Characterization of wind farm energy production in a zone by artificial neuronal networks

Joaquín Mur   Ángel Antonio Bayod

Abstract—**This paper studies the effect of a number of close wind farms, where there is a relationship among the farms. First of all, data are statistically analysed to check main relationships between farms.**

The second analysis goes further in the characterisation of wind farm relationships. Artificial Neural Networks (ANN) provide a powerful tool to classify data automatically. They are very useful in finding patterns of productions of a zone. The patterns are obtained only with data of the power of wind farms through competitive neural networks and self-organizing feature maps (SOFM).

Time variations of wind farm output have been studied in the third analysis. Only hourly data is available, so only slow variations can be analysed (such as down and dusk variations). Two tools have been used in the time study: correlogram and spectrogram of each wind farm.

In addition to this, those patterns found by ANN have been compared in the forth analysis with meteorological data of a nearby station.

Index Terms— **wind energy, wind farm, artificial neural network.**

I. INTRODUCTION

The following paper treats in deep about power generated in one year (registered each hour) in the wind farms in an area in Spain. The position of the wind farms is in the middle of a valley, and they are less of 170 km apart one to another. Because of these short distances, their performance is supposed to have a good correlation.

Their powers are evaluated and averaged each hour, and all the wind farms are grouped in nine groups (some farms that are very near are considered to have only one value). In the final part of this study, data from a meteorological tower located in the middle of the valley are also used.

We only have hourly data, so it is not possible to study neither fast fluctuations nor gusts in the power production. It is possible, however, to study the oscillations of several hours of duration, as the diurnal/nocturnal variations in wind speed and these due to the meteorological evolution in the area.

So, the study presented here is very useful to study the stationary performance of wind farms in the power system (i.e. to calculate the power flow in a small number of cases type that can be considered as representatives of the most probable states in the wind farms). This study is also useful as a very first point of departure to estimate the power production of wind farms in order to plan the generation in an electrical power system.

II. STATISTICAL ANALISYS

The analysed data are expressed in per unit, so the productions of all the wind farms are of the same magnitude. Some errors have been detected in the original data (outliers). The adopted solution consisted in eliminate this entries (8233 hourly data are left).

Another difficulty in the study of the data is that some wind farms are been enlarged, and some are quite new (therefore, maintenance stops are more probable to happen). It is also necessary to consider that the stopping due to fails and maintenance can cause some distortion in the data.

Nevertheless, some of these wind farms are going to present bigger productions in relation with their nominal power. This can be seen in the figure 1, in which one can observe that the wind farms with less production in per unit are numbers 6, 7 and 8. This can be due to have a location with a greater wind resource, or to dispose of WT with larger energy production / rated power ratio. Other point is that measured data are apparent power, not active power. However, wind farms in Spain operate with power factors very close to 1, so all conclusions derived from apparent power can be assumed to be valid for the active power.
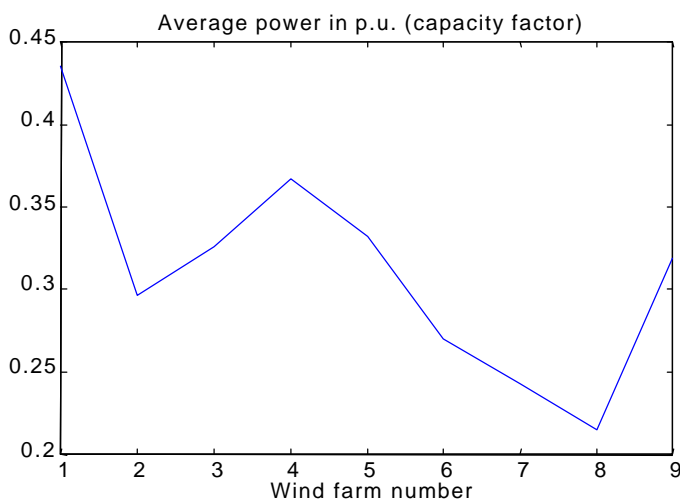

Figure 1: Average of apparent power of wind farm in per unit (with reference to rated power) .

Figure number 1 gives us an idea of the average productions. With the histogram of the figure number 2, we can see the distribution of the productions en each wind farm (i.e. how much time the farm is not generating, with low production and with full generation).

For instance, looking at the figure number 2 we can conclude that the wind farms 6,7 and 8 are not generating the 30% of the time (besides, these farms are the ones that have less production). Wind farm number 9 is producing at 10% $\pm$ 5% of his nominal power during the 22% of the time.

