

# Improving the operation and maintenance of wind farms: determination of wind turbine performance

A. Llombart-Estopiñan

Fundacion CIRCE; Department of Electrical Engineering, Zaragoza University

C/ Maria de Luna 3, 50018 Zaragoza (Spain)

Phone/Fax number:+34 976 762398, e-mail: [llombart@unizar.es](mailto:llombart@unizar.es)

## Abstract:

To improve the economic and technical wind farm operation is necessary an accurate determination of wind turbine performance. To do so it is necessary to estimate the unavailability and the efficiency of each wind turbine.

The first step consists in filtering the data set of the period to be studied. Two data sets are obtained, a first set where all data corresponds to time periods in which the wind turbines have worked properly, and the rest of data correspond to periods in which the WT have had some problems, and have to stop during part of the time.

The second set is useful to get the unavailability measured both in time and power, and the first one to get the efficiency.

In both cases a model to get the relationship between power and wind speed is needed. The CIRCE's AIRE (Renewable Energy Integral Analysis) group is working in automatic methods for filtering wrong data and characterizing the wind turbine behaviour with the objective of improving the operation and maintenance of wind farms.

## Key words

Wind energy, power curve, power performance, data filtering, robust filtering

## 1. Introduction

The cost effective operation and maintenance of wind farms requires accurate and informed condition monitoring of the wind turbines. For the latest wind turbines, e.g. >700kW, there is limited data available to assess the long-term power performance of these machines. In addition, in Spain there are concerns in relation to the long-term operation in sites with highly complex terrain and high levels of turbulence intensity.

To assess the performance of a wind turbine two parameters have to be studied: unavailability and efficiency.

The unavailability can be measured both in time and in power. The first one is quite easy to get having the SCADA records, but in the second case it is necessary a production model in order to estimate the power losses.

To get the efficiency of a wind turbine the relationship between the wind mean speed taken either at the meteorological mast or at the nacelle and the power generated by each wind turbine must be found. So in both cases it is necessary to get a production model that relates power and wind speed.

But the best method to do this is not clear. There are some factors that complicate this relationship:

1. The time delay in the wind speed propagation and how this affects the correlation between mast and turbine wind speed;
2. The topography of the wind farm site;
3. The presence of nearby obstacles.

Several studies have been carried out on this topic. In [1-3] some methods to estimate wind turbine power are proposed. In [5-6] different ways to characterize the power curve are discussed and tested. All this methods need good data in order to get the parameter of the different models proposed. In fact, this is one of the most critical steps.

As was stated in [9], experience have shown that real measurements are prone to wrong data, being essential to have some sort of data check to protect the models against the influence of erroneous measurements. They use a range check, a stationarity check and a confidence check with predefined confidence bands to detect wrong data.

Taking into account the abovementioned considerations, the process to analyze the performance of a WT is the following:

1. Data filtering: to set of data are got, the wrong data are used to analyze the unavailability and the right ones to get the efficiency (or the model if historical data are considered)
2. Analyzing the unavailability: getting the percentage of time that the WT has been working right and getting the power losses (that implies the need of a P-v model)
3. Analyzing the efficiency: by comparing the production of the period to be analyzed with the theoretical production got from the P-v model

To do these tasks automatically some needs arise: getting an automatic filtering technique and getting a P-v model. These needs are the objectives of this paper.

## 2. The need of filtering

Focussing the problem in wind power characterization, the most important data to consider are:

- Wind speed
- Power production

Wind speed can be affected by shade effect due to the presence of some obstacles. This effect makes the wind speed to decrease. In figure 1 it can be observed this phenomenon clearly (data in zone 1). Considering the power production, if the value of a power data is the mean value of the power produced during ten minutes by a wind turbine, it must be assured that the wind turbine has worked properly during those ten minutes. In other case, the data doesn't represent the normal working of the turbine (figure 1, data in zone 2). Sometimes, the effect of the different problems is not so clear. In Figure 1, data in zone 3 and zone 4 are probably bad, but this cannot be assured without considering more information.

