# Statistical learning for wind power: a modeling and stability study towards forecasting

Aurélie Fischer, Lucie Montuelle,
Mathilde Mougeot & Dominique Picard

Laboratoire de Probabilités et Modèles Aléatoires,
Université Paris Diderot, 75013 Paris, France

We focus on wind power modeling using machine learning techniques. We show on real data provided by the wind energy company Maïa Eolis, that parametric models, even following closely the physical equation relating wind production to wind speed are outperformed by intelligent learning algorithms. In particular, the CART-Bagging algorithm gives very stable and promising results. Besides, as a step towards forecast, we quantify the impact of using deteriorated wind measures on the performances. We show also on this application that the default methodology to select a subset of predictors provided in the standard random forest package can be refined, especially when there exists among the predictors one variable which has a major impact.

*Keywords: wind power; data mining; random forests; bagging; modeling; stability; forecasting.*

## 1 Introduction

In France, wind energy represents today 3.9% of the national electricity production. The United Nations Conference on Climate Change COP21 has set a goal of 30% renewable energy in the overall energy supply in the country by 2020, and more precisely, the French wind production should double by 2020 [19].

Since electricity can hardly be stored, forecasting tools are essential to appropriately balance the production of the different renewable energies. Today, in France, wind energy is produced by more than 1400 wind farms scattered all over the country. The production of each wind farm is highly dependent of the meteorological conditions and especially of the wind. It is well known that the behavior of the wind is very different from one region to another, and this seems especially significant in France, where several quite different climates are present despite the relatively small area of the country [19]. So, to be accurate, the global wind electricity forecast should rely on local models, dedicated to each wind farm. Consequently, an important first step is to quantify the modeling performances of wind production in the different French regions, using real operational data.

Two kinds of framework are usually investigated today for wind power prediction. On the one side, physical models rely on the modeling of each wind turbine based on equations [5]. On another side, a trend of new mathematical tools tends to model the power production by learning the phenomenon directly on the data, without integrating any knowledge on the physical behavior of the wind turbines. Such techniques using statistical models and data mining methods have been investigated in many complex situations, for instance considering short term prediction. Among others, parametric regression models, Support Vector Machines for regression, regression trees, random forests, neural networks have been considered. For instance, the use of neural networks has been investigated in [18, 12] and in [17]. A special network, called extreme learning machine, has been used in [21] for probabilistic interval forecasting. In [15], the $k$-nearest neighbor algorithm is used for probabilistic forecasts in the frame of the Global Energy Forecasting Competition 2014. Support vector machines for regression have been proposed in this context in [11], whereas [13] provides a comparison between several data-mining approaches. Besides, time series-based models have also contributed to the field of wind power forecast (see, e.g., [16, 22]). For an overview of different modeling and forecasting methods for wind power, the reader may further refer to the surveys [5, 8, 9, 10].

In the present paper, adopting the second point of view, we investigate and compare different techniques for modeling the electrical power for several wind farms in France. For each farm, we first model the electrical power of each wind turbine of the farm using local inputs coming from sensors directly installed on each wind turbine. The predictive power of the farm is then given by the sum of the predictive powers computed for each wind turbine. In a second step, we quantify the modeling performances by using more global inputs as may be provided by a meteorologist forecaster as for example, Météo France. This approach helps to quantify the performance of the different models running in an operational environment, using only average input information at a farm scale.

The CART-Bagging algorithm appears to perform the best on our data and gives very satisfactory predictions.

The paper is organized as follows. In Section 2, we thoroughly describe the data set at hand. Section 3 introduces the different methods investigated in our study. Section 4 presents and discusses the modeling performances obtained using the local information on each turbine. The results found when replacing this information by the more global one, relying on averages, are given in Section 5.