However, the first wind farm is producing about 100% of its nominal power during the 10% of the time. One can observe even that in certain periods, the wind farms 1, 4 and 9 produce slightly above their nominal power.
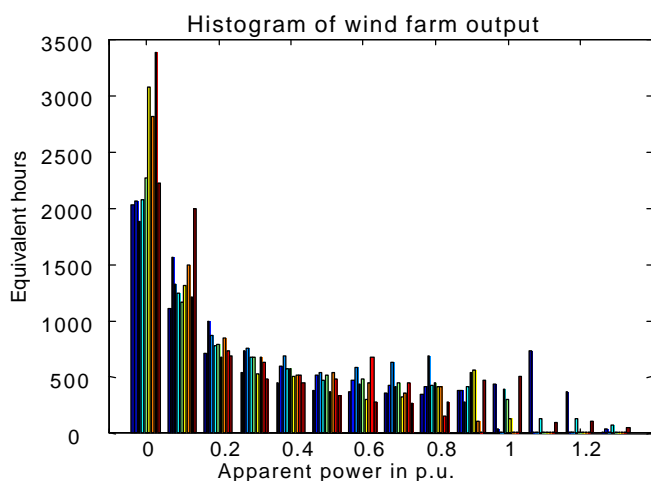

Figure 2: Histogram of apparent power of wind farms.

We can also study the relationship between the productions in each wind farm. Every element of the correlation's matrix shows how similar the productions of two wind farms are. If the productions of both parks are identical, the correlation will be the unity (100%, brownish red colour in figure 3), and if the productions are not correlated at all, the factor will be zero (blue colour) –full colour version of the figures can be seeing in the CD-ROM proceedings of ICREP '03–. That is why the elements in the main diagonal of the matrix are all one. The matrix is symmetrical because of it is equivalent to compare the park A with the B or vice versa.

In the following figure we can see the correlation's matrix. The squares with hot colours (the red ones) show that the relationship between the wind farms indicated by the row and the correspondent column have a high correlation. However, a cold colour indicates that the relationship between these two farms is weak.
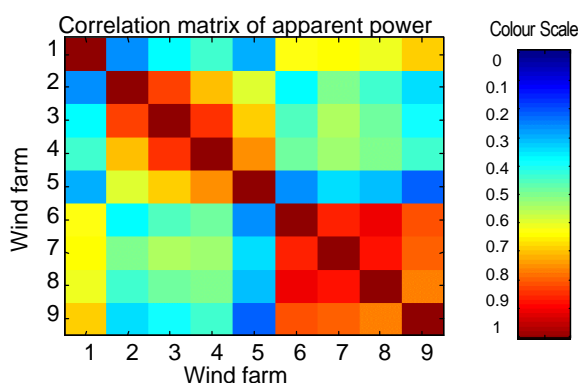

Figure 3: Correlation matrix, codified by colors.

Attending to this graphic, we can classify the wind farms in two groups:

- The wind farms number 2 to 5 show a great similitude between them, because of in the intersection of the correspondent rows and columns have hot colours, whereas in the rest of the rows the correlation is lower (colder colors). In this group, the most representative

wind farm are the third and the fourth ones, and the one that present a more different performance is number 5.
- The wind farms 6, 7, 8 and 7 show also a very good similitude among them. The least representative wind farm in this group is number 9.
- The wind farm number 1 presents behaviour more similar to farm 9. However, this wind farm is the less related to the rest.

Another conclusion that we can obtain from this data is that wind farm 1 and 5 are the ones that present more different behaviour from the average.

III. SEARCH OF OPERATION PATTERNS

A. CLASSIFICATION IN 10 PATTERNS

If the relationship among the wind farms is linear, the statistical study is sufficient. However, ANN have been used to detect non-linear ways of functioning. In figure number 4 we can see the operation patterns obtained from clustering all data in 10 patterns, codified by colours. Each row corresponds to one pattern and each column corresponds to the production in each one of the nine farms analysed. So, if we see in one row that the correspondent column to one farm is red, that farm has a high production in this pattern. If we observe the blue colour in the column of a farm, the production of this farm is very small in this pattern.

Although it is necessary to use the bias in the first phase of training to obtain a balanced network (each pattern with a similar occurrence frequency) an without "dead neurones". After that, the bias is forced to zero and a second phase of training has been realized to obtain a network with a lower error of classification (at dispense of that all the parameters do not have the same relative frequency). Another advantage of having a bias zero is that the correspondent pattern to a production is that one that is nearer, i.e., the one that has a lower Euclidean –geometric– distance. Consequently, to have a bias zero is more intuitive.

After a first training with bias, a second training without bias has been used in order to check its effect. The patterns have no changed significantly in the second training. The small differences are due to the fact that very similar patterns have evolved to situations of production slightly more specifics. As shown in figure 5, the first and last patterns are now more frequents (without production and with high production in al the wind farms).
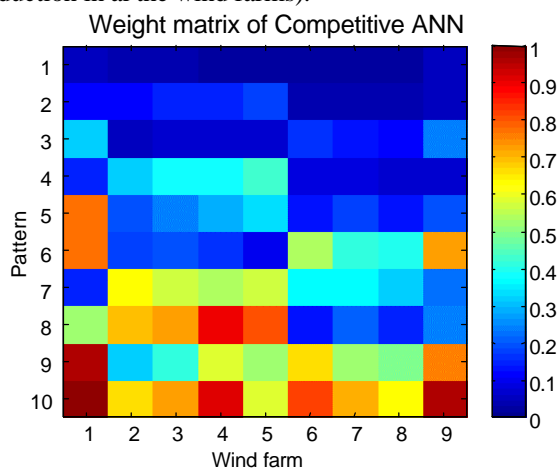

