Data mining for wind power forecasting

3 authors, including:

George N. Kariniotakis
MINES ParisTech
230 PUBLICATIONS   **5,000** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Storage-Enabled Sustainable Energy for Buildings and Communities (SENSIBLE) View project

Project   Fractal Grid: towards the future smart grid View project

European Wind Energy Conference - Brussels, Belgium, April 2008

Data mining for wind power forecasting

Lionel Fugon , Jérémie Juban and George Kariniotakis

École des Mines de Paris

B.P.207, F-06904 Sophia-Antipolis, France

jeremie.juban@ensmp.fr; georges.kariniotakis@ensmp.fr

Abstract

Short-term forecasting of wind energy production up to 2-3 days ahead is recognized as a major contribution for reliable large-scale wind power integration. Increasing the value of wind generation through the improvement of prediction systems performance is recognised as one of the priorities in wind energy research needs for the coming years. This paper aims to evaluate Data Mining type of models for wind power forecasting. Models that are examined include neural networks, support vector machines, the recently proposed regression trees approach, and others. Evaluation results are presented for several real wind farms.

1   Introduction

Wind power has been undergoing a rapid development in recent years. Several countries have reached already a high level of installed wind power capacity, such as Germany, Spain and, Denmark, while others follow with fast rates of development. Such large-scale integration of wind power is challenging in terms of power system management. Indeed, wind is a variable resource that is difficult to predict. As an example, traditionally, additional reserves are allocated to manage this uncertainty.

This increases the overall cost of the produced energy and limits the benefits of using such a renewable energy resource. A way of reducing the uncertainty associated to wind power production is to use forecasting tools. Development of such tools has been ongoing for more than 15 years [1]. These tools are multi-step ahead forecasting models that provide information for several horizons ahead. In the same way, the continuous improvement of computers and the constant increase in databases capacity permitted the development of a new scientific investigation field called Data Mining. Data Mining has been defined as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data (Fayyad, 1996)". Several algorithmic techniques issued from Data Mining have already been adapted to the wind power forecasting [2] to provide a single expected value for each forecast horizon, called deterministic, spot or point forecast.

The paper initially introduces the different algorithms used. Then the real-world data used to evaluate the models are presented. They are from French wind farms located in different terrain complexity and climatic conditions. Finally an evaluation and comparison of the models performance for each wind farm follows. The paper ends with some conclusions and remarks.

2   Data Mining Models Used

Data Mining encompasses different algorithms from many scientific fields (statistics, artificial intelligence...) for building models (supervised methods) : $Y = \phi(X) + \epsilon$ where Y is the variable to predict or explain and $X = (X_1, X_2, ..., X_n)$ represents the vector of the n-explanatory variables. ϵ is the error of the model. The aim is to approximate the function ϕ and minimize the error. For that purpose, we used both a linear and a non-linear approach.

2.1   Linear Models

Two versions of linear regression models have been considered: one base version, which is used as a simple reference model and a second one that includes interactions. Interactions basically consist in combining the input variables to create extra variables used in the same linear settings. This has the advantage of considering non-linearities while keeping a simple linear setting. However, this can enlarge considerably the dimension of the problem.

The linear model without interaction can be presented by the following formula:

$$Y = X\beta + \epsilon \qquad (1)$$

where $\beta = (\beta_1, \beta_2, ..., \beta_n)$ represents the vector of linear regression coefficients to estimate.

If we take account of different hypothesis on ϵ (such as i.i.d.), we approximate the β-vector by least square minimization on the learning set:

$$\sum_{i=1}^{n} y_i - \beta_0 - \beta_1 x_i^{1} - \beta_2 x_i^{2} - ... - \beta_p x_i^{p}$$

$$= \|y - X\beta\|^2 \quad (2)$$

For the linear regression with interaction, cross products between variables are added to the input variables. Then, the same method is applied on this augmented input set.

2.2 Non-Linear Models

In this part, several non-linear models for wind power forecasting are considered. Such models are better suited to account for the non-linearity of the wind to power conversion. First, neural networks are considered. Then, we consider models based on classification and regression trees with a simple bagging version and a more advanced version called random forest. Finally, we present a last approach based on support vector machines.

