

Short-Term Wind Power Forecast Based on Cluster Analysis and Artificial Neural Networks

Javier Lorenzo¹, Juan Méndez², Modesto Castrillón¹, and Daniel Hernández¹

¹ Universidad de Las Palmas de Gran Canaria
Instituto Universitario SIANI

Campus Universitario de Tafira- 35017 Las Palmas - Spain

² Universidad de Las Palmas de Gran Canaria
Departamento de Informática y Sistemas

Campus Universitario de Tafira- 35017 Las Palmas - Spain

Abstract. In this paper an architecture for an estimator of short-term wind farm power is proposed. The estimator is made up of a Linear Machine classifier and a set of k Multilayer Perceptrons, training each one for a specific subspace of the input space. The splitting of the input dataset into the k clusters is done using a k-means technique, obtaining the equivalent Linear Machine classifier from the cluster centroids. In order to assess the accuracy of the proposed estimator, some experiments will be carried out with actual data of wind speed and power of an experimental wind farm. We also compute the output of an ideal wind turbine to enrich the dataset and estimate the performance of the estimator on one isolated turbine.

1 Introduction

Nowadays there exists a wide consensus about the global warming [6]. The implication of the human activity in this climatological phenomenon has not been proved because there exists a cyclic component too. However, it is clear that the burning of fossil fuel increases the concentration of greenhouse gases that accelerates the warming effect, hence the growing interest in the use of renewable energy sources that will reduce the emission of them. Among the available alternatives, wind is a promising one, showing a steady expansion.

Unfortunately, the wind is not constant and it can be considered as a chaotic system whose predictability is limited. This fact along with the increase in installed power capacity have made that in many countries research groups have been granted to develop forecasting systems [4,17].

Depending on the forecast horizon, models can be divided into very short-term, short-term and long-term models. In each country, the Transmission System Operator has to deal with the management of the electric system in the different control and planning levels and also with the power production schedules in power plants. So the very short-term and short-term forecasting of wind power production becomes essential [1].

The autoregressive (AR) based models such as ARMA, ARX and Box-Jenkins methods have been used historically for very short-term wind forecasting up to few hours ahead [1]. Artificial neural networks (ANN) have been also used for wind or power forecast due to their ability of dealing with non linearities unlike AR models. Mohandes et al. [14] presented a comparison between AR model and neural networks for wind speed prediction and conclude that the ANN model outperforms the AR model in both one and several days horizon. Another comparison between regression and ANN models was presented by Li et al. [11] using as input the speed and direction of the wind measured in two meteorological towers. They found that Multilayer Perceptron ANN model outperforms the best regression model, which is a 3rd degree polynomial. More recent works have also confirmed the validity of ANN models for power forecasting [13,10].

Some researchers have proposed hybrid models. Ramirez-Rosado et al. [16] compared two systems for wind power forecast: FORECAS and SGP. In FORECAS the Power Curve Model (PCM) of each turbine was obtained with a Multilayer Perceptron neural network whose inputs were numerical weather prediction forecast values. The output of the system was obtained from a neural network that combines the PCM and AR outputs. On the other hand, SGP system is made up of 12 different models: nine neural networks used to forecast in the very short-term horizon and three additional models based on Elman and Modular neural networks. Thus, depending on the forecast horizon, the models were selected to forecast the hourly mean of electric power generated. Kusiak and Li [9] presented an approach of several power predictor which were the results of a previous clustering analysis. In their work, instead of using a combination of NN for each horizon, they customize a NN for each subspace of the input domain.

The aim of this paper is to present a hybrid architecture for short-term wind power forecast. The proposed approach differs from one presented in [9] in two aspects. On the one hand, the estimation is done based on data that are easily available in a wind farm, namely, wind speed and generated power. On the other hand, a short-term power prediction scenario is considered instead of a very short-term scenario. To evaluate the validity of the proposed architecture some tests are carried out with actual data from an experimental wind farm in the North of Spain.

The paper is organized as follows: in section 2 the estimator proposed in this work is explained. Experiments are shown in section 3 and in section 4 conclusions and further works are presented.