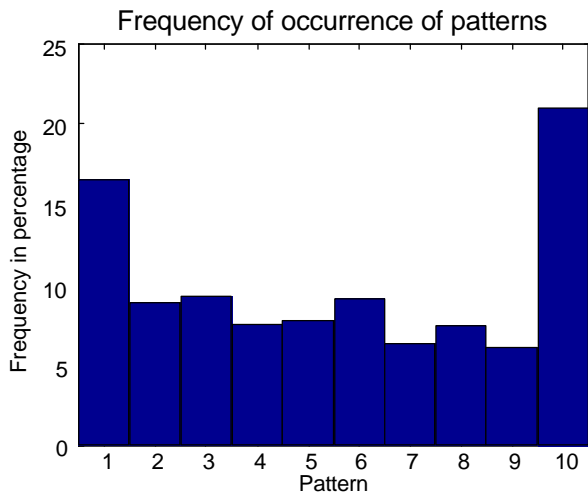Figure 4: Weight matrix of the Competitive ANN (patterns of wind farm power).

Figure 5: Frequency of occurrence of patterns of the Competitive ANN.

It is possible to measure the error committed for each wind farm in the classification (difference between the production of the nearer pattern and the real production). This value permits to know which farm is worst adjusted to the patterns (this can be used to consider if a wind farm is independent of other ones). The total classification error will be the geometric distance (in a space of dimension N=9 wind farms) between the measured point and the nearest pattern.

A.1 Low winds

The first conclusion that we can obtain is that the first four patterns correspond to no production in the nine wind farms. This result can be foreseen because the greatest part of the time, the wind farms do not inject energy to the power system (see Fig. 2).

As a curiosity, we can observe that in the first four patterns, the wind farm with a lower production is usually the number 8 because the blue colour is darker (in the histogram, this one was the wind farm that showed a greater frequency in the bin 0,05-0,15% p.u. power and besides, it is the one that has a lower average production).

The three first patterns correspond to productions of every wind farm below 12%. As we have three patterns of a total of ten, we can assign a 3x10% = 30% of the time to calm or null production. These patterns would correspond to calm periods, of weak geostrophic wind (geostrophic wind is that that exists in the layer of the atmosphere that is not in contact with the terrestrial surface, but to a high height that is not strongly affected by the orography, for instance to one height of 800 kPa –approximately 1200 meters above the sea level–).

The fourth case is surprising because, in the conditions in which the wind farms 6 to 9 are still, the number 5 presents a production of the 33% (and in a lesser extent, the farms 2,3 and 4). Studying the matrix of correlation we saw that the farms could be divided in two groups: farms 2 to 5 in one hand, and farms 6 to 9 in the other). In this case it would be necessary to verify if the WT installed in wind farm 5 can take advantage of lower winds or if there exist some orographic characteristics that reinforce geostrophic wind with a certain direction.

A.2 High winds

The pattern 9 and 10 correspond to high production. The pattern 10 corresponds to the maximum production, characterised by high productions in every farm. A remark: although the productions in pattern 10 are high, some farms are above 90% of rated production.

The wind farms that present a higher production are the farm 1 and 9. This can be observed in the histogram directly. For instance, the farm 1 stays during more than 1000 hours a year producing above 105% p.u. and wind farm 9 stays for more than 1000 hours above 85%. (The maximum power of these farms is lower than the nominal power. This can be due to the increase of the installed power in the farm without having updated the nominal power in the database).

B. CLASSIFICATION IN 50 PATTERNS

Once classified the usual functioning of a group of farms, we are going to make a search of a bigger number of patterns. This will permit us the visualization of some critic case, that although is presented only a 2% of the time (some 44 hours of functioning type in one year), it has a special interest (for instance, very high productions in every farm).

In the case of 50 patterns, the topology of the competitive neuronal network is no efficient enough and the process of learning is quite sensitive to the parameters. , the algorithm of training has some difficulties to choose a high number or patterns, increasing the number of neurones in the competitive network. As a consequence, the polarization or bias of the neurones increases significantly in order to not produce "dead neurones". Having a high bias, the classification is not so good.

For the election of 50 patterns of functioning, we have chose to use self-organised maps (SOFM). This solution is more logic, due to the fact that these types of networks are more appropriated to solve problems of automatic classification when the number of pattern to extract is elevated.

We have decided to use the one-dimensional connection among the neurones. If the wind farms analyzed corresponded to two or three meteorologically independent zones, a two-dimensional or a three-dimensional scheme would have been more appropriate. To compensate the low connectivity of the one-dimensional network, a parameter of vicinity of 3 is used instead of the usual 2. The result of the self-organised map can be seen in graphic number 6. Besides, one can observe that using a linear connection, the patterns are organized. The average error of the classification of a production to a pattern is 0.0045 p.u.
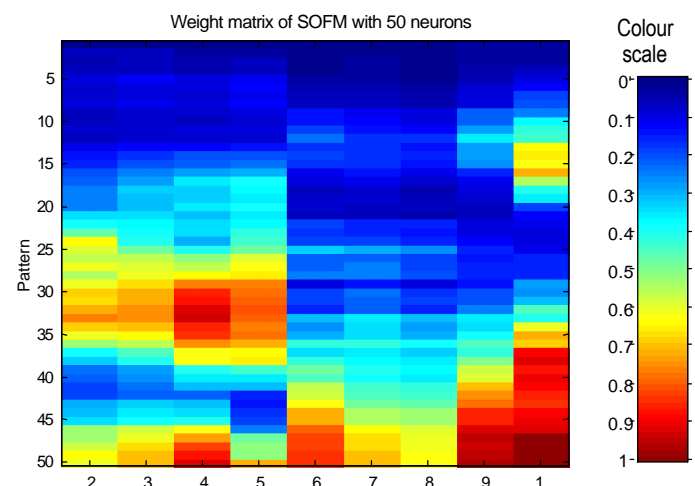