2.2.1 Neural Networks

A neural network is an ensemble of neurons (or nodes) connected along several levels called layer inspired by the structure of the human brain. It is generally composed of three layers of neurons, namely an input layer corresponding to explanatory variables, an output layer that provides a response and one or several hidden layers. Neurons of a same layer are never connected between themselves.

Each neuron is affected a weight and passes signals based on a specific transfer function. The most used transfer function is the sigmoid, which receives as input a weighted linear combination of the output of the neurons of the previous layer:

$$s(\theta) = \frac{1}{1 + exp(-\theta)} \qquad (3)$$

with $\theta = \Sigma_i \beta_i X_i$.

The learning step of the neural network is based on a back-propagation algorithm. The principle is to adjust the different neuron weights progressively by back-propagating the error

from the output layer to the input layer. The main advantage of the neural network is its capacity to model complex structures and account for non-linear relations between the explanatory variables and the output. Although, one of the main drawbacks, is that the final performance of a neural network is very sensitive to the design of the network. The choice of the number of hidden layers and neurons contained in each layer is very important. For instance integrating a high number of neurons in the network permits to model more complex relations but can also lead to overfitting of the data and to poor out of sample performances. In this paper, in order to overcome this problem, a neural network architecture optimisation algorithm based on the parameter decay principle has been used (regularisation of the problem). A second drawback is that it is difficult to get insight on the learnt relation simply by looking at the information stored in the network. In this way, neural networks are generally considered as "black box" algorithms.

2.2.2 Random Forests

Bagging for Bootstrap Aggregating, is a method for generating an ensemble of models constructed from samples bootstrap replicates [3]. These replicates are obtained by sampling uniformly with replacement from the original samples. The predictors are then combined by voting for classification or averaging for regression [3]. The main advantage of averages of predictions from several models (like bootstrap sampling) is that it reduces the variance and prediction error.

The base method used in the models hereafter named "Bagging" and "Random Forest" is classification and regression trees (CARTs) [4]. The goal of CARTs is to divide a sample of data using binary rules making the child nodes less heterogeneous than the parent nodes. Once a tree is grown it is possible to extract information from the tree structure, which makes it also a tool for data analysis. The main advantages of CARTs is that it permits to perform a regression or a classification with high dimensional inputs and the major disadvantage of the later is that it is unstable i.e. a small change in the training sample can generate large changes in the learned predictor (classification or regression) [3]. The Bagging algorithm has been used in the specific case of binary tree and Random Forest is a version more sophisticated of Bagging because it adds a random input selection which consists in selecting at random, at each node, a small group
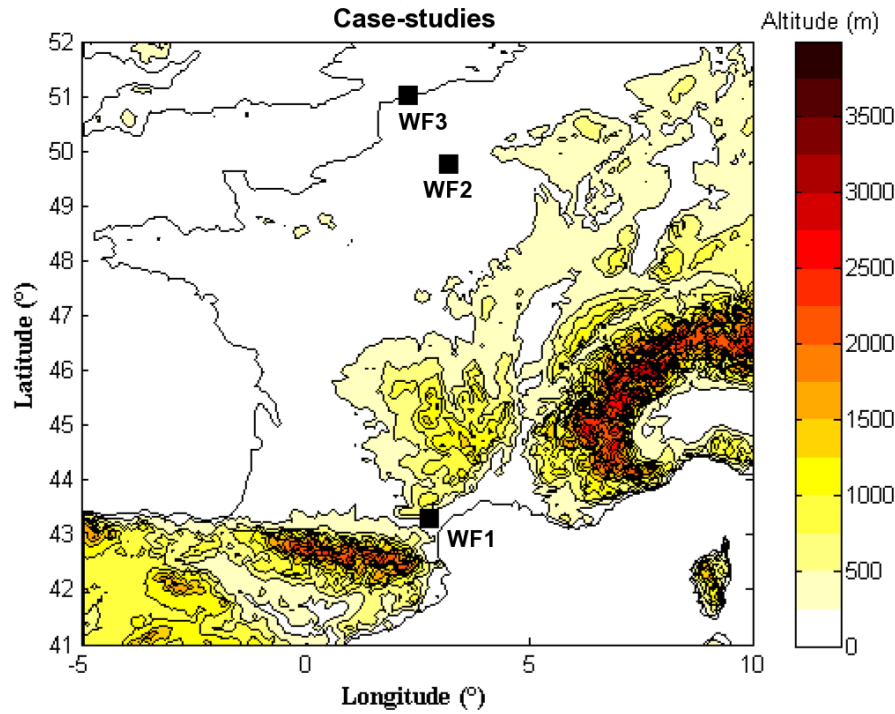
2