## 2 Data set

The data set has been provided by the wind energy company Maïa Eolis. In a farm, each wind turbine provides 10 minute measurements of electrical power, wind speed, wind direction, temperature, as well as an indicator of the working state of the turbine. The electrical power output of the whole farm is also provided on a 10 minute basis. All measures are recorded simultaneously. Data is available for 3 different farms made up of 4 to 6 turbines, in the North and East of France, from 2011 to 2014.

To detect freeze, wind speed is measured on each turbine both by a classical anemometer and a heated one. Since more measures are available from the heated anemometer, the study has been conducted with this data. Wind direction is provided by a weather vane and has been recoded into two variables corresponding to the cosine and the sine of the angle. The state of the turbine may correspond to start, stop or full working of the turbine, depending on the wind speed and maintenance operations. For the sake of simplicity, this study focuses on fully operating times. Besides, the data has been averaged over 30 minutes in order to slightly smooth the signals. However, it should be stressed that most often the results obtained on a 10 minute basis are quite similar to those presented in the sequel.

Taking advantage of the 30 minutes averages, two new variables have been introduced: the variance of the wind speed, and the variance of the wind direction over 30 minutes. The second variable (complex-valued), has been decomposed into its real part and its imaginary part, leading to a total of 7 explanatory variables.

## 3 Predictive methods

In this section, all the measures are assumed to be observed in real time. Based on this data, our aim is to *model* the farm power. More precisely, the variables are observed at time $t$ and the sum of the power of each turbine of the farm at time $t$ is predicted. We recall that the studied model is applied at each turbine, providing an estimate of its power. Then the estimated farm power is computed by summing the estimated turbine powers. The error is calculated at the farm scale.

Our intention is to compare parametric statistical methods closely reflecting the related physical equation, to more elaborated techniques inspired from machine learning. These methods are especially designed to learn a phenomenon in a completely agnostic way and may be suitable for high dimensional data or complex data. In particular, they can easily accommodate non-linear modeling as well as dependence between variables, which is the case here.

## 3.1 Theoretical equation

According to theoretical studies on wind turbines (see, e.g., [14]), the delivered power obeys the following equation :

$$P(W) = \frac{1}{2}\rho S c_p W^3, \tag{1}$$

where $W$ is the wind speed, $\rho$ the air density, $S$ the rotor surface, which is the area swept by the blades, and $c_p$ the power coefficient, corresponding to the fraction of wind energy that the wind turbine is able to extract. Thus, as expected, the power significantly depends on the wind speed and a good approximation of the power curve could lead to good predictions using wind speed measurements. Figure 1 shows the raw observations and the fit to the theoretical curves for a wind turbine. Figure 1a plots power versus wind-speed, whereas Figure 1b plots power versus the cube of the wind speed. The two plotted theoretical curves correspond to two different values of $c_p$: the maximal theoretical value (16/27, red curve), and a more realistic value given in Table 8 of [4] (blue curve). The third curve (in green) is provided by the turbine builder, based on his experiments.

Notice that the cloud of observations is quite dispersed and we can already anticipate difficulties for prediction.

In particular, it should be underlined that the parameters of the physical equation (1) are in practice difficult to guess, so that the theoretical curves may not fit very well. Furthermore, to better reflect the observations, the theoretical formula is often used only for a range of wind speeds, outside which the power is assumed to be constant. However, the knowledge of this range requires the estimation of both endpoints of the interval.

Although these curves correspond to some trend, there is obviously room for improvement to produce a better prediction.

## 3.2 Parametric methods

Several methods have been tested to approximate the power curve and model the production. In this section, we present the parametric statistical methods, directly inspired from the physical equation.

**Parametric modeling according to the wind speed only** We first investigated the simplest parametric models, namely linear regression and logistic regression, with the wind speed as unique explanatory variable. If the predicted power at time $t$ is denoted by $\hat{Y}_t$, these models are given by

$$\hat{Y}_t = a_0 + a_1 W_t,$$

and

$$\hat{Y}_t = \frac{C}{1 + \exp(a_0 + a_1 W_t)},$$

where the parameters $a_0$, $a_1$, $C$ are estimated using the associated methodology.