Figure 6: Weight matrix of SOFM with 50 neurons

In figure 7 we can see the frequency of occurrence of each pattern (with a number as high of patterns is difficult to obtain that all of them have a similar frequency of occurrence).
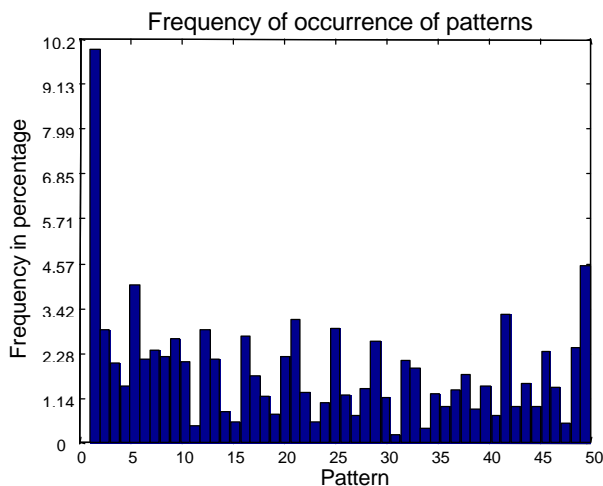

Figura 7: Frequency of occurrence of patterns of SOFM.

A remark: in figure 6 the wind farm 1 have been moved next to the farm 9 in order to produce a nicer scheme. This is due to the fact that the productions of both farms are relatively correlated.

IV. FREQUENTIAL STUDY

In order to study the evolution of the power in each farm, a joint time-frequency analysis has been used, by means of the spectrogram. This tool gives us, in each farm, the distribution of the oscillations of the power as time varies. The horizontal axis is the axis of time and the vertical axis is the one of frequencies. A hot colour (red) indicates that in that moment an oscillation of power have been produced, with the frequency associated to the vertical axis. The colour at a point indicates the amplitude (in dB) of the fluctuation content of that frequency at that time.

To understand better this graphic, we are going to see some examples. A frequency of 1 would indicate that a daily oscillation of power has been produced. This frequency appears in every farm because of usually at dawn the wind speed is minimum and maximum at dusk due to the warm-up of the surface by the sun (for instance at 7 h. the power usually is lower and at 19h. the power is usually bigger).

A frequency less than 1 would be associated to changes in the meteorology. A frequency of 0.25 would indicate that in a period of 4 days the power has lowered and has increased again of vice versa. In the spectrogram one can observe that the slow fluctuation of power (daily oscillations) is frequent because they have red colours associated.

You can also observe in the spectrograph that the intraday oscillations (the ones that are produced several times a day) are very few frequents. This means that although the production of the wind farms is continuously changing, these hourly variations have a random character; they have no periodicity nor a determined pattern. If the power oscillates several times a day during a significance period, this would be probably due to a periodic disconnection in the machines (disconnection by overheating in the gearbox with a high wind speed, etc.). These are the reasons why the superior part of the spectrograms has a colder colour (variations of some hours of period rarely occur), whereas the bottom part has a hotter colour (daily variations and meteorological changes).

There can be also observed that when an abrupt change in the production of the farm (for instance due to a disconnection of the farm or to an erroneous data), this produces a column of hotter colours. This is due to the fact that a steep change in the power, no cyclic, generates a wide spectrum of frequencies.
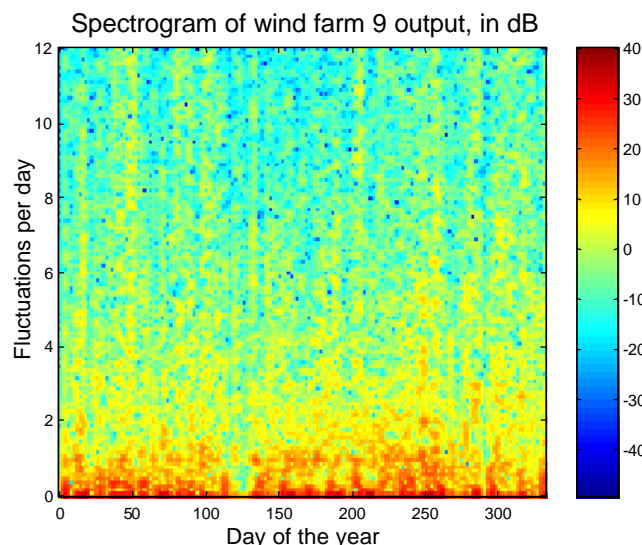

Figure 8: Spectrogram of power of wind farm No. 9,

Given the characteristics of the used data, it is not possible to study the fast oscillations of power due to gusts, disconnections and sudden changes of production. To this type of study it would be necessary data sampled several times by second and the period of study could be diminished proportionally.