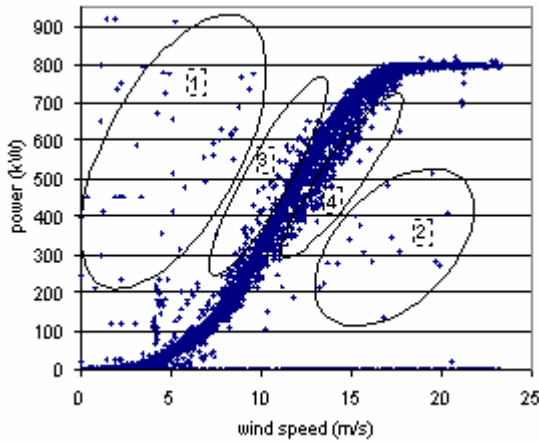


Figure 1: Original set of data considered

It can be thought that the problem of data power quality can be solved accurately. There are alarm records in the SCADA system of the wind farms. But usually this records don't exist for the whole period to be treated and other times it cannot be assured the accuracy of the alarms recorded.

In figure 2 it can be observed that some data en zones 1 and 2 have been not well filtered. So, despite of considering the alarms register some wrong data still remain.

There are a huge amount of data filtering techniques, but their application in wind systems it is not well referenced. Moreover, data coming from wind systems are not common, i.e. the changes in the mean values of the wind speed are not easily characterized and this is reflected in the power data so the filtering of that data is not obvious, therefore, most people in the wind sector filter the data manually.

CIRCE's AIRE group is working in a method to filter wind data using the Least Median of Squares (LMedS)

method. It allows detecting and rejecting bad measures simultaneously to the wind power curve estimation, working in a robust way.

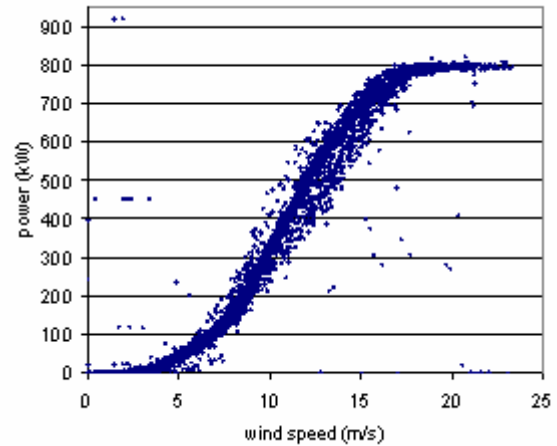


Figure 2: Data set after filtering those affected by the alarms recorded at the SCADA system

## 3 The robust statistic method

We have redundant data within unknown wrong information that must be filtered out to compute the best solution. This circumstance is similar to some situations in computer vision applications where, robust techniques have being used with very good results, and are actually mandatory in practice.

### 3.1 Robust estimation

Usually we have many measures and an estimation method can be used to process all of them, exploiting redundancy to get a better result. The classical method is to use the Least Mean of Squares (LMS) estimator. The LMS method assumes that all the measurements can be interpreted with the same model, which makes it to be very sensitive to out of norm data, or outliers. It has a breakdown of 0% of spurious data, which means that a sole outlier can destroy the fitting [4]. LMS minimizes the sum of the squares over all the measurements and, if a measurement is far away the correct value, its squares error prevail in the summation, and therefore prevails in the fitting. In figure 3 we can see a simple example where a line is fitted from noisy points. Using LMS a sole outlier point can destroy the fitting.

Several people have tried to make this estimator robust by replacing the square by something else like the absolute value, but not touching the summation sign. However, the key issue is to try that outliers have not any influence on the result.

Robust estimators provide well founded methods to detect the outliers, obtaining trustworthy results even if a certain amount of data is contaminated [4]. From the existing robust estimation methods [7,8], we have chosen the LMedS method. Comparing with the LMS method, the LMedS method replaces the sum by the median

which is very robust, but unfortunately it has not analytical solution, and an iterative solution is required.

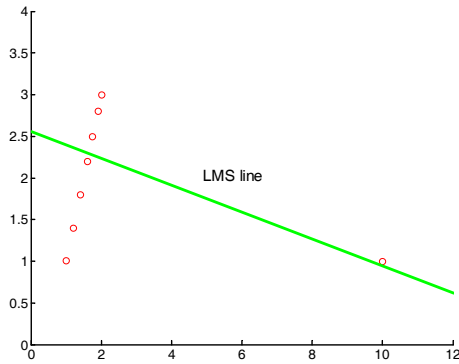


Figure 3: Fitting of a line from noisy points by LMS method.