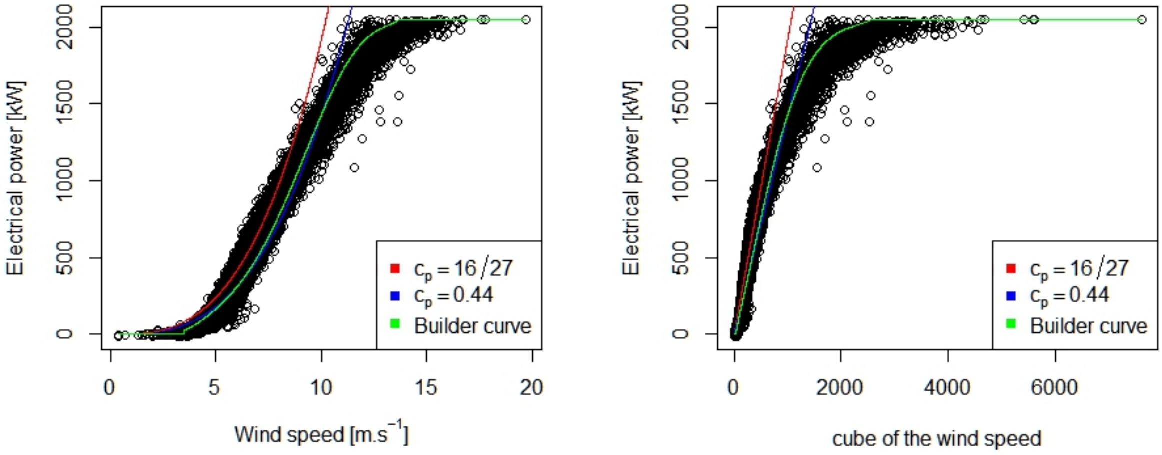


Figure 1: Empirical observations for a wind turbine and theoretical power curves for different power coefficient values, compared to the curve provided by the turbine builder.

Introducing a third degree polynomial of the wind speed in the logistic regression has also been considered to mimic more closely Equation (1). More precisely, the model is then defined by:

$$\hat{Y}_t = \frac{C}{1 + \exp(a_0 + a_1 W_t + a_2 W_t^2 + a_3 W_t^3)},$$

where $a_i$, $i = 0, \ldots, 3$ and $C$ are estimated parameters. This model is denominated in the sequel as polynomial logistic regression.

**Parametric modeling using more variables** Linear regression, logistic regression and polynomial logistic regression with more variables, using not only wind speed as a predictor, but also wind direction, (coded by its cosine and sine : $D^{\cos}$ and $D^{\sin}$), temperature $T$ and the variances of the wind speed $W^S$ and direction, $D^{S,\mathfrak{Re}}$ and $D^{S,\mathfrak{Im}}$, have also been studied.

The Lasso method, which simultaneously performs variable selection and regularization through the least squares criterion penalized by the $\ell^1$ norm of the regression coefficients has been investigated as well (see for instance [20]). The model is defined by

$$\hat{Y}_t = a_0 + a_1 W_t + a_2 D_t^{\cos} + a_3 D_t^{\sin} + a_4 T_t + a_5 W_t^S + a_6 D_t^{S,\mathfrak{Re}} + a_7 D_t^{S,\mathfrak{Im}},$$

with $a_0, \ldots, a_7$ minimizing

$$\frac{1}{n} \sum_{i=1}^{n} \left( Y_i - a_0 - a_1 W_i - a_2 D_i^{\cos} - a_3 D_i^{\sin} - a_4 T_i - a_5 W_i^S - a_6 D_i^{S,\mathfrak{Re}} - a_7 D_i^{S,\mathfrak{Im}} \right)^2 + \lambda \sum_{j=1}^{7} |a_j|.$$

### 3.3 Non-parametric methods and machine learning algorithms

It is well-known that non-parametric and non-linear methods are very useful to model complex phenomena. The following algorithms do not generally lead to closed formulas as in the previous section. We will describe them briefly and refer to the literature for more details.