To compare the variability of the power in several farms is preferable to observe the spectrum averaged during all the year. In figures 6 and 7 it is possible to observe that all the farms show a similar behaviour, although in some of them the power is slightest more fluctuating that in others. For instance, the daily modulation is lower in the farm 3 and it is superior in 1 and 9.
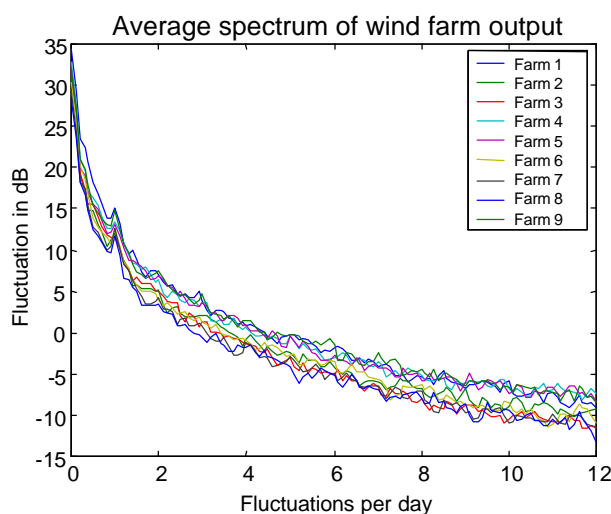

Figure 9: Power Spectrum Density (one year modulus average) of farm outputs. A peak can be noticed in 1 fluctuation/day.

V. STUDY OF THE TEMPORAL CORRELATION

The correlogram and the spectrogram give similar information and are related mathematically [1, 2]. Although

the spectrogram is more adequate for clearly periodic signals, the correlogram is more adequate for the study of random processes and time series. Another possibility to identify the system is to use parametric means to the identification of systems, but these have the disadvantage that it is necessary to make suppositions about the model of the system [3, 4].

The correlation coefficients have been used in the graphics to compare the results in wind farms that has different variance. Due to the fact that the average of power is not subtracted from the data, the correlogram is not nullified with increasing lag times. You can also observe that the graphic present some fluctuations every 12 hours due to a meteorological cycle every 12 and 24 h (each two peaks of 12h there is other one more noticeable due to the daily variation).

In figure 10 one can observe that the correlogram presents a sharp decrease in the first hours and then the decay is slower. This can be interpreted as the power in the farms varies between a 20 and a 28% of its variance in the first 5 hours (the correlation coefficient is between 0.8 an 0.72 depending of the farm). With a lag of 10 hours, the production has changed between a 28 and a 35% of its variance.

As in the frequential study, the farm with bigger power variations is the number 1, and the one of the lower variations is number 8. However, this difference is clearer in the correlogram.
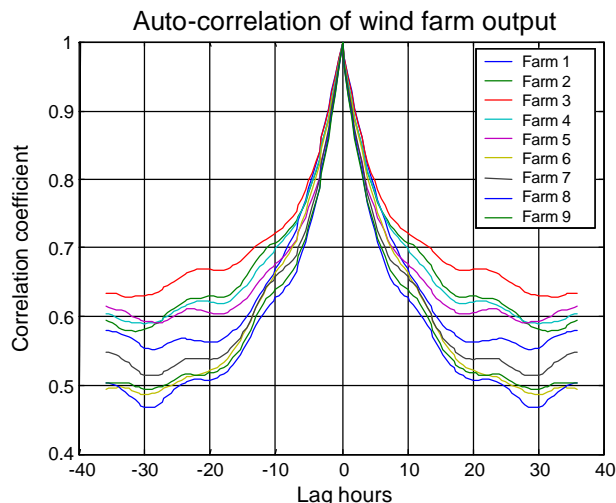

Figure 10: Auto-correlogram of wind farm output.

To summarize, the power of a farm usually varies in the 5 first hours in an interval of ±20% to an ±28% of its variance (depending on the farm) with an confidence interval of 68% (supposing that the generated power by the farms is a random process that follows a normal distribution). From the 10 h, these intervals slowly increase. One can even observe that the following day, to the same hour, the interval is of ±34% to ±50% of the variance of each farm, with the same confidence interval. To obtain a confidence interval of 95% in p.u. power, we must multiply the previous percentages by 1.96 and by the variance of each farm.

