EXPERIMENTS IN SHORT-TERM WIND POWER PREDICTION USING VARIABLE SELECTION

Javier Lorenzo
Institute SIANI, Univ. de Las Palmas de Gran Canaria, Campus Tafira, 35017 Las Palmas, Spain

Juan Méndez
Dept. of Informática y Sistemas, Univ. de Las Palmas de Gran Canaria, Campus Tafira, 35017 Las Palmas, Spain

Daniel Hernández, Modesto Castrillón
Institute SIANI, Univ. de Las Palmas de Gran Canaria, Campus Tafira, 35017 Las Palmas, Spain

Keywords: Machine learning, Neural networks, k-NN, Short-term wind farm power prediction.

Abstract: In this paper some experiments have been realized to test how the introduction of variable selection has an effect on the predictor performance in short-term wind farm power prediction. Variable selection based on Kraskov estimation of the mutual information will be used due to its capability to deal with sets of continuous random variables. A Multilayer Percetron and a k-NN estimator will be the predictor based models with different topologies and number of neighbors. 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 effect of variable selection on one isolated turbine. This will allow us to define four different settings for the experiments which vary in the nature of the inputs to the model, wind speed, wind farm or isolated wind turbine power, and the predicted variable, wind farm or isolated wind turbine power.

1 INTRODUCTION

Society is very worried about the impact of the human activities on the environment being pollution a result of those human activities. Electricity production consumes a big amount of fossil fuel, which produces carbon dioxide, so the use of renewable energy sources will reduce the emission of it. Among the renewable energy sources, wind is a promising alternative with a increasing installed power capacity.

However, 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 (Landberg, 2001; Focken et al., 2002; Sánchez, 2006).

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 have 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 is essential (Costa et al., 2008).

The statistical models such as ARMA, ARX and Box-Jenkins methods have been used historically for very short-term wind forecasting up to few hours ahead (Costa et al., 2008). 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. (Mohandes et al., 1998) present a comparison between AR model and neural networks for wind speed prediction and obtain that the ANN model outperforms the AR model in both one day and some days horizon. Another comparison between regression and ANN models was presented by Li et al. (Li et al., 2001) using as input the speed and direction of the wind 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

Lorenzo J., Méndez J., Hernández D. and Castrillón M..
EXPERIMENTS IN SHORT-TERM WIND POWER PREDICTION USING VARIABLE SELECTION.
DOI: 10.5220/0003182703700375
In *Proceedings of the 3rd International Conference on Agents and Artificial Intelligence* (ICAART-2011), pages 370-375
ISBN: 978-989-8425-40-9

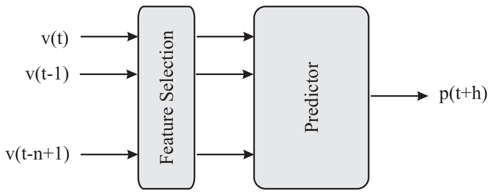

Figure 1: Wind power prediction based on previous wind speed with variable (feature) selection.

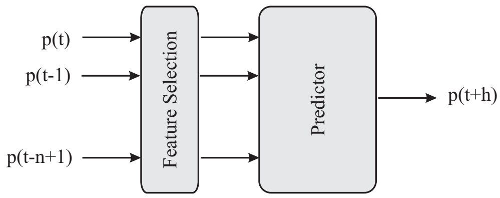

Figure 2: Wind power prediction based on previous power of the wind farm with variable (feature) selection.

models for power forecasting (Méndez et al., 2009; Kusiak et al., 2009).

The selection or extraction of features has demonstrated that can improve the performance of the induced model in classification tasks. In prediction of time series, feature selection has also been investigated. Kusiak et al. (Kusiak et al., 2009) propose both feature selection and extraction in short-term and long-term prediction based on NWP data provided in 16 locations. Feature selection is used to select the locations that have more influence on the prediction. Feature extraction is carried out with PCA to reduce the high dimensionality of the data provided by the models, 10 and 12 respectively for each location. Interactive Neural Filter (INF) (Crone and Kourentzes, 2010) is a method that remove the seasonality, trends and noise of a time series that do not explain it thus removing any bias to the predicting model. In (Yoon et al., 2005) a method called CLeVer based on PCA and descriptive common principal components (DCPC) is presented to select features in multivariate time series.