**SVM for regression** The SVM method for regression maps the inputs into a non-linear feature space, using a kernel representation, for example a Gaussian kernel (see for instance [6]). A non-linear regression function is computed by minimizing the sum of the losses on the points giving rise to an error exceeding some threshold. The threshold parameter is here calibrated using a grid.

**KNN** The $k$-nearest-neighbour procedure consists in computing the average power corresponding to the $k$ nearest neighbours in the feature space (see for instance [7]). As for the previous method, the number $k$ of neighbours is optimized on a grid.

**CART, Bagging and RF** Tree-based methods like CART [3] and Random Forests [2] are also applied. CART grows a binary tree by choosing the cut minimizing the intra-node variance, over all variables and corresponding thresholds. To avoid over-fitting, the tree is usually pruned. The prediction is provided by the value associated to the leaf in which the observation falls.

Another way to reduce variance and avoid over-fitting is to use Bagging [1]. Bagging consists in generating bootstrap samples, fitting a method on every sample (here growing a full tree by CART) and averaging the predictions.

To produce more diversity in the trees to be averaged, an additional random step may be introduced in the previous procedure, leading to Random Forests. In the Random Forests procedure, each tree is grown following the same principle as in CART (with no pruning), but, here, the best cut is chosen among a much smaller subset of randomly chosen variables. The predicted value is the mean of the predictions of the trees.

In the next section, all the experiments have been conducted using the R software. The previous procedures are implemented respectively in the packages `lars`, `kernlab`, `FNN`, `rpart` and `randomForest`. For the Random Forests, the default parameters, advocated by Breiman, were used: 500 trees were grown in each forest and the size of the subset of randomly chosen variables, commonly denoted by *mtry*, is the floor of the third of the number of variables. Note that the CART-Bagging algorithm is a particular case of Random Forests where *mtry* equals the total number of variables.

### 3.4 The naive method

Finally, the so-called "persistence method" uses the last observation as prediction: if $Y_t$ denotes the electric production at time $t$, the predicted production at time $t$ is given by $\hat{Y}_t = Y_{t-1}$. It is interesting to introduce this very naive method as a benchmark in comparison to more sophisticated methods to precisely quantify their gain.

### 3.5 From turbine to farm modeling

As mentioned, the evaluation of the performances is made at the farm scale. Therefore, each turbine is modeled using the evaluated method, then the estimation of each wind turbine power is provided on test points. Finally, the estimated power of the farm is computed by summing theses estimations. More precisely, if the farm comprises six turbines and the linear regression is considered, six linear regression models are adjusted, then predictions for the test set are computed on each turbine: $\hat{Y}_{t,1}, \hat{Y}_{t,2}, \hat{Y}_{t,3}, \hat{Y}_{t,4}, \hat{Y}_{t,5}, \hat{Y}_{t,6}$, and finally the estimation of the farm power is given by $\hat{Y}_t = \sum_{i=1}^{6} \hat{Y}_{t,i}$.

## 4 Modeling performance results

As we are interested in evaluating the predictive power of each method, the data set is split as usual into a training and a test set. In order to quantify the variability of the predictive ability, several test sets are used. An average performance, as well as a standard deviation, are then computed.

More precisely, the procedures are trained on around 8000 instant-points and 10 data sets of 724 points are used to evaluate the performances. The error criterion is the Root Mean Squared Error (RMSE), defined between a vector of predictions $\hat{Y}$ and a vector of observed wind power productions $Y$ by

$$RMSE(\hat{Y}) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2}.$$

A quantity which is also of interest for industries is the error in term of percentage of the installed power (% of IP in the results tables), defined by the average RMSE divided by the theoretical power of the farm. For example, if the farm is composed of 6 turbines of theoretical power 2.05 GW (specified by the turbines builder), the error in term of percentage of the installed power is

$$\text{\% of IP } = 100 \times \frac{mean(RMSE)}{6 \times 2050}\%.$$

This quantity sometimes appears under the denomination Normalized RMSE.