The LMedS method makes a search in the space of solutions obtained from subsets of minimum number of data. Let us suppose that a polynomial regression with 4 coefficients is used to adjust the power curve. If we need a minimum of 4 measures to adjust the model for the power curve, and there are a total of  $n$  measurements, then the space of solutions will be obtained from the combinations of  $n$  elements taken 4 by 4, giving  $m$  possible solutions:

$$m = \frac{n!}{4!(n-4)!} \quad (2)$$

The algorithm to obtain an estimate with this method can be summarized as follows:

- 1) Calculate the  $m$  subsets of the minimum number of measurements required to adjust your curve.
- 2) For each subset  $S$ , we compute a power curve in closed form  $\mathbf{P}_S$ .
- 3) For each solution  $\mathbf{P}_S$ , the median  $M_S$  of the squares of the residue with respect to all the measurements is computed.
- 4) We store the solution  $\mathbf{P}_S$  which gives the least median  $M_S$ .

When having many measurements, this exhaustive search is computationally too expensive, and a solution in practice is to select randomly enough subsets, of the minimum number of measurements required to adjust the power curve, to warrant a reasonable probability of not fail. Thus, the first step is substituted by a Monte-Carlo technique to randomly select  $m$  subsets of 4 measurements.

A selection of  $m$  subsets is good if at least in one subset all the measurements are good. If  $P_{ns}$  is the probability that a measurement is not spurious, and  $P_m$  is the assumed probability of missing the computation, not reaching a good solution, the number of subsets to consider can be computed as:

$$m = \frac{\log P_m}{\log(1 - P_{ns}^4)} \quad (3)$$

For example, if we accept a small probability of fail  $P_m = 0.001$ , with an estimation of the probability of good measurements  $P_{ns} = 75\%$ , the number of subsets  $m$  should be 19.

In the example of figure 3, the LMedS method obtains a good line if the outlier point is not randomly selected in at least one of the subsets of two points.

### 3.2 Rejection of wrong data

The search in the space of solutions using the median gives a robust solution where the spurious data or outliers have not any influence. Besides that, we can detect easily the outliers as those of higher residue assuming the noise for the inliers is Gaussian. The standard deviation of the error can be estimated from the least median of the residues  $M_S$  as [4]:

$$\hat{\sigma} = 1.48[1 + 5/(n-4)]\sqrt{M_S} \quad (4)$$

which allows to define a threshold to select the inliers and the outliers from them. Taking, for example, a probability of 99% of accepting a measure being good, the threshold will be fixed to  $2.57\hat{\sigma}$ . A more detailed explanation of this method can be seen in [4].

### 3.3 Models to be considered

As it is explained in section 3.1, the robust statistic method needs a model in order to work. So, we have considered two different models to estimate the power curve as a function of the wind speed.

The first one considers a polynomial regression and a wind speed partition. The power curve of a wind turbine has three clearly defined parts. Firstly, between 0 m/s and the cut-in wind speed the production is zero, then the power output increases following in theory a cubic relationship with wind speed until the rated wind speed, and finally, depending on the wind turbine model, the power output is regulated to stay almost constant until the cut-out or furling wind speed. This three-part shape suggests that different polynomial relationships should be used for different wind speed ranges.

We consider a third order polynomial to adjust the power ( $P$ ) in function of the wind speed ( $V$ ) in each part.

$$P = a_0 + a_1V + \dots + a_3V^3 \quad (1)$$

For an ideal power curve, the partitioning of the curve for the purposes of fitting a polynomial is fairly clear. However, if a turbine is a significant distance from the meteorological tower, the effective power curve observed may not be so clearly partitioned and may differ from turbine to turbine depending on the distance.

The second model is based on the bin method, which is outlined in the IEC 61400-12 standard. In this case the data are grouped using bins of 0,5 m/s of size. We consider a continuous polygonal curve considering in

each bin a linear relation between the power and the wind speed. At the moment partitioning by wind direction has not been considered because there are not data of wind direction from the anemometer at the turbine, but only from the meteorological tower, where the wind speed does not predict so well the power of the wind turbine [5].

## 4. Modelling the P-v relationship

Several studies have been carried out on this topic using Neural Network Techniques. These studies compare the use of NNs with relatively simple stochastic methods and conclude that the NN approach is superior, however, the stochastic methods chosen are overly simplistic.

In order to have a clearer model different stochastic methods that can be used to solve the problem are reviewed (bin method and polynomial regression). And, some modifications to all of them are proposed to get more accuracy.