In this paper, we study the effect of feature selection in very-short term wind power forecast using mutual information. The prediction for different forecast horizons is computed both from wind speed as previous power generated using Multilayer Perceptron (MLP) and k-NN interpolation. The experiments have been carried out from real wind and power data of a wind farm. The period of time considered has been of six months which gives a better understanding of the variable selection incidence. To evaluate the results, the reference measure described in (Nielsen et al., 1998) is used.

The paper is organized as follows: in section 2, the methodology is presented. Experiments are shown in section 3 and in section 4 conclusions and further works are presented.

2 METHODOLOGY

In this work, the inputs to predict the wind farm power at a horizon h, $\hat{p}(t+h)$, are wind speeds from SCADA,

$$\hat{p}(t+h) = f(v(t), \ldots, v(t-n+1)) \qquad (1)$$

or alternatively, previous power generated by the wind farm,

$$\hat{p}(t+h) = f(p(t), \ldots, p(t-n+1)) \qquad (2)$$

Usually n is estimated with the aid of the autocorrelogram.

In these models, the assumption that all the n previous measures have the same relevance is done. However as in many other applications of forecasting, the ranking of the variables according to a relevance measure can improve the performance of the predictor. As relevance measure Mutual Information has been chosen because it has proved good results in feature selection (Chow and Huang, 2005). In time series, Ji et al. (Ji et al., 2005) propose the use of the mutual information to rank variables in time series prediction. Guillen et al. (Guillén et al., 2010) also make use of mutual information to prototype and variable selection in a decremental approach, where variables are removed while the mutual information is above a threshold. Here the estimation of mutual information proposed by Kraskov et al. (Kraskov et al., 2004), which is based on k-nearest neighbor statistics, is used. MILCA is a implementation of this method that can be downloaded from (MILCA, 2010).

A filter feature selection approach with a sequential forward selection strategy is going to be implemented (Figures 1 and 2), where a feature selection stage comes before the forecast module. To assess if any improvement (Imp_{RMSE}) is achieved, a comparison with a reference model (Nielsen et al., 1998) is carried out using,

$$Imp_{RMSE} = \frac{RMSE_{reference} - RMSE_{model}}{RMSE_{reference}} 100\% \qquad (3)$$

As reference model it is used the one proposed by Nielsen (Nielsen et al., 1998) which is an improvement over the pure persistence model because it also includes long-term information using the linear expression, $\hat{y}(t+k) = b + ay(t)$.

Table 1: Imp_{RMSE} for MLP model (8 neurons in the hidden layer) with different topologies and horizons (in hours) for the four settings.

Prediction horizon	*Setting A*	*Setting B*	*Setting C*	*Setting D*
$h=1$	-313.32 ± 2.89	-185.20 ± 3.32	50.10 ± 16.86	-0.59 ± 1.15
$h=2$	-176.02 ± 2.25	-110.72 ± 1.75	63.62 ± 11.92	3.76 ± 0.40
$h=3$	-122.67 ± 2.37	-82.11 ± 1.60	71.19 ± 11.34	6.74 ± 0.23
$h=4$	-91.52 ± 1.67	-67.05 ± 1.22	56.24 ± 9.89	8.03 ± 0.33
$h=5$	-70.08 ± 1.57	-55.53 ± 1.49	47.24 ± 6.90	9.24 ± 0.20
$h=6$	-54.93 ± 1.36	-48.08 ± 1.78	45.34 ± 9.72	10.24 ± 0.33

Table 2: Imp_{RMSE} for k-NN ($k=8$) for six prediction horizons (in hours) in the four settings.