2 Proposed Hybrid Estimator

Some works have demonstrated the validity of Multilayer Perceptrons (MLP) as function approximators of scalar continuous functions [2,8]. The aim is to find a mapping between a set of samples (\mathbf{x}_i, y_i) where $y_i = f(\mathbf{x}_i)$, being f the unknown scalar underlying function: $f : R^n \rightarrow R$. The set of all samples made up the training dataset, $D = \{(\mathbf{x}_i, y_i)\}$, under the assumption that they are independently and identically distributed.

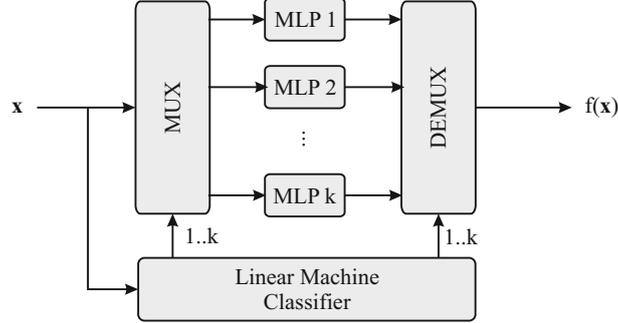


Fig. 1. Architecture of the proposed combined estimator

The set of nonlinear weight equations that relates the neural network's adjustable parameters to the data can be obtained with optimization-based techniques [12,7,5]. These methods minimize the error function, normally the Root-mean-square error (RMSE), over the whole training dataset. If the underlying function f is too complex due to the nature of the problem as in short-term wind power prediction, the complexity of the MLP must be increased to approximate better f from the training dataset. Consequently the optimization methods can get trapped in undesirable local minima.

Instead of using the original training set, the proposed architecture is based on a splitting of the input space into subspaces under the hypothesis that customized models in each subspaces yield higher prediction accuracy. Unsupervised learning methods are designed to split a set of samples into different subsets (clusters) where the samples of each cluster are similar among them and different from the samples of other clusters. Thus the initial training dataset D is divided into k non-overlapping subsets, $D = \cup_{j=1}^k D^j$, using the k-means clustering method [3], where each subset D^j is characterized by its centroid \mathbf{Z}^j .

The clustering process is equivalent to define k classes in the input space so when the power must be estimated for an unseen sample, it must be firstly classified to feed the corresponding estimator. This classification is done with a Linear Machine classifier $d(\mathbf{x})$,

$$d(\mathbf{x}) = \max_{j=1..k} \{d^j(\mathbf{x})\} \quad (1)$$

where $d^j(\mathbf{x})$ is the linear discriminant function associated to each cluster

$$d^j(\mathbf{x}) = \omega_0^j + \sum_{l=1}^n \omega_l^j x_l^j \quad (2)$$

whose weights ω_l^j are computed from the centroids \mathbf{Z}^j as

$$\omega_0^j = -\frac{1}{2} \|\mathbf{Z}^j\|^2 \quad (3)$$

$$\omega_l^j = \mathbf{Z}_l^j; \quad l = 1 \dots n \quad (4)$$

where \mathbf{Z}_l^j is the l -th centroid element and n is the problem dimension (lag).

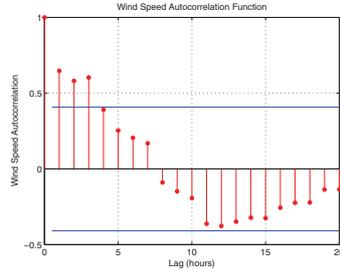


Fig. 2. Autocorrelation of the wind speed

In the wind power forecast problem, the samples \mathbf{x} are wind speed or power measured in n intervals previous to current time t . The function of interest is the estimated power at a horizon h , $\hat{p}(t + h)$. For each cluster D^j , a MLP is trained getting k estimators customized for each subspace of the input domain. The architecture of the estimator is shown in Figure 1. When the power for a new sample must be estimated, a first classification stage is carried out by the Linear Machine to decide which MLP will be in charge of the estimation.

