 20, 25, 29, 31 - 33, 35	4, 6, 19, 24, 30

Attributes of Cluster A are defined by a medium installed capacity (mean value equal 13,6 MW with a standard deviation of 5,9) and a predominance of technology based on Induction Generators, including Squirrel Cage (SCIG) and Doubly Fed (DFIG) generators (76,2%). As previously mentioned, these wind turbine generators require multi-stage gearboxes, which translates in high maintenance costs and reduced reliability. It should be noted that technology based on SCIG presents a solution with a poor aerodynamic efficiency, which is augmented to moderate values when the DFIG is used.

Cluster B is characterized by an higher installed capacity (mean value 27,7 MW with a standard deviation of 9,3) and the dominant technology (77,7%) is based on direct driven Synchronous Generators. This technological concept using full variable speed range improves substantially the efficiency of the system. Moreover, the absence of the gearbox component increases the reliability and allows reduced maintenance schemes which, together with the wide speed range operation, results in an increased capture of the disposable wind energy, in the order of 10% [7].

Finally, cluster C includes both variable speed technologies (DFIG and SG), presents the higher capacity factors of the data set for the lower mean value of the installed capacity (10,3 MW with a standard deviation of 4,6). The improved capacity factors of these wind farms may be related with high wind speeds in the geographical areas they are located. This hypothesis should be verified, introducing a wind availability variable into the model.

Subsequently, it has been performed the descriptive statistics including the mean, minimum and maximum values, as well as the standard deviations of the clustering variables. Table II shows the results for the variables into the different clusters, Capacity Factor ( $C_F$ , in %), and Net Production ( $E$ , in GWh), for the years 2010, 2011 and 2012.

TABLE II. SUMMARY OF DESCRIPTIVE STATISTICS BY CLUSTER AND VARIABLES.

Cluster	Variable	$n$	Minimum	Maximum	Mean	Std. Deviation
A	$C_F$	21	12,60	28,31	24,4246	3,27511
	$E$		6,50	51,83	28,8794	13,13830
B	$C_F$	9	27,81	30,59	28,7806	0,84963
	$E$		51,27	105,80	74,4333	18,46653
C	$C_F$	5	29,74	42,77	33,9700	5,06650
	$E$		15,63	51,47	30,4867	12,97580

From these results, it is possible to verify that cluster A contains lower values for both variables. On the other hand, cluster C comprises higher values of the capacity factor

when compared with the other clusters, being the net production lower than the one of cluster B, due to its low mean installed capacity.

In contrast to hierarchical clustering, the K-means outputs provide the ANOVA (ANalysis Of VAriance) of the cluster centers, as presented in Table III, indicating which variable contributes to the cluster solution. Variables with large mean square errors (and lowest  $F$  test statistics) provide the least help in differentiating between clusters. From these results, it is possible to verify that all the clustering variables' means differ significantly across at least two of the three clusters, because the null hypothesis is rejected in every cases (p-values less than 5% of significance level).

TABLE III. ANOVA BETWEEN GROUPS.

Year	Variable	Cluster		Error		$F$	p-value
		Mean Square	$df$ ( $k-1$ )	Mean Square	$df$ ( $n-k$ )		
2010	$C_F$	7,176	2	0,614	32	11,686	<0,001
	$E$	10,876	2	0,383	32	28,415	<0,001
2011	$C_F$	10,423	2	0,411	32	25,355	<0,001
	$E$	11,099	2	0,369	32	30,097	<0,001
2012	$C_F$	8,949	2	0,503	32	17,785	<0,001
	$E$	11,489	2	0,344	32	33,357	<0,001

The previous view shows an overview of the variables' overall importance for the clustering solution, which provides the same result as the Hierarchical Cluster Analysis.

## V. CONCLUSIONS

The main objective of this research was to identify homogeneous groups within a data set of wind farms of a major promoter acting in the energy sector in Portugal, based on two multivariate analyses: Hierarchical Cluster Analysis and Non Hierarchical Analysis (K-means clustering). Based on both methodologies, from the obtained results, it has been possible to identify three clusters whose attributes are the installed capacity and the technological trend based on the wind turbine generator type.

From the attributes of the identified clusters, it is possible to infer that technological concepts using squirrel cage and doubly fed induction generators are grouped together, presenting the lowest mean values of capacity factor and productivity for a medium installed capacity. On the other hand, the technological trend within wind farms with higher installed capacity is the gearless variable speed synchronous generator wind turbine. This technology is able to improve the mean capacity factor when compared with the previous one.

Finally, an identified third cluster with the lowest mean value of the installed capacity, contains wind farms with the higher capacity factors and includes both variable speed technologies: partial speed range systems, using doubly fed induction generators, and full-range variable speed drives

based on synchronous generators. The improved capacity factors of these wind farms, disregarding the variable speed technological trend, may be linked with an higher wind speed into the geographical areas they are located. Therefore, future work should consider the inclusion of the variable wind speed, as it translates the physical input “fuel” in the production function. It should also be considered an increased data set of wind farms and an extended time span in order to establish a comparative analysis with previous results.