Prediction horizon	*Setting A*	*Setting B*	*Setting C*	*Setting d*
$h=1$	-101.78	-10.20	95.98	-5.73
$h=2$	-50.62	-5.02	95.70	0.43
$h=3$	-32.94	-4.97	95.57	2.38
$h=4$	-22.14	-4.46	95.82	4.88
$h=5$	-14.25	-4.32	95.32	5.95
$h=6$	-8.91	-4.36	94.98	6.93

3 EXPERIMENTS

Experiments were made with actual wind speeds and wind farm power obtained from the website of 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 set was split in two subset, the train (2/3) and test (1/3). Two models to predict the very-short term wind power are considered: Multilayer Perceptron (MLP) and K-Nearest Neighbor (kNN). Due to the random initialization of the MLP weigths, we provide the mean and the standard deviation obtained from 25 training trials as: $\mu \pm \sigma$, 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 ($p_{vesta}(t)$) whose transfer function has 5 and 12.5 m/sec cut-off values is included in the experiments. So, four different settings are going to be considered in the experiments depending on the predicted variable and the inputs that feed the models:

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

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

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

Setting D. The predicted variable is the ideal turbine output computed from previous ideal turbine outputs, $\hat{p}_{vesta}(t+h) = f(p_{vesta}(t), p_{vesta}(t-1), p_{vesta}(t-2), \ldots))$

Variables in the previous settings are the wind speed, wind farm power or ideal turbine output. Coefficients of the enhanced persistence model (3) for wind farm power are: $A_0 = 0.9487$ and $B = 37.5692$; and for ideal turbine output: $A_0 = 0.8947$ and $B = 0.0281$.

3.1 Results without Variable Selection

To allow an analysis of the effect of variable selection in the problem addressed in this paper, at first the improvement over the reference model (eq. 3) without variable selection is carried out. The number of variables used for the model induction is set to four because we have auto correlated the power with itself and cross correlated it with the wind speed and concluded that the highest values are for offsets until the range of 4-6 hours back.

In the experiments three different topologies for the MLP were tested: 4, 6 and 8 neurons in the hidden layer, however results for the 8 neurons in the hidden layer are shown in Table 1 due to space constraints. In *Setting A* the MLP model performance is lower than the reference model and this behavior does not depend on the number of neurons in the hidden layer. Results are a bit better as the horizon is far away although they are still worse than the reference model. In *Setting B* the MLP exhibits a better performance

Table 3: Imp_{RMSE} for MLP model (8 neurons in the hidden layer) with different topologies and horizons (in hours) for the four settings using variable selection.

Prediction horizon	*Setting A*	*Setting B*	*Setting C*	*Setting d*
$h = 1$	-314.34 ± 3.00	-185.80 ± 3.32	46.97 ± 14.49	0.23 ± 0.79
$h = 2$	-177.31 ± 1.41	-112.11 ± 1.76	68.81 ± 9.82	4.00 ± 0.29
$h = 3$	-126.23 ± 2.54	-85.20 ± 1.34	71.70 ± 9.29	5.77 ± 0.33
$h = 4$	-95.29 ± 1.78	-70.33 ± 2.04	67.34 ± 7.33	6.97 ± 0.29
$h = 5$	-75.33 ± 2.19	-59.68 ± 2.61	56.29 ± 17.40	7.87 ± 0.43
$h = 6$	-59.26 ± 2.35	-52.13 ± 2.85	50.39 ± 20.27	8.92 ± 0.26

Table 4: Imp_{RMSE} for k-NN ($k = 8$) for six prediction horizons (in hours) in the four settings using variable selection.

Prediction horizon	*Setting A*	*Setting B*	*Setting C*	*Setting d*
$h = 1$	-90.16	-10.41	93.91	-6.74
$h = 2$	-40.98	-6.74	95.74	-1.38
$h = 3$	-26.62	-5.02	96.63	2.05
$h = 4$	-19.41	-5.06	94.61	2.59
$h = 5$	-12.62	-5.92	94.37	3.72
$h = 6$	-7.65	-7.01	95.73	4.48

but it is still worse than the reference model. No influence of the number of neurons in the hidden layer is observed and the farther the horizon is, the better the results are. When in *Setting C* the wind farm power is used as input variable, MLP model clearly surpasses the reference model with a maximum improvement of 71.19% for a horizon of 3 hours. Unlike previous settings, here the number of neurons in the hidden layer has influence over the results but there is no clear pattern because it also depends on the prediction horizon. In the latter setting, the MLP model achieves better performance than the reference model although not so high as in *Setting C* due to the absence of external interferences. Only for 1 hour horizon the reference model is slightly better than the MLP model which is explained to the high inertia of the atmosphere that is gathered in the ideal turbine output.