3 Experiments

Experiments were made with actual wind speed and wind farm power data from Sotavento Galicia project. The wind speed series comprises from August 5th, 2009 until February 4th, 2010 with time steps of 10 minutes. Data were preprocessed to obtain mean hourly wind speed which yield a total of 4416 values. The data are divided into the training dataset (2/3) and test dataset (1/3).

The training dataset is used to split the input space into k clusters using the k-means algorithm, training a MLP for each cluster. After that, a Linear Machine Classifier is obtained which assigns every test sample to one of the k MLP. The output of the estimator for each test sample is the output of the MLP selected by the classifier (Figure 1).

Due to the random initialization of the k-means clustering method, 200 runs were done and the one with the lowest value for the sum of within-cluster point-to-centroid distances for the whole clusters is considered. Also to minimize the effect of the random initialization of the MLP weights, we provide the mean obtained from 25 training trials in order to reduce the uncertainty of the results.

As it is not possible to access to the power produced by only one turbine, the output of an ideal wind turbine whose transfer function has 5 and 12.5 m/sec cut-off values was included in the experiments. So, four different scenarios were considered in the experiments depending on the predicted variable and the inputs that feed the estimator:

Table 1. Number of neurons in the hidden layer of the MLP

	Prediction Horizon					
	1 hour	2 hours	3 hours	4 hours	5 hours	6 hours
<i>Scenario A</i>	4	7	6	6	9	6
<i>Scenario B</i>	8	8	7	8	6	8
<i>Scenario C</i>	2	2	1	2	2	2
<i>Scenario D</i>	2	1	4	4	6	4

Scenario A. The predicted variable is the wind farm power computed from the wind speeds, $\hat{p}(t+h) = f(v(t), v(t-1), v(t-2), \dots, v(t-n+1))$

Scenario B. The predicted variable is the wind farm power computed from previous wind farm power values, $\hat{p}(t+h) = f(p(t), p(t-1), p(t-2), \dots, p(t-n+1))$

Scenario C. The predicted variable is the ideal turbine output computed from the wind speeds, $\hat{p}_{vesta}(t+h) = f(v(t), v(t-1), v(t-2), \dots, v(t-n+1))$

Scenario D. The predicted variable is the ideal turbine output computed from previous ideal turbine outputs, $\hat{p}_{vesta}(t+h) = f(p_{vesta}(t), p_{vesta}(t-1), \dots, p_{vesta}(t-n+1))$

Instead of using the RMSE as measure to assess the accuracy of the results, the improvement over a reference model is carried out using,

$$Imp_{RMSE} = \frac{RMSE_{reference\ model} - RMSE_{proposed\ model}}{RMSE_{reference\ model}} 100\% \quad (5)$$

As reference model, the one proposed by Nielsen [15] is used which is an extension over the pure persistence model that also includes long-term information as a linear expression: $\hat{y}(t+k) = b + ay(t)$. Coefficients of the enhanced persistence model for wind farm power are: $a = 0.9487$ and $b = 37.5692$; and for ideal turbine output: $a = 0.8947$ and $b = 0.0281$.

An important parameter of the MLP topology for the experiments is the number of inputs n , that is, the lag time for the power prediction. To find it, the autocorrelation of the wind speed is obtained (Figure 2) and it is found that beyond 4 hours the influence is negligible so a lag time of 4 hours is considered. Bayesian regularization [5] was used to train the network. It uses as goal function a combination of squared errors and weights, and then determines the correct combination so as to produce a network that generalizes well.

The other parameters that define the topology of the network are the number of hidden layers and neurons in each layer. A one hidden layer topology is considered with hyperbolic tangent sigmoid activation functions in the hidden layer and linear activation function in the output layer. To decide the number of neurons, 25 training trails for 1 to 16 neurons in the hidden layer for each scenario and horizon were carried out. After that, an ANOVA statistical test was performed to find the number of neurons that yields better performance and the results are shown in Table 1.