### 4.1 Bin Method

The analysis of the Bin method has been divided in three different tests:

#### A. Bin method characterization

The bin method has been applied to each wind turbine as it is explained in paragraph 2, but, opposite to that, all the data have been taken as valid independently of their wind direction.

The power curve that characterizes the wind turbine behaviour is composed by the link of the different points obtained. So, the theoretical power that corresponds to each wind speed, can be obtained through the interpolation between the two nearest bins.

To obtain the mean squared error it has been calculated the error between the theoretical and the real power for all the data in the full year.

#### B. Bin method characterization with normalized power

It is clear that the relationship between wind speed and power output, is not linear. If the bin characterised by wind speeds between 5 and 5,5 is observed, and a cubic relationship between the different variables is assumed, the theoretical power output associated with the wind speed at the maximum value of the bin is 33,1 % higher than that associated with the wind speed at the minimum value of the bin.

To take into account the “non linear” relationship between the wind speed and the power, a new method of obtaining the  $V_i$  and  $P_i$  values is proposed:

$$V_i = \frac{V_{i,H} - V_{i,L}}{2} \quad (1)$$

$$P_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \frac{P_{n,i,j} \cdot V_{n,i,j}^k}{V_i^k} \quad (2)$$

where

$V_{i,H}$  is the maximum value of wind speed in bin  $i$   
 $V_{i,L}$  is the minimum value of wind speed in bin  $i$   
 $k$  is a factor that must be adjusted for each direction sector and each wind speed bin

To get the theoretical power that corresponds to any wind speed ( $V$ ) these steps must be followed:

- 1) Select the bin that corresponds to the wind speed chosen.
- 2) Apply the following equation, in order to consider the relationship shown in eq. 2

$$P_v = \frac{P_i \cdot V^k}{V_i^k} \quad (3)$$

where

$P_v$  is the theoretical power corresponding to  $V$   
 $V_i$  is the normalised and averaged wind speed in bin  $i$   
 $P_i$  is the normalised and averaged power output in bin  $i$

To get the characterization some  $k$  values have been tested from 0,5 to 3, in steps of 0,5. Each bin have been characterized for the  $k$  value that optimises the mean squared error, obviously the data taken to calculate the error are only those that belongs to this bin.

Finally, the mean squared error for the whole year has been calculated.

#### C. Bin method characterization with sector partition

We have carried out studies to analyse the effect of wind direction in the characterisation process. The data in each bin are grouped as a function of the direction and then  $V_i$  and  $P_i$  are calculated, for each sector.

The approach that we have been considered is a partition in 4 sectors. Taking as reference the direction in which the turbines are aligned, we have considered 4 sectors as it is shown in figure 1 (the sectors are defined as indicated in the bottom right corner).

This method has been used in combination with the method shown in paragraph B.

Each sector has been analysed separately and then the mean squared error for the whole year has been calculated.

The different characterization methods exposed can be applied using either the wind speed at the meteorological mast or that at the nacelle of each wind turbine.

Obviously if the wind direction must be considered the only data available are those at the meteorological tower, that have been used independently of the wind speed considered.

## 4.2 Polynomial method

We have considered a number of approaches to improve the polynomial method, those approaches are described below. Validation of the different methods described is carried out by comparing the predicting wind turbine power outputs with the actual power outputs and by calculating the mean squared error.

### A. All data

Initially, the data were analysed as a group and the relationship between mast and turbine determined. This was used to test the influence of distance and as a base case against which to compare the different methods described below.

The approach proposed limits the polynomial calculated to the fifth order and, due to the limitation of the tool utilized (Excell by Microsoft), all the different terms of the polynomial have been taken into consideration, that is, the different expressions obtained are as that showed below:

$$P = a_0 + a_1V + \dots + a_5V^5 \quad (1)$$

### B. Partitioning the power curve by wind speed

For an ideal power curve, the portioning of the curve for the purposes of fitting a polynomial is fairly clear. However, if a turbine is a significant distance from the mast, the correlation between the wind speeds at mast and turbine reduce and the effective power curve observed may not be so clearly partitioned and may differ from turbine to turbine depending on the distance.

As a first step, the partitions for each turbine used in the study have been chosen identically for all the cases. This has allowed comparing the results easily.

In the next figure it can be seen the limits proposed, these have been chosen using practical methods (some try and error probes around the theoretical ranges).