For the data used in this set of experiments it can be concluded that short-term wind power prediction based on a MLP model only improves a reference model based on persistence when the inputs are also power measures and not wind speeds.

The other model under study is the k-NN interpolation. In this model there is no random initialization of parameters so one run for each experiment is carried out. As the number of neighbors k used in the approximation can produce different outcomes, the results were obtained for values of k from 1 to 10. Due to space requirements only results for 8 neighbors is shown in Table 2.

The same four settings are used and the results achieved with the k-NN model (Table 2) display the same behavior than the results achieved with the MLP model. For *Setting A*, the k-NN model has a worse performance than the reference model but unlike in MLP in this model as the number of neighbors increases the result also improves although it never surpasses the reference model. In *Setting B*, k-NN neither improves the performance of the reference model although the results are better than in *Setting A*. Here the model also yields better results as the number of neighbor increases as in previous setting. k-NN model exhibits better improvement in *Setting C* than in the other settings. Here the improvements over the reference model are normally around the 95%. Unlike the *Setting A* and *Setting B*, in this case results do not always got better as the number of neighbors increases. Finally for *Setting D*, the results are better as the horizon and number of neighbors increase.

Comparing the performance of the k-NN with MLP without variable selection, both models improve the reference model in the same settings but k-NN model gives better performance than MLP model when both surpass the reference model. This can be explained due to fact that k-NN is a non parametric model and exhibits the ability of modeling any distribution if the noise level is low. The noise effect is played down as the number of neighbors is larger.

3.2 Results with Variable Selection

In this section, we repeat the experiments presented in the previous section using as inputs to the models the variables selected with Mutual Information (Figures 1 and 2). A time window of twelve hours before to the prediction time is considered and the subset of the four variables with highest mutual information is used to feed the two models. Tables 5, 6, 7 and 8 show the

Table 5: Selected variables for the *Setting A*.

	Selected Variables
$\hat{p}(t+1)$	$v(t)$, $v(t-1)$, $v(t-4)$, $v(t-9)$
$\hat{p}(t+2)$	$v(t)$, $v(t-5)$, $v(t-9)$, $v(t-11)$
$\hat{p}(t+3)$	$v(t)$, $v(t-4)$, $v(t-9)$, $v(t-11)$
$\hat{p}(t+4)$	$v(t)$, $v(t-3)$, $v(t-7)$, $v(t-9)$
$\hat{p}(t+5)$	$v(t)$, $v(t-4)$, $v(t-8)$, $v(t-11)$
$\hat{p}(t+6)$	$v(t)$, $v(t-5)$, $v(t-9)$, $v(t-11)$

Table 6: Selected variables for the *Setting B*.

	Selected Variables
$\hat{p}_{vesta}(t+1)$	$v(t)$, $v(t-1)$, $v(t-6)$, $v(t-8)$
$\hat{p}_{vesta}(t+2)$	$v(t)$, $v(t-4)$, $v(t-10)$, $v(t-11)$
$\hat{p}_{vesta}(t+3)$	$v(t)$, $v(t-3)$, $v(t-8)$, $v(t-11)$
$\hat{p}_{vesta}(t+4)$	$v(t$, $v(t-2)$, $v(t-5)$, $v(t-10)$
$\hat{p}_{vesta}(t+5)$	$v(t)$, $v(t-1)$, $v(t-5)$, $v(t-11)$
$\hat{p}_{vesta}(t+6)$	$v(t)$, $v(t-5)$, $v(t-9)$, $v(t-10)$

selected variables for each setting. As expected, in all settings the closest value to the prediction instant ($v(t)$, $p(t)$, $p_{vesta}(t)$) is always selected owing to the inertia of the atmosphere.