Figure 1: Map representing the location of the three considered wind farms WF1, WF2 and WF3.

of input variables to split on. That way, the trees built are more independent.

In the Random Forest approach, the conditional mean $E[Y|X = x]$ is approximated by the averaged prediction of K single trees, each constructed with an i.i.d. vector θ_k, k=1..K, which represents the tree parameters defining how the tree is grown (e.g. split points).

The main drawbacks of these two algorithms are the important computing time of the learning step and trees storage but the advantages are that an insensibility to overfitting thanks to out-of-bag error and few parameters to adjust. Moreover, Random Forest provides information about the frequency of variables appearance in trees, which can be used to determine the importance of each input variable.

2.2.3 Support Vector Machines (SVM)

Support vector machines (SVM) are a recent supervised learning methods used, initially for classification, and generalized later for regression. They are based on Vapnik's research about learning theory. Support vector machines for classification are based on two ideas. First, a principle of maximum margin, which is the distance maximizing the separation frontier and nearest elements called support vector. The learning step is the optimization of this frontier

and can be presented like a quadratic optimization problem. Secondly, the input dimension space is transformed into a higher dimensional space, thanks to a kernel function, where a maximal separating hyperplane is constructed. The goal is to transform a complex (non-linear) low dimension problem into a simple (linear) high dimensional problem.

The SVM can be also used to predict a quantitative variable: "the Support Vector method can also be applied to the case of regression, maintaining all the main features that characterise the maximal margin algorithm: a non-linear function is learned by a linear learning machine in a kernel-induced feature space while the capacity of the system is controlled by a parameter that does not depend on the dimensionality of the space" Cristianini and Shawe-Taylor (2000).

The main advantage of SVM is that the regularisation technique makes the model very resistant to overfitting. The main drawback of SVM is that the computing time required might be very high when compared to other non-linear learning approaches.

3

3 Case study description

Three wind farms in France, denoted as WF1, WF2, and, WF3, are considered. They are representative of various terrain and climate conditions. WF1 is situated on a complex terrain and WF2, WF3 on a flat terrain. Hourly power production time series are considered spanning a period of 18 months from July 2004 to December 2005. For the same period, numerical weather predictions (NWPs) by the ARPEGE model of Meteo France are used. The forecasts are provided once a day for horizons 0 to 60 hours ahead, with a 3-hour resolution, i.e. 20 values for each meteorological variable are provided per run.

The meteorological variables considered in this study are 50 meter above ground level wind speed and gust wind direction. These meteorological variables were found to be the most informative for these case study [5].

The variable to be predicted Y_t is the hourly average power production of each wind farm. The explanatory variable vector (X_t) contains the predicted wind speed and wind direction by the NWP model, the last measured wind power and the forecast horizon. These two last variables permit to improve forecasts for the first forecast horizons. The horizons of power predictions are the same as that of NWPs, which range from 0 to 60 hours ahead, with a 3-hour resolution. The available dataset is divided into a learning-set and a test-set comprising 1 year and 6 months of data respectively. The 1 year learning-set permits to integrate all seasonal variations.