### C. Fitting different polynomials by wind direction

If the wind farm site is complex, the correlation between mast and turbine wind speed will be different for different direction sectors. This suggests that the power curve should be fitted using a different polynomial curve for different direction sectors.

### D. Partitioning by both wind speed and direction

Clearly, partitioning of the power curve can be made for both wind speed and these partitions can be changed for different direction sectors before polynomials are fitted.

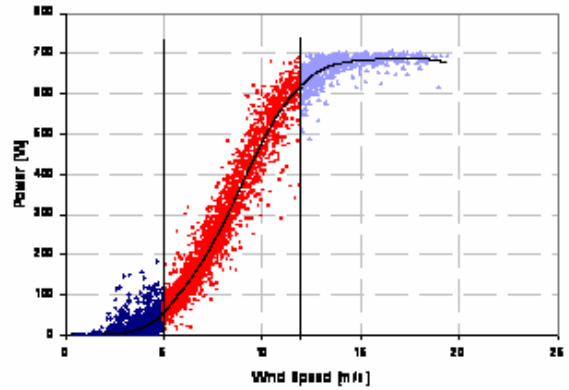


Figure 4.-P-V curve for wind turbine 1 and wind speed partition

## 5 Experimental results

### 5.1 Methodology

To test the goodness of the filtering method proposed, a year of historical data from a wind farm has been used.

The wind farm chosen to perform the different tests is situated in Aragon (Spain). Due to confidentiality issues it is not allowed to present it in detail but, in order to understand the type of the terrain and to present the situation of the wind turbines whose data are used in the study, a simplified map is presented (Fig. 5).

In this map, it can be seen that the terrain is extremely complex, the turbines are disposed in two lines (following, more or less, the straight lines represented in the map) and the main wind direction is indicated by the big arrow in the top left corner.

The analysis of the filtering technique has been divided in different tests:

#### A.- Filtering technique

##### 1 Filtering using the alarm records at the SCADA system

This filtering method consists in marking as wrong data, all those data that are affected by the alarm record of the SCADA system.

##### 2 Filtering using the alarm records at the SCADA system plus a classical statistical method

Once the alarm records have been considered, the data are divided in bins. The mean value ( $\mu$ ) and the standard deviation ( $\sigma$ ) are calculated for each bin. Then, all the data that are bigger than  $\mu + 3\sigma$  or smaller than  $\mu - 3\sigma$  are considered as wrong data.

##### 3 Filtering using the alarm records at the SCADA system plus a robust statistical method

Once the alarm records have been considered, the robust method, explained in section 4, is applied.

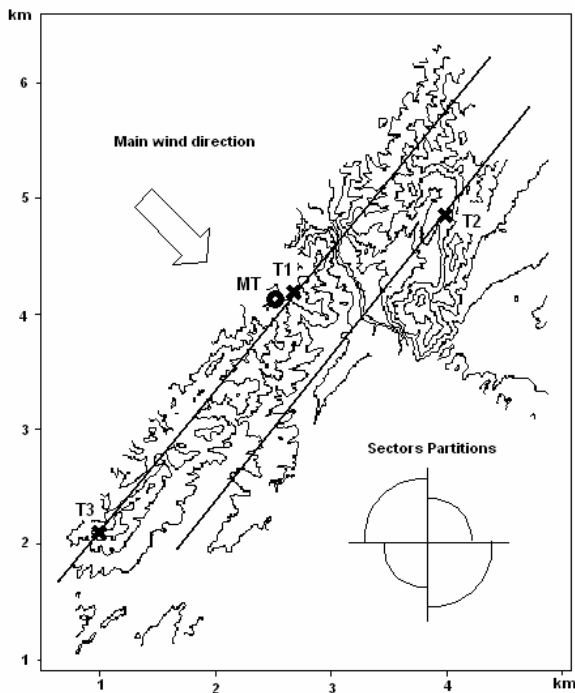


Figure 5. A simplified view of the wind farm

#### 4 Filtering using only a classical statistical method

Taking into account the original data set, the mean value ( $\mu$ ) and the standard deviation ( $\sigma$ ) are calculated for each bin. Then, all the data that are bigger than  $\mu + 3\sigma$  or smaller than  $\mu - 3\sigma$  are considered as wrong data.

#### 5 Filtering using only a robust statistical method

Taking into account the original data set, the robust method, explained in section 4, is applied.

In both tests with robust filtering, the threshold taken in practice to reject outliers is  $3\sigma$ .