For *Setting A*, predicting wind farm power from wind speed (Table 5), it can be observed that selected variables, except the first one, are around $v(t-4)$, $v(t-9)$ and $v(t-11)$. In *Setting B* (Table 6) there is no clear pattern in selected variables and for each horizon the selected variables differ each others. The most clear pattern in selected variables appears in *Setting C* where variables around $p(t-1)$,$p(t-2)$ and $p(t-8)$ are always selected for the different horizons(Table 7). Finally in Table 8 there is no clear pattern in the selected variables.

To assess if variable selection has an effect in the performance of the MLP model we have to compare Table 1 with Table 3. For *Setting A*, it can be observed that the introduction of variable selection has no effect in the performance of the MLP. For the second setting, the introduction of variable selection is not relevant for horizons of 1 and 2 hours and even worse for far away horizons, where the performance of the MLP without variable selection is better for all the topologies. In *Setting C*, selection increases the performance of the MLP (Tables 1 and 3), observing that for horizons from 2 to 6 hours, the improvement of the MLP over reference model is higher with variable selection but at the expense of a high variance. Finally in *Setting D*, the introduction of variable selection does not improve the MLP performance over reference model because a slight decrease in improvement values is observed for high horizons. After this analysis we can consider that variable selection has only a positive effect in *Setting C* where the wind farm

Table 7: Selected variables for the *Setting C*.

	Selected Variables
$\hat{p}(t+1)$	$p(t)$, $p(t-3)$, $p(t-7)$, $p(t-9)$
$\hat{p}(t+2)$	$p(t)$, $p(t-1)$, $p(t-2)$, $p(t-8)$
$\hat{p}(t+3)$	$p(t)$, $p(t-1)$, $p(t-2)$, $p(t-7)$
$\hat{p}(t+4)$	$p(t)$, $p(t-1)$, $p(t-2)$, $p(t-8)$
$\hat{p}(t+5)$	$p(t)$, $p(t-1)$, $p(t-2)$, $p(t-7)$
$\hat{p}(t+6)$	$p(t)$, $p(t-1)$, $p(t-3)$, $p(t-8)$

Table 8: Selected variables for the *Setting D*.

	Selected Variables
$\hat{p}_{vesta}(t+1)$	$p_{vesta}(t)$, $p_{vesta}(t-1)$, $p_{vesta}(t-4)$, $p_{vesta}(t-7)$
$\hat{p}_{vesta}(t+2)$	$p_{vesta}(t)$, $p_{vesta}(t-2)$, $p_{vesta}(t-8)$, $p_{vesta}(t-11)$
$\hat{p}_{vesta}(t+3)$	$p_{vesta}(t)$, $p_{vesta}(t-2)$, $p_{vesta}(t-10)$, $p_{vesta}(t-11)$
$\hat{p}_{vesta}(t+4)$	$p_{vesta}(t)$, $p_{vesta}(t-4)$, $p_{vesta}(t-9)$, $p_{vesta}(t-10)$
$\hat{p}_{vesta}(t+5)$	$p_{vesta}(t)$, $p_{vesta}(t-4)$, $p_{vesta}(t-10)$, $p_{vesta}(t-11)$
$\hat{p}_{vesta}(t+6)$	$p_{vesta}(t)$, $p_{vesta}(t-4)$, $p_{vesta}(t-9)$, $p_{vesta}(t-10)$

power is predicted from previous measures of generated power.

Once the effect of the variable selection has been analyzed for the MLP model, we repeat the same analysis for the k-NN model. In this case two comparisons can be done: one with MLP with variable selection and another with k-NN without variable selection. The comparison of k-NN model and the MLP model, both with variable selection, yields the same results as the comparison of both models without variable selection. k-NN model increase in improvement is higher when both surpass the reference model and the explanation is the same that it was given above (Sec. 3.1).

When comparing k-NN model with and without variable selection is found that in general there is not a remarkable effect of the variable selection. For *Setting A* (Tables 2 and 4) and *Setting B* the improvements of the k-NN model are lower than the reference model and it increases as the horizon and the number of neighbors raise. In *Setting C* and *Setting D*, the effect of the variable selection in the results of the k-NN model is negligible because they are almost the same. On sight of the results it can be thought that k-NN is less affected by the variables because the non parametric nature of the model which adapts better to the underlying distribution of the values than the MLP model.