Let us comment the main conclusions drawn thanks to Table 1 and Figure 2.

**General observations** As can be observed in Table 1, the learning algorithms have been investigated either using the wind speed variable only, in which case the emphasis is on the non-linear added value of the method, or using all variables, insisting then on both the non-linear and regularization aspects.

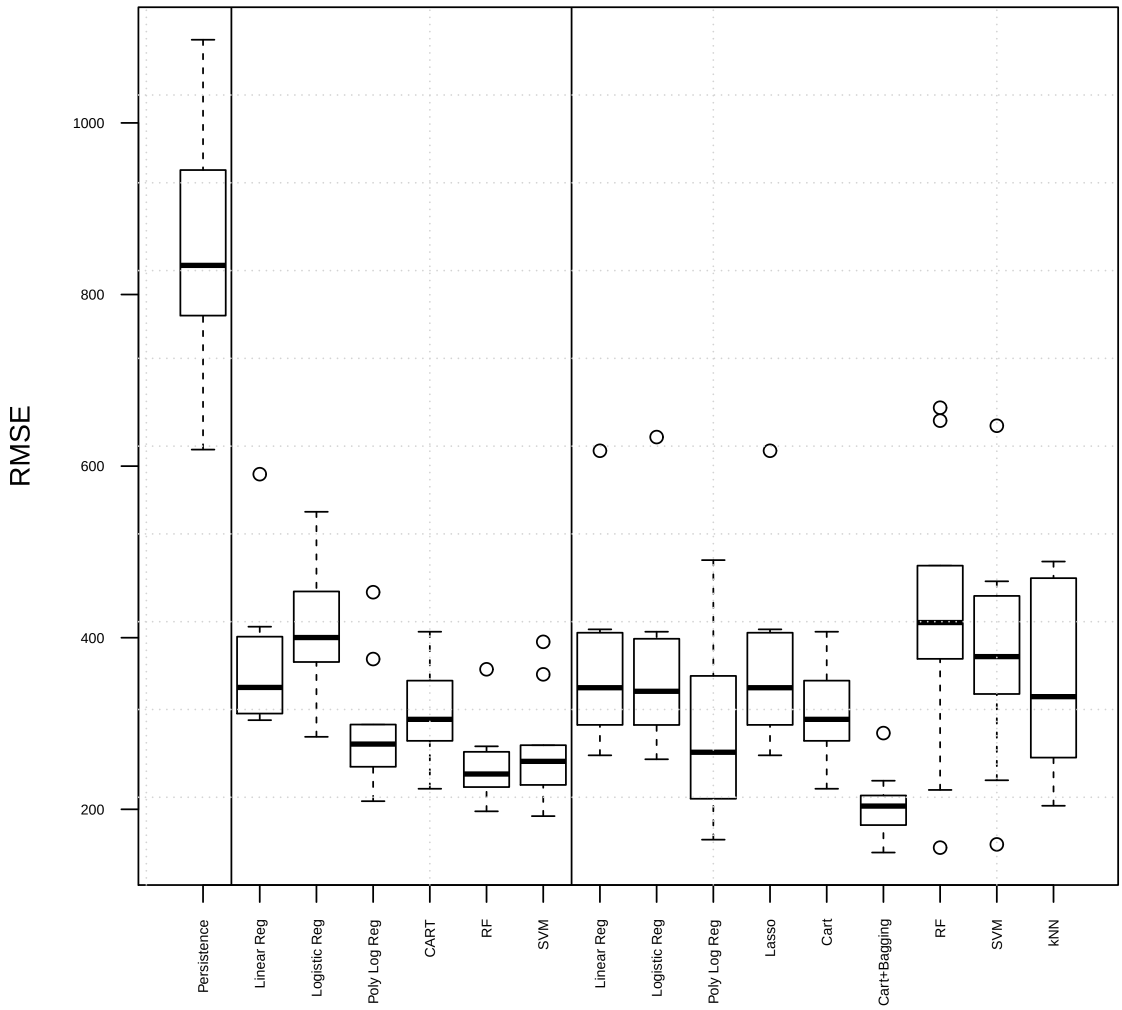


Figure 2: Boxplots of the RMSE for the different procedures using local measures for one farm.

Table 1: Modeling performances using local measures for one farm (IP= Installed Power).

| | Method | Mean of RMSE | Sd of RMSE | % of IP |
|---|---|---|---|---|
| | Persistence | 855.52 | 141.14 | 6.96 |
| using wind speed only | Linear Regression | 373.61 | 86.91 | 3.04 |
| | Logistic Regression | 404.86 | 76.74 | 3.29 |
| | Polynomial Log. Reg. | 290.36 | 73.87 | 2.36 |
| | CART | 314.46 | 57.74 | 2.56 |
| | **CART-Bagging (=RF)** | **250.52** | **46.52** | **2.04** |
| | SVM for regression | 269.94 | 64.21 | 2.19 |
| using all variables | Linear Regression | 364.21 | 102.39 | 2.96 |
| | Logistic Regression | 362.76 | 107.58 | 2.95 |
| | Polynomial Log. Reg. | 292.57 | 100.53 | 2.38 |
| | LASSO | 364.21 | 102.39 | 2.96 |
| | CART | 314.46 | 57.74 | 2.56 |
| | **CART-Bagging** | **203.50** | **39.72** | **1.65** |
| | RF | 425.78 | 161.53 | 3.46 |
| | SVM for regression | 382.16 | 134.34 | 3.11 |
| | kNN (k=2) | 355.29 | 109.96 | 2.89 |

We first observe that all the methods investigated show a much better performance than the naive persistence method, substantially reducing the mean error, with a standard deviation almost always better.

**Wind speed only** Concerning the methods using only the wind speed as predictor, their performances are pretty good, more than twice better than persistence.

The polynomial logistic regression shows a very good performance, which was expected since this model is directly inspired from the physical equations as illustrated in Figure 1. However, the variability of the prediction is a bit high.