In figure 11 the correlation of the farm 1 is shown, with respect to the rest of the farms. The corresponding coefficient to zero lag hours corresponds to the first row of the matrix of correlation.
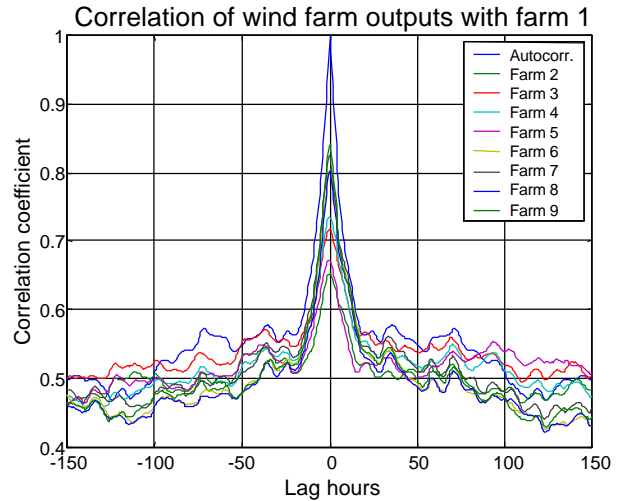

Figure 11: Correlogram of wind farm output respect wind farm 1.

The farm with a more similar behaviour to the 1 is the 9, and the more different is number 2. You can observe, as in figure 10, that there is a fast decay of the correlation in the first 24 h. From this level, the correlation tends to the stationary value. The ripple due to the cyclic behaviour is noticeable: the production presents patterns that are reproduced every 12 and 24 h.

A.1 Additional considerations about the measurements

The period of integration has a filtering effect on the power fluctuations. Fluctuations due to the tower shadow, gusts, etc. is studied from power measured at least at a sample rate of 10 Hz. However, in this work the data integrated each hour are treated and so these fluctuations (in the range of seconds and milliseconds) do not appear.

If the prediction of the production is desired, it is desirable to have in mind the meteorological prediction and the stopping for maintenance. Nowadays, there is a big effort in predicting the wind (for WT control) and the farm output (for electrical dispatch plan) [5, 6, 7, 8, 9].

VI.   RELATIONSHIP BETWEEN THE ENCOUNTERED PATTERNS AND THE METEOROLOGICAL PARAMETERS..

A meteorological station located approximately in the centre of the farms has been used to study the meaning of the found production patterns. Although the wind speed is measured at some 10 m. above a roof, the data are valid enough to our aim.
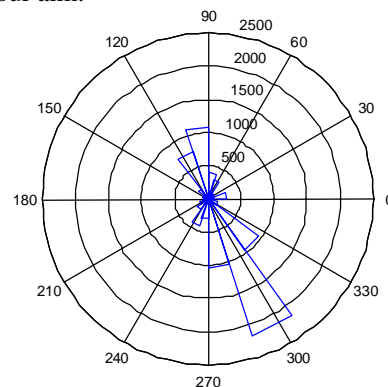

Figure 12: Wind direction rose of a nearby meteorological station.
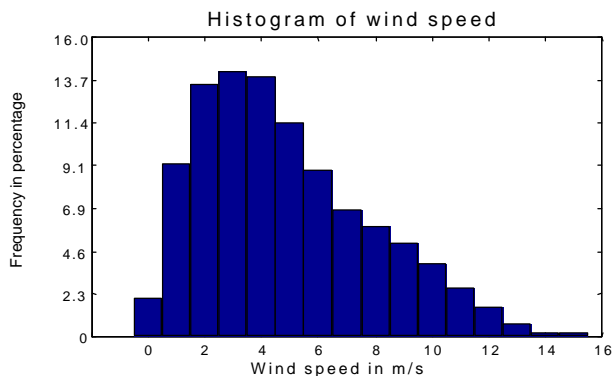
Figure 13: Histogram of wind speed at the meteorological station.

A. PATTERNS CHOSEN FOR THE METEOROLOGICAL STUDY

The patterns chosen for the meteorological study are shown in figure 14. Basically, they are the same as those of figure 4. The patterns 1 and 2 have been substituted here by a null production pattern and the pattern 10 has been added, corresponding to a 100% production in all farms.

This classification does not significantly increase the classification error and it is very natural to consider no production and full production schemes.
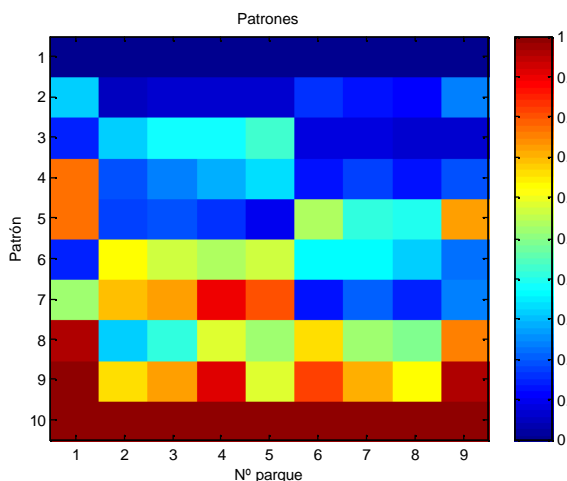

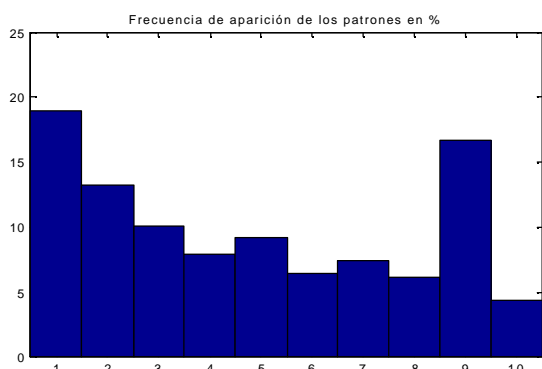Figure 14: Patterns of wind farm output utilized in the meteorological study.