4 CONCLUSIONS

In this work we have presented a comparative study of the use of variable selection in short-term wind farm power prediction. The variable selection is done with the Mutual Information that is estimated with the method proposed by Kraskov. Two models are considered to study the effect of the variable selection. One is a Multilayer Perceptron with different topologies and the other is the k-NN model for different values of the number of neighbors *k*.

Four different set of experiments were proposed with different input and predicted variables. To assess the quality of the results instead of the RMSE value, the improvement of the RMSE over an improved persistence model is used.

From the obtained results it can be concluded that the k-NN model performs better than MLP model for the different considered horizons and this is more emphasized as the number of neighbors increase. An interesting conclusion is that the wind farm power prediction is better done when power is used as predicting variables 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 MLP and k-NN models yield better performance.

With respect to the introduction of a previous stage of variable selection, in the experiments carried out there is no evidence of a remarkable effect in the results. In the MLP this effect is a bit noticeable when predicting the wind farm power from previous measures of generated power and negligible in the k-NN model.

ACKNOWLEDGEMENTS

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

REFERENCES

Chow, T. W. S. and Huang, D. (2005). Estimating optimal feature subsets using efficient estimation of high-dimensional mutual information. *IEEE Transactions on Neural Networks*, 16(1):213–224.

Costa, A., Crespo, A., Navarro, J., Lizcano, G., Madsen, H., and Feitosa, E. (2008). A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews*, 12(6):1725–1744.

Crone, S. F. and Kourentzes, N. (2010). Feature selection for time series predictiona combined filter and wrapper approach for neural networks. *Neurocomputing*, 73:1923–1936.

Focken, U., Lange, M., Monnich, K., Waldl, H., Beyer, H., and Luig, A. (2002). 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.

Guillén, A., Herrera, L., Rubio, G., Pomares, H., Lendasse, A., and Rojas, I. (2010). New method for instance or prototype selection using mutual information in time seriesprediction. *Neurocomputing*, 73:2030–2038.

Ji, Y., Reyhani, N., and Lendasse, A. (2005). Direct and recursive prediction of time series using mutual information selection. In *Computational Intelligence and Bioinspired Systems (IWANN 2005)*, volume 3512/2005 of *Lecture Notes in Computer Science*, pages 1010–1017.

Kraskov, A., Stgbauer, H., and Grassberger, P. (2004). Estimating mutual information. *PHYSICAL REVIEW E*, 69(6):066138.

Kusiak, A., Zheng, H., and Song, Z. (2009). Wind farm power prediction: A data-mining approach. *Wind Energy*, 12:275–293.

Landberg, L. (2001). Short-term prediction of local wind conditions. *Journal of Wind Engineering and Industrial Aerodynamics*, 89:235245.

Li, S., Wunsch, D. C., Ohair, E. A., and Giesselmann, M. G. (2001). Using neural networks to estimate wind turbine power generation. *IEEE Transactions on Energy Conversion*, 16(3):276–282.

Méndez, J., Lorenzo, J., and Hernández, M. (2009). Experiments and reference models in training neural networks for short-term wind power forecasting in electricity markets. In *Bio-Inspired Systems: Computational and Ambient Intelligence*, volume 5517/2009 of *Lecture Notes in Computer Science*, pages 1288–1295.

MILCA (2010). Mutual information least-dependent component analysis library. http://www.klab.caltech.edu/∼kraskov/MILCA/. Last visited july, 2010.

Mohandes, M. A., Rehman, S., and Halawani, T. O. (1998). A neural networks approach for wind speed prediction. *Renewable Energy*, 13(3):345 – 354.

Nielsen, T. S., Joensen, A., Madsen, H., Landberg, L., and Giebel, G. (1998). A new reference for wind power forecasting. *Wind Energy*, 1(1):29–34.

Sánchez, I. (2006). Short-term prediction of wind energy production. *International Journal of Forecasting*, 22:43–56.

Yoon, H., Yang, K., and Shahabi, C. (2005). Feature subset selection and feature ranking for multivariate time series. *IEEE Transactions on Knowledge and Data Engineering*, 17(9):1186–1198.