## REFERENCES

- [1] ENEOP. (2014, Access Date: January, 2014). *Eólicas de Portugal, S.A.* Available: <http://www.eneop.pt/>
- [2] M. Denault, D. Dupuis, and S. Couture-Cardinal, "Complementarity of Hydro and Wind Power: Improving the Risk Profile of Energy Inflows," *Energy Policy*, vol. 37, pp. 5376-5384, 2009.
- [3] R. Karki and R. Billinton, "Cost-Effective Wind Energy Utilization for Reliable Power Supply," *IEEE Transactions on Energy Conversion*, vol. 19, pp. 435-440, 2004.
- [4] M. Ali, I. S. Ilie, J. V. Milanovic, and G. Chicco, "Wind Farm Model Aggregation Using Probabilistic Clustering," *IEEE Transactions on Power Systems*, vol. 28, pp. 309-316, 2013.
- [5] H. Arabian-Hoseynabadi, P. J. Tavner, and H. Oraee, "Reliability Comparison of Direct Drive and Geared Drive Wind Turbine Concepts," *Wind Energy, Wiley Online Library*, vol. 13, pp. 62-73, 2010.
- [6] H. Polinder, F. F. A. van der Pijl, G. J. de Vilder, and P. J. Tavner, "Comparison of Direct-Drive and Geared Generator Concepts for Wind Turbines," *IEEE Transactions on Energy Conversion*, vol. 21, pp. 725-733, September, 2006 2006.
- [7] P. O. Fernandes and A. P. Ferreira, "Wind Farms Model Aggregation Using Probabilistic Clustering," in *11th International Conference of Numerical Analysis and Applied Mathematics 2013, ICNAAM 2013*, Rhodes, Greece, AIP Conference Proceedings, 2013, pp. 618-621.
- [8] EWEA, *Wind Energy – The Facts*: Routledge, 2009.
- [9] A. D. Hansen, F. Iov, F. Blaabjerg, and L. H. Hansen, "Review of Contemporary Wind Turbine Concepts and their Market Penetration," *Wind Engineering*, vol. 28, pp. 247-263, 2004.
- [10] P. Mutschler and R. Hoffmann, "Comparison of Wind Turbines Regarding their Energy Generation," in *IEEE 33rd Annual Power Electronics Specialists Conference, PESC*, 2002, pp. 6-11.
- [11] F. Blaabjerg, F. Iov, T. Kerekes, and R. Teodorescu, "Trends in Power Electronics and Control of Renewable Energy Systems," in *14th International Power Electronics and Motion Control Conference (EPE-PEMC 2010)*, Ohrid, Republic of Macedonia, 2010, pp. k1-k19.
- [12] S. Mathew and G. S. Philip, *Advances in Wind Energy and Conversion Technology*: Springer, 2011.
- [13] J. A. Baroudi, V. Dinavahi, and A. M. Knight, "A Review of Power Converter Topologies for Wind Generators," *Renewable Energy, Elsevier*, vol. 32, pp. 2369-2385, 2007.
- [14] Enercon. (2011, Access Date: June, 2011). *Enercon - Technology*. Available: <http://www.enercon.de/en-en/21.htm>
- [15] J. F. Conroy and R. Watson, "Low-Voltage Ride-Through of a Full Converter Wind Turbine with Permanent Magnet Generator," *Renewable Power Generation, IET*, vol. 1, pp. 182-189, 2007.
- [16] GE\_Energy. (2011, Access Date: June, 2011). *Wind Turbines*. Available: [http://www.ge-energy.com/products\\_and\\_services/products/wind\\_turbines/](http://www.ge-energy.com/products_and_services/products/wind_turbines/)
- [17] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*: Wiley-Interscience, 2005.
- [18] W. Hardle and L. Simar, *Applied Multivariate Statistical Analysis*, 3th ed. Berlin Heidelberg: Springer, 2012.
- [19] J. Ward, "Hierarchical Grouping to Optimize an Objective Function," *Journal of the American Statistical Association*, vol. 58, pp. 236-244, 1963.
- [20] W. H. Greene, *Econometric Analysis*, 7th ed. New Jersey: Prentice Hall, 2011.
- [21] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed.: Pearson, 2007.
- [22] K. Singh, D. Malik, and N. Sharma, "Evolving Limitations in K-means Algorithm in Data Mining and their Removal," *International Journal of Computational Engineering & Management*, vol. 12, pp. 105-109, 2011.
- [23] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, Calif., 1967, pp. 281--297.
- [24] EDP. (2014, Access Date: January, 2014). *Energias de Portugal*. Available: <http://www.edp.pt/>