The SVM and RF methods for regression show the best results with the best stability. Note that in that particular case, RF and the CART-Bagging procedure coincide.

**All variables** Regarding the parametric methods, the results show that adding more variables, namely the wind direction, the variances of the wind speed and direction, and the temperature, do not lead to any substantial improvement. Among these procedures, polynomial logistic regression shows the best performances.

The LASSO procedure is not very promising. This is probably due to multiple factors : the method uses the predictors in a linear way – compared to SVM or CART, which are bringing different kinds of non-linearity – and the predictors are highly correlated. We observe also that the results are the same for LASSO and the classical linear regression due to the fact that no selection has been in fact performed by the method.

The CART algorithm does not take advantage of the additional variables and seem to choose its cuts only according to the wind speed. This may be explained by the prevailing importance of the wind speed over other measures.

Among the agnostic machine learning algorithms, the SVM shows one of the poorest performances. It should be noted that several tested kernels were not able to compete with the polynomial logistic regression for example.

The KNN method has a performance similar to the SVM procedure.

The CART-Bagging algorithm outperforms all the investigated statistical models. The case of Random Forests is quite interesting and has to be discussed separately. Looking at the Table 1 and Figure 2, we can observe that RF surprisingly seem less efficient than other methods and especially CART when dealing with all variables. However this poor result has to be refined.