4 Results

The chosen evaluation criteria are the Normalized Mean Absolute Error:

$$NMAE(k) = \frac{\sum_{t=1}^{N} |\varepsilon(t+k|t)|}{N} \qquad (4)$$

and the Normalized Root Mean Square Error:

$$NRMSE(k) = \sqrt{\frac{\sum_{t=1}^{N} (\varepsilon(t+k|t))^2}{N}} \qquad (5)$$

where ε is the normalized prediction error, k is the horizon and N is the number of samples in the testing set. These forecasts are compared to persistence, which is used as base line reference model, and simply consists in using the latest
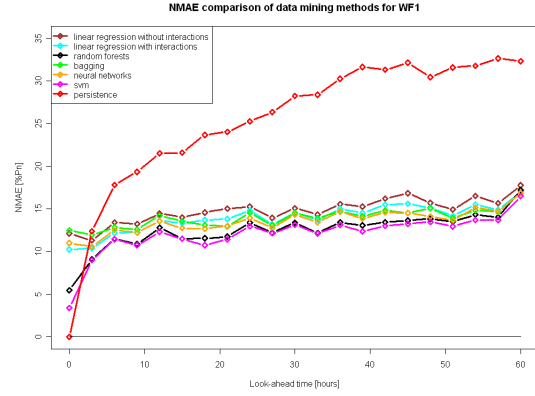

Figure 2: Comparison of NMAE results in the case of the wind farm WF1 situated on a complex terrain
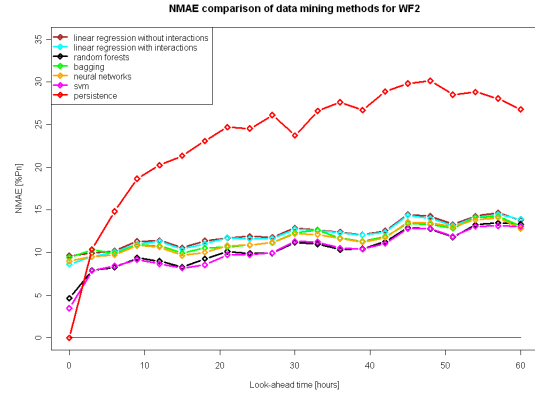

Figure 3: Comparison of NMAE results in the case of the wind farm WF2 situated on a flat terrain
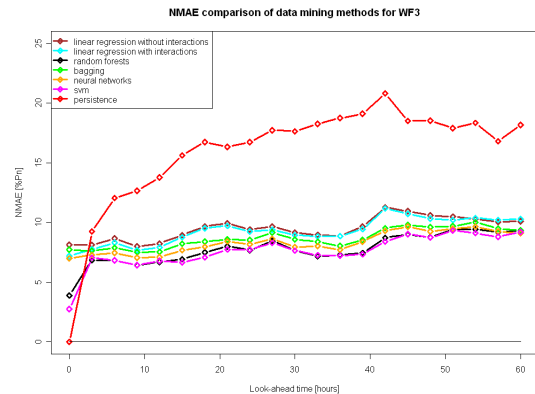

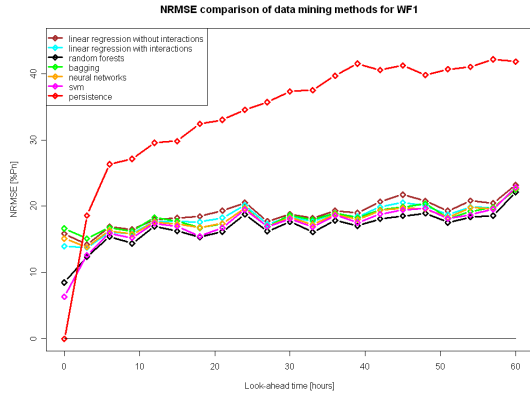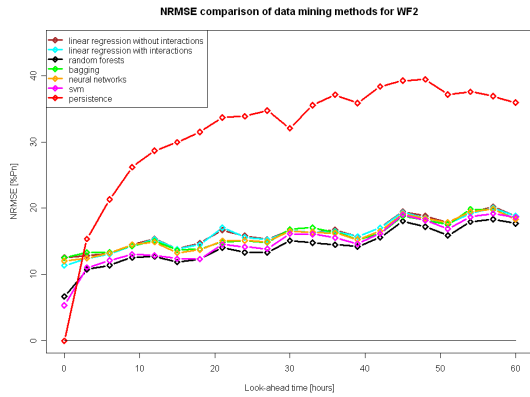Figure 4: Comparison of NMAE results in the case of the wind farm WF3 situated on a flat terrain