Figure 15: Frequency of occurrence of patterns.

The relative frequency of patterns (see fig. 15) varies because the bias has been nullified in the second training, two patterns have been removed and other two patterns have been added.

Full production (1 p.u. production in all farms) is the least frequent pattern (4,4 %). Nevertheless, pattern 9, being very

similar to 10 but having less production in farms 2,3, 5 and 8, shows a high frequency (16.6 %). Although full production has a probability inferior to 5 %, it is worthy to study because it can result in electric overload of some tie lines and transformers.

B. METEOROLOGICAL PARAMETERS

In order to obtain the relationship between the meteorological parameters and the patterns, the histogram of the meteorological parameters have been computed for every pattern. The vertical axes of the graphics are the relative frequency in % of the pattern minus the relative frequency in % of the whole data. In other words, vertical axes are the deviation of the histogram whereas the horizontal axes are the values of the meteorological variable under study.

On the one hand, if the histogram of a meteorological variable has the same frequency distribution than the average of data, then the variable is not related to the pattern. On the other hand, if the histogram deviates significantly from the average, then the variable is related to the pattern.

If the difference between the pattern histogram and the average of the data is positive in a range, the pattern will happen more likely if the meteorological parameter is in this range. Correspondingly, a negative difference means that the probability of this pattern to appear is lower when the meteorological variable is in this range.

B.1 Influence of wind speed

In order to study the wind speed associated to each of the patterns, the deviation with respect to the average has been calculated. The horizontal axis of figure 16 is wind speed and vertical axis is the deviation of occurrence frequency in %.
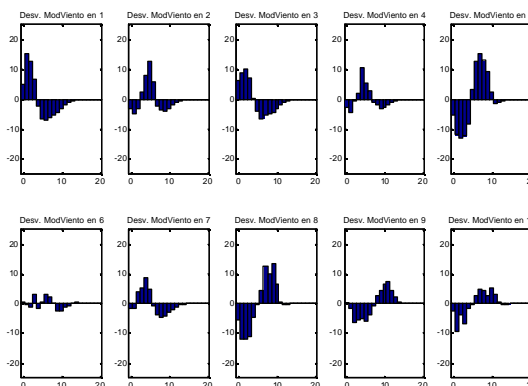

Figure 16: Deviation of histogram of wind speed for each pattern.

Patterns 1 and 3 are associated with low winds and calms at the meteorological station because there is a positive deviation in winds lower than 3,5 m/s. The null production pattern (1) is associated with the lowest wind speeds.

Patterns 2, 4 and 7 are associated to medium speeds, between 5 and 7 m/s. Surprisingly, pattern number 7 is associated to moderate wind speed although the production is high. Pattern number 5 is related to winds in the 5 to 10 m/s range.

The patterns corresponding to the highest production (number 9 an 10) are associated to the strongest winds. Nevertheless, the full production pattern is associated to wind which are not as strong (9 to 11.5 m/s) as those associated to pattern 9 (8 to 14 m/s). This is due to the fact that fixed blades, fixed speed turbines yield its maximum

output at its nominal wind speed. If the wind surpasses that speed, the production decreases due to aero dynamical stall.

Pattern number 6 shows a frequency distribution quite similar to the average (the deviation from the average is low, as it can be seen in Fig. 16). This fact indicates that this pattern is not related to a specific range of wind speed. We will see in the next sub-section that pattern # 6 is clearly connected to wind direction.

B.2 Influence of wind direction

Wind direction is a meteorological variable strongly interrelated to the patterns. This strong relationship is noticed by the high deviation of frequency of occurrences of patterns from average (more than 25 % deviation in some cases). A positive deviation indicates that the pattern is more likely to happen if that wind direction is measured at the meteorological station. If the deviation is negative, the pattern scarcely happens at that wind speed.

The predominant wind of the valley corresponds to NO direction (from 270 to 310º); we will call it hereafter wind A. Wind B is a less usual wind and blows in the opposite direction. Wind B has South direction and it typically corresponds to lower speeds than A. Wind called C has SO direction (from 210 to 260º) and it is the least frequent.
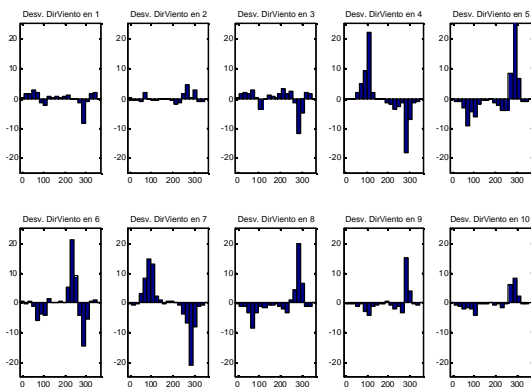

Figure 17: Deviation of histogram of wind direction for each pattern.