## 5.2 Results

The different methods explained in section 5.1 have been applied, and the results obtained are shown in different figures.

The results obtained using the first method are shown in figure 2. As it has been explained in section 1 this method it is not suitable to comply the task required.

The second method proposed gets much better results (see figure 6), but there are some data that are clearly wrong, and they have not been detected. These wrong data belongs mainly to zones 1 and 2 of figure 1.

The third method which includes the robust method fits quite well with the objective proposed (figure 7).

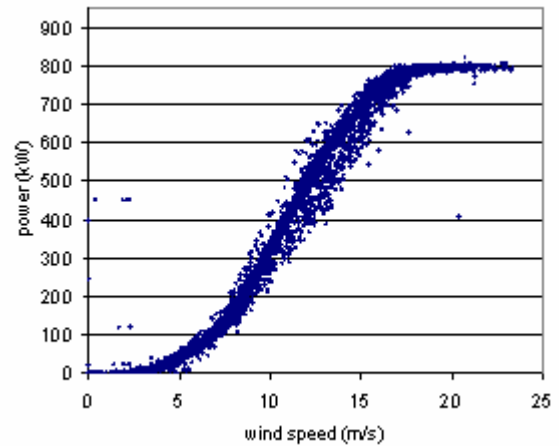


Figure 6: Data set after applying the filtering method 2

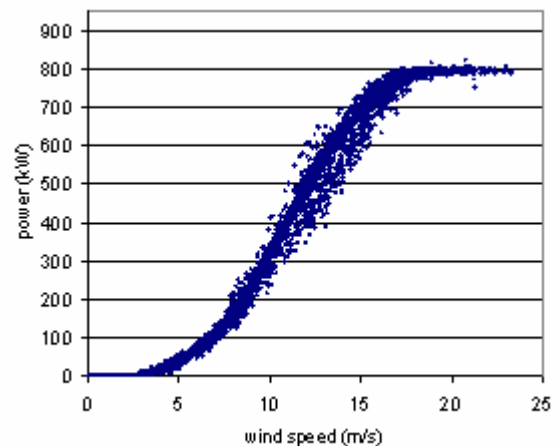


Figure 7: Data set after applying the filtering method 3

Once it has been demonstrated the suitability of the robust statistical method proposed, we want to show the special robustness of it. To do so, the classical statistical filter and the robust filter have been applied over the original data set. The results are shown in figures 8 and 9 respectively.

It is clear that the method proposed has a good behaviour even in extremely bad conditions, in which the classical method works not quite well.

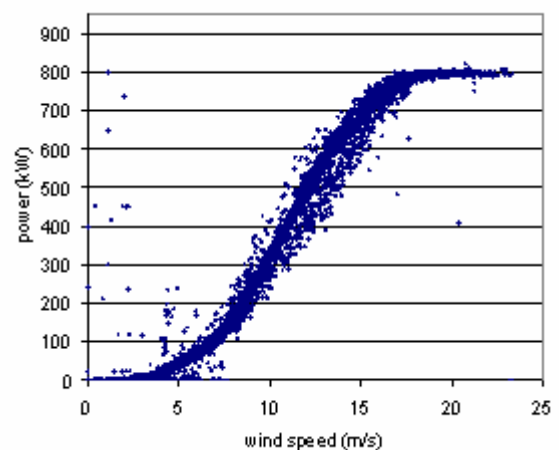


Figure 8: Data set after applying the filtering method 4

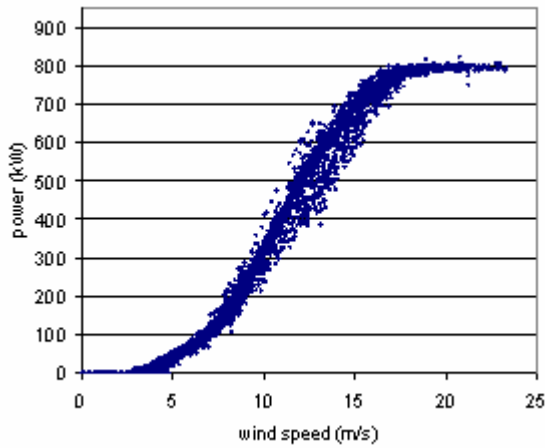


Figure 9: Data set after applying the filtering method 5

The different characterization methods explained in section 4 have been applied to the two turbines using both the wind speed taken at the meteorological mast and the wind speed taken in the nacelle of each turbine. These results are shown:

	Turbine 1			
	Using wind speed		Using nacelle speed	
	MSE	Improv. (%)	MSE	Improv. (%)
Method IEC	1359.08		81.74	
Method A1	1357.48	0.12%	-	-
Method A2	1217.59	10.41%	77.23	5.52%
Method B1	1383.26	-1.78%	118.19	-44.59%
Method B2	1358.41	0.05%	81.26	0.59%
Method B3	1247.05	8.24%	108.51	-32.75%
Method B4	1215.91	10.53%	78.11	4.44%

TABLE I. – Characterization results for turbine 1 (close to the met mast)

	Turbine 2			
	Using wind speed		Using nacelle speed	
	MSE	Improv. (%)	MSE	Improv. (%)
Method IEC	2730.49		119.1	
Method A1	2717.29	0.48%	-	-
Method A2	1837.47	32.71%	93.89	21.17%
Method B1	2773.11	-1.56%	164.04	-37.73%
Method B2	2732.31	-0.07%	118.53	0.48%
Method B3	1882.83	31.04%	127.23	-6.83%
Method B4	1851.17	32.20%	94.25	20.86%

TABLE II. – Characterization results for turbine 1 (further from the met mast)

There is a table for each wind turbine; in these tables the results of each method and speed are shown. The mean squared error is presented and for the different modifications also is presented the improvement measure as a percentage of the first method.

Looking at the results got in the tables it can be said that the two methods are similar for the task of representing the performance of the power in wind turbines.

It can be said that, for this wind farm, the wind speed taken at the nacelle is more suitable to characterize the power performance of the wind turbine.

In relation with the different modifications proposed, the consideration of different sector always improve the

results obtained, modification in the polynomials regression, based on the wind speed partition is important in the case that the wind speed at the nacelle is considered.

## 5. Conclusion

Characterisation of the wind turbine power curves in relation to the wind speed recorded in real scenarios is a problem that must be solved in order to optimise the operation and maintenance of the wind farm.

To fit the models some data must be used and their quality must be assured. The Least Median Squared (LMedS) method has been proposed in order to filter the crude data taken from a real wind farm. The results obtained are good and the method has shown a very good robustness to reject bad data.

Modifications to the basic bin method and polynomial method have been proposed which take account of distance and complexity of terrain. The value of these methods has been determined using actual wind farm data; and the results show that reduce between 13% and 33% the error. Generalizing these results for any windfarm is difficult, however results obtained show that considering wind speed at wind turbine nacelle, in the same way that considering different sectors, is more suitable for characterization purposes.

## References

- [1] S. Li, D. C. Wunsch, E. A. O’Hair and M. G. Giesselmann, “Wind turbine power estimation by neural networks with kalman filter training on a SIMD parallel machine”, International Joint Conference on Neural Networks, Vol. 5, pp 3430 – 3434, July 1999
- [2] S. Li, D. C. Wunsch, E. A. O’Hair and M. G. Giesselmann, “Using Neural Networks to Estimate Wind Turbine Power Generation”, IEEE trans. On Energy Conversion, Vol. 16, N° 3, pp 276-282, September 2001
- [3] S. Kélouani, K. Agbossou, “Nonlinear Model Identification of Wind Turbine with a Neural Network”, IEEE trans. On Energy Conversion, Vol. 19, N° 3, pp 607-612, September 2004
- [4] P. Rousseeuw and A. Leroy, Robust Regression and Outlier Detection (John Wiley, New York, 1987).
- [5] A. Llombart, S. J. Watson, D. Llombart and J.M. Fandos; “Power Curve Characterization I: improving the bin method” ICREPQ, 2005.
- [6] A. Llombart, S. J. Watson, J.M. Fandos and D. Llombart; “Power Curve Characterization II: modelling using polynomial regression” ICREPQ, 2005.
- [7] Z. Zhang, “Parameter Estimation Techniques: A tutorial with Application to Conic Fitting,” *Rapport de recherche RR-2676*, I.N.R.I.A., Sophia-Antipolis, France (1995).
- [8] M. A. Fischler and R. C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,” *Comm. of the ACM* vol. 24, pp. 381-395, 1981.
- [9] T. S. Nielsen, H. Madsen and J. Tofting, “Experiences with statistical methods for wind power prediction ,” in *Proceedings of the European Wind Energy Conference*. 1999 Nice, France, pp1066-1069.