4

NRMSE comparison of data mining methods for WF1

Figure 5: Comparison of NRMSE results in the case of the wind farm WF1 situated on a complex terrain

NRMSE comparison of data mining methods for WF2

Figure 6: Comparison of NRMSE results in the case of the wind farm WF2 situated on a flat terrain

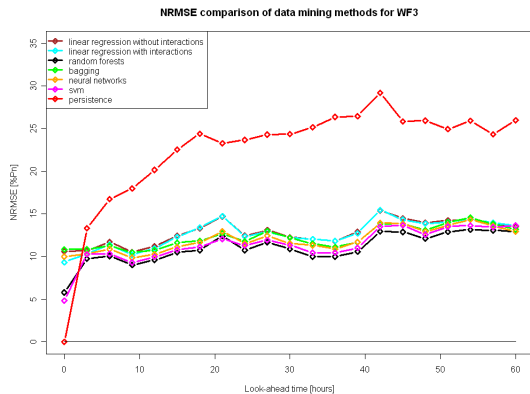**NRMSE comparison of data mining methods for WF3**

Figure 7: Comparison of NRMSE results in the case of the wind farm WF3 situated on a flat terrain

observation as forecast for all horizons. Persistence is commonly used as a benchmark model in wind power forecasting.

The first common conclusion is that all models outperform persistence (excepted for 0-horizon which is a particular case corresponding to nowcasting) and the level of accuracy depends of the terrain. Thus, we can observe a higher error level for WF1, which is situated on a complex terrain, compared to WF2 and WF3, which are on flat terrains. Then, for the three wind farms, we observe a global superiority of the non-linear models over the linear ones. This could be easily explained by the non-linear relationship between wind and power.

Even if the non-linear models presented here permit to improve the global performance when compared to linear ones, it should be noticed that the performances of the simple linear approach is still reasonably good when compared to the persistence reference model. It should also be noticed that the results presented here are comparable to results found in the literature for wind farms located in similar terrains.

5 Conclusions

This paper presents a comparison of the performance of various data mining algorithms applied to short-term wind power forecasting. The interest of non-linear methods is illustrated for which the performance is equivalent to that found in the literature for wind farms located in similar terrains. The comparison has revealed that Random Forest outperforms the rest of the models. This model, originally applied here for wind power forecasting, is interesting since it does not require a long architecture optimisation step, only the number of trees in the forest has to be optimised.

Moreover, a generalization of Random Forests, Quantile Regression Forests give a non-parametric way of estimating conditional quantiles for high-dimensional predictor variables [6, 5]. Thus provides an additional information on the uncertainty of the predictions for performing efficiently functions such as reserves estimation, unit commitment, trading in electricity markets, a.o. Such prediction (deterministic with prediction intervals) is named probabilistic forecasting and are, nowadays, an important research field.

5

6 Acknowledgements

References

[1] G. & Brownsword R. Giebel, G.; Kariniotakis. The state-of-the-art in short-term prediction of wind power - from a danish perspective, 2003.

[2] S. Santoso, M. Negnevitsky, and N. Hatziargyriou. Data mining and analysis techniques in wind power system applications: abridged. *Power Engineering Society General Meeting, 2006. IEEE*, pages 3 pp.–, 18-22 June 2006.

[3] Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123140, August 1996.

[4] Charles J. Stone Leo Breiman, Jerome Friedman and R.A. Olshen. *Classification and Regression Trees*. Chapman & Hall/CRC, 1984.

[5] Jeremie Juban, Lionel Fugon, and George Kariniotakis. Probabilistic short-term wind power forecasting based on kernel density estimators. In *Proceedings of the European Wind Energy Conference*, Milan, Italy, 7-10 May 2007.

[6] Nicolai Meinshausen. Quantile regression forests. *Journal of Machine Learning Research*, 7:983999, June 2006.