From Fig. 17, it can be seen that patterns number 5, 8 and 9 correspond to wind A. Patterns number 4 and 7 correspond to moderate speed wind B. Pattern 6 corresponds to wind C.

Low production patterns 1, 2 and 3 are not related to specific directions. This makes sense because in calm, the direction of the wind is not significant.

Full production pattern 10 yield a speed distribution similar to average, perhaps a little bit more likely with wind A than wind B because wind A normally has a higher speed.

B.3 Influence of relative humidity

Relative humidity is not a relevant parameter (see Fig. 18), except for pattern 7. This pattern is associated to relative humidity close to 100% (that means that this pattern occurs more often in rainy, cloudy or foggy days).
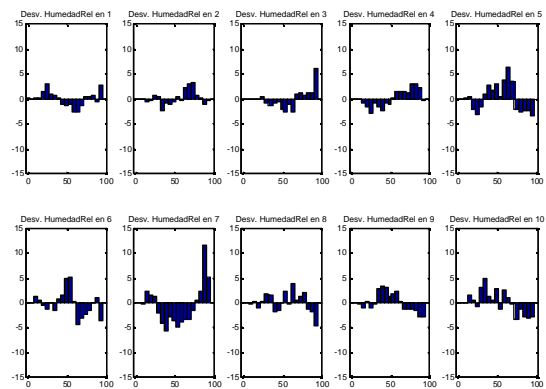

Figure 18: Deviation of histogram of relative humidity for each pattern.

B.4 Influence of the pressure

Even though pressure is not very significant, pattern 7 (and to a less extent patterns 3 and 6) are associated to low atmospheric pressures. Moreover, for pattern 7 the relative humidity is high, the temperature is moderate (about 10º) and the wind type is B.

Pattern 8 (and to a lesser extent patterns 9 and 5) correspond to high pressures and strong A wind.
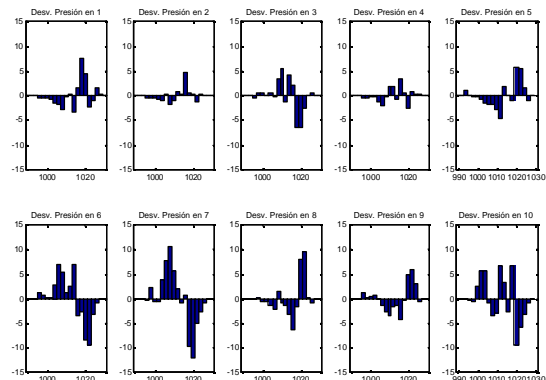

Figure 19: Deviation of histogram of pressure for each pattern.

B.5 Influence of the temperature

The temperature is a variable of secondary relevance in this study. Its deviations with reference to the average rarely reach 5%. The main connection is due to the fact that wind B is associated to a higher temperature and humidity than A or C winds.

VII. CONCLUSIONS

Artificial Neuronal Networks has proved a suitable tool to classify wind farm production. There are strong connections among the output of the wind farms of a zone and they are difficult to establish in an analytical way.

When the wind farms are quite related, as the case presented in this paper, a few number of patterns is enough to characterise wind farm output.

Competitive Neural Networks and Self-Organised Feature Maps are the most suitable topologies for wind farm output classification. Whereas the Competitive Network is preferred when are classifying in a reduced number of patterns, auto organized maps are more suitable for a higher number of parameters. Another advantage of these networks is that they have a direct physical meaning, which has allowed understanding the behaviour of wind farms in a given area.

Competitive networks and SOFM converged easily during the training. However, other topologies as radial basis network and LVQ were sensitive to the training parameters.

The meteorological parameters most related to the patterns (those with more than 20% of deviation from average) are speed and wind direction. Medium correlations (up to 10%) are found for pressure and temperature. Low influence has been found for relative humidity (around 5%).

REFERENCES

[1] Box, G.E.P. and Jekins, G. M. "Time Series Analysis: Forecasting and Control", 2nd Ed., Holden-Day Series, 1976

[2] Análisis de series temporales. Modelos ARIMA. Ezequiel URIEL. Ed. Paraninfo.

[3] Ljung L. And T. Glad. *Modeling of Dynamic Systems*, Prentice Hall, Englewood Cliffs, N.J. 1994.

[4] Lennart Ljung. Systen Identification Toolbox For Use with MATLAB. User's Guide. 1997

[5] L. Landberg. "Short-term prediction of the power production from wind farms" J. Wind Engng. Ind. Aerodyn., 1998.

[6] T. S. Nielsen, A. Joensen, L. Landberg et alt. "A new reference for wind power forecasting". Wind Energy, vol. 1, pp. 29-34. 1998

[7] A.G. Dutton, G. Kariniotakis, J.A. Halliday and E. Nogaret. "Load and Wind Power Forecasting Methods for the Optimal Management of Isolated Power Systems with High Wind Penetration" Wind Engineering, Vol 23/2. 2000.

[8] D.A. Bechrakis and P.D. Sparis. "Wind Speed Prediction Using Artificial Neural Networks". Wind Engineering, Vol 22/6. 1999.

[9] Y. D. Song. "A New Approach for Wind Speed Prediction". Wind Engineering, Vol 24/2. 2001.