Table 2. Imp_{RMSE} for different number of clusters and horizons (in hours) for the *Scenario A*

Prediction horizon	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8
$h = 1$	-92%	-78%	-67%	-74%	-69%	-66%	-65%	-67%
$h = 2$	-44%	-43%	-40%	-34%	-40%	-32%	-29%	-33%
$h = 3$	-25%	-25%	-26%	-23%	-24%	-18%	-18%	-21%
$h = 4$	-16%	-17%	-19%	-20%	-14%	-11%	-11%	-12%
$h = 5$	-9%	-15%	-15%	-13%	-8%	-8%	-6%	-6%
$h = 6$	-5%	-9%	-10%	-9%	-5%	-5%	-3%	-2%

Table 3. Imp_{RMSE} for different number of clusters and horizons (in hours) for the *Scenario B*

Prediction horizon	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8
$h = 1$	83%	93%	91%	93%	93%	90%	91%	0%
$h = 2$	84%	91%	89%	90%	92%	84%	88%	0%
$h = 3$	72%	91%	88%	90%	90%	90%	92%	91%
$h = 4$	70%	74%	70%	76%	76%	77%	79%	79%
$h = 5$	54%	66%	64%	73%	73%	73%	72%	72%
$h = 6$	52%	63%	62%	70%	70%	70%	70%	66%

Now the results of every experiment are analyzed. Table 2 shows the results obtained for the *Scenario A* where the wind speed of the previous 4 hours are used as input, $\hat{p}(t+h) = f(v(t), v(t-1), v(t-2), v(t-3))$. In this scenario the MLP model performance is lower than the reference model for all the configurations. Analyzing the effect of the cluster stage in the performance of the estimator it can be observed that the performance increases for almost all the horizons with a maximum in $k = 7$. With more clusters the performance decreases because the obtained MLP are overspecialized and the generalization capability of the estimator is worse.

Results for *Scenario B* are shown in Table 3 where the 4 previous hourly mean wind farm power values are used as predictor variables, $\hat{p}(t+h) = f(p(t), p(t-1), p(t-2), p(t-3))$. Unlike *Scenario A*, in this scenario the estimator surpasses the reference model giving better accuracy for all the horizons and number of clusters. The splitting of the input space gives better performance because for all the columns the results are higher than those of the column labeled as $k = 1$.

Tables 4 and 5 show the results considering an ideal turbine. When the power is estimated from previous wind speed (Table 4), the splitting of the input space with clustering analysis improves slightly the results obtained with the initial training dataset (column $k = 1$). On the other hand, the use of cluster analysis for the *Scenario D* makes slightly worse the results. However in these two scenarios where the output of an ideal turbine was used, the difference between the best and worst estimation is not so noticeable as in scenarios where actual data are used.

Table 4. Imp_{RMSE} for different number of clusters and horizons (in hours) for the *Scenario C*

Prediction horizon	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8
$h = 1$	-8%	-7%	-10%	-8%	-11%	-11%	-10%	-15%
$h = 2$	-6%	-6%	-7%	-7%	-4%	-4%	-3%	-4%
$h = 3$	1%	0%	0%	-1%	-1%	-1%	-1%	-2%
$h = 4$	-1%	-2%	-1%	-3%	-3%	-2%	-3%	-3%
$h = 5$	-2%	-1%	-2%	-2%	-2%	-2%	-3%	-2%
$h = 6$	-2%	-3%	-1%	-1%	-1%	-3%	-2%	-2%

Table 5. Imp_{RMSE} for different number of clusters and horizons (in hours) for the *Scenario D*

Prediction horizon	k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8
$h = 1$	1%	0%	0%	0%	0%	0%	-1%	-2%
$h = 2$	4%	4%	4%	4%	4%	3%	3%	1%
$h = 3$	7%	6%	5%	6%	6%	5%	5%	4%
$h = 4$	8%	7%	6%	7%	6%	7%	6%	6%
$h = 5$	9%	8%	8%	8%	7%	8%	7%	7%
$h = 6$	10%	9%	8%	9%	9%	8%	8%	8%

4 Conclusions

This paper proposes a hybrid estimator for power prediction. The estimator is composed of a Linear Machine and a set of customized MLP. The Linear Machine classifies the samples into one of several subsets which has been previously obtained with a clustering analysis. The proposal has been tested on actual data from a wind farm and data from an ideal wind turbine, both for different prediction horizons.

From the obtained results it can be concluded that for the actual data, which are of interest to industry, the proposed estimator increases the estimation accuracy compared to a single MLP. On the other hand, for the data of an ideal wind turbine the proposed estimator performance is similar to a single MLP. An interesting conclusion is that the wind farm power prediction is better done when power is used as predicting variable instead of wind speed. Another fact that the experiments has brought up and that it is in consonance with the nature of the persistence model, it is that as the horizon goes farther the proposed estimator yields better performance.

Acknowledgements

This work has been partially supported by the Canary Islands government through projects SolSubC200801000137 and ProID2010-0062, and by the Spanish government and FEDER through the project TIN2008-06068.

References

1. Costa, A., Crespo, A., Navarro, J., Lizcano, G., Madsen, H., Feitosa, E.: A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews* 12(6), 1725–1744 (2008)
2. Cybenko, G.: Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)* 2(4), 303–314 (1989)
3. Duda, R.O., Hart, P.E., Stork, D.G.: *Pattern Classification*, 2nd edn. Wiley Interscience, Hoboken (2000)
4. Focken, U., Lange, M., Monnich, K., Waldl, H., Beyer, H., Luig, A.: Short-term prediction of the aggregated power output of wind farms - a statistical analysis of the reduction of the prediction error by spatial smoothing. *Journal of Wind Engineering and Industrial Aerodynamics* 90, 231–246 (2002)
5. Foresee, F., Hagan, M.: Gauss-newton approximation to bayesian regularization. In: *Proceedings of the 1997 International Joint Conference on Neural Networks* (1997)
6. Hansen, J., Ruedy, R., Sato, M., Lo, K.: Global surface temperature change. *Reviews of Geophysics* 48, 1–29 (2010)
7. Haykin, S.: *Neural Networks; A comprehensive Foundation*, 1st edn. Macmillan, New York (1994)
8. Hornik, K., Stinchcombe, M., White, H.: Multilayer feedforward networks are universal approximators. *Neural Networks* 2, 359–366 (1989), <http://portal.acm.org/citation.cfm?id=70405.70408>
9. Kusiak, A., Li, W.: Short-term prediction of wind power with clustering approach. *Renewable Energy* 35, 2362–2369 (2010)
10. Kusiak, A., Zheng, H., Song, Z.: Wind farm power prediction: A data-mining approach. *Wind Energy* 12, 275–293 (2009)
11. Li, S., Wunsch, D.C., Ohair, E.A., Giesselmann, M.G.: Using neural networks to estimate wind turbine power generation. *IEEE Transactions on Energy Conversion* 16(3), 276–282 (2001)
12. Marquardt, D.: An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics* 11, 431–441 (1963)
13. Méndez, J., Lorenzo, J., Hernández, M.: Experiments and reference models in training neural networks for short-term wind power forecasting in electricity markets. In: Cabestany, J., Sandoval, F., Prieto, A., Corchado, J.M. (eds.) *IWANN 2009*. LNCS, vol. 5517, pp. 1288–1295. Springer, Heidelberg (2009)
14. Mohandes, M.A., Rehman, S., Halawani, T.O.: A neural networks approach for wind speed prediction. *Renewable Energy* 13(3), 345–354 (1998)
15. Nielsen, T.S., Joensen, A., Madsen, H., Landberg, L., Giebel, G.: A new reference for wind power forecasting. *Wind Energy* 1(1), 29–34 (1998)
16. Ramirez-Rosado, I.J., Fernandez-Jimenez, L.A., Monteiro, C., Sousa, J., Bessa, R.: Comparison of two new short-term wind-power forecasting systems. *Renewable Energy* 34(7), 1848–1854 (2009)
17. Sánchez, I.: Short-term prediction of wind energy production. *International Journal of Forecasting* 22, 43–56 (2006)