As explained above, the RF algorithm, instead of considering all the variables to grow a tree (as CART does), operates a random selection among these variables. The default choice for this random selection is the uniform distribution to choose a subset of the original variables, of size *mtry*, the floor of the third of the number of predictors. In the CART-Bagging procedure, all the variables are selected. In our data set, obviously, the importance of the wind speed prevails over all other variables: for instance, CART performs nearly all its cuts according to the wind speed. Therefore, if the wind speed variable is often not selected in a random sample, the resulting cut is often not appropriate. Choosing more variables increases the probability to select a specific variable, namely, here, the wind speed. Very different performances are then observed between RF with the default parameter for *mtry* and the CART-Bagging method, corresponding to RF with *mtry* equal to the number of predictors.

Comparing CART and CART-Bagging highlights the advantages of bootstrapping and averaging. This step allows to reduce the error by a third, when dealing with all the predictors.

Note that, according to the renewable energy union [19], French industries obtain a root mean squared error of 2.4 % of the installed power of farm productions, which illustrates the benefits of using CART-Bagging (1.65 %).

**Comparison of different farms** The results given in the previous paragraphs concern a farm in the East of France. Data from two different sites in the North of France were also available. For every farm, the hierarchy between procedures is quite similar, the procedure ranking first most often is CART-Bagging.

To make a fair comparison between the farms, a new experiment has been conducted. A common test set, with observed variables available at the same time for each farm, with at least one turbine fully operational, has been drawn. The test set has been divided into ten subsets of 1440 instant-points, each covering a period of around thirty days, to quantify the average performance and its variability. The training set consists in around 7200 instant-points, satisfying a ratio of 83% of the data dedicated to learning and 17% used for test.

Only the best procedure, CART-Bagging, has been applied. We also compare the results with the turbine builder's power curve, used on each turbine to model the farm. Figure 3 highlights the good results of CART-Bagging on the first and the third farms.

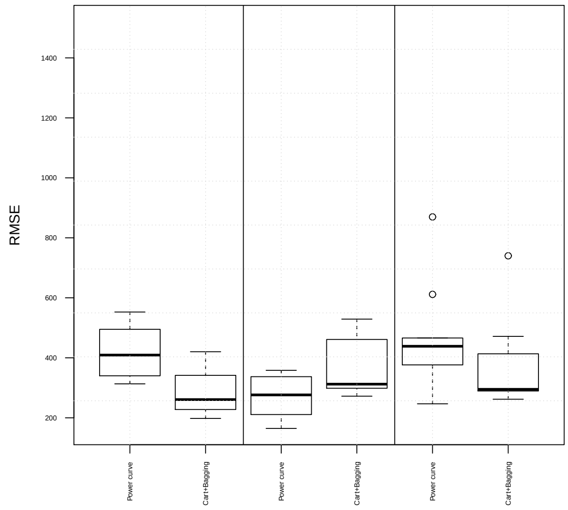


Figure 3: Comparison of the RMSE for the turbine builder's power curve and the CART-Bagging procedure on several farms using local measures.

It performs reasonably well on the second farm, but is not as good as the power curve's builder. It may be explained by the difference between the wind speed in the training sample and in the test set. Few high wind speed levels are observed in the training sample on the second farm compared to the test sample, so the CART-Bagging prediction may not be accurate.

# 5 Towards forecast : a stability investigation

On a daily use, many observations are recorded in real time on each wind turbine. For example, as already mentioned, each turbine has its own anemometer and vane, which provide very localized information about wind speed and wind direction. In the previous section, we have shown that using this kind of observations, accurate models can be proposed. In a forecast framework, however, this local information is not available. More specifically, the French Weather Agency Météo France can provide forecast of wind and temperature based on numerical simulations. The finest grid resolution is brought by the AROME model, which proposes a resolution of about 1.5 km. It should be noted that, in general, two wind turbines are at a distance of about 300 m from each other. Consequently, an interesting question is to quantify the predictive power not using localized information, but information on a much broader scale.

To mimic Météo France data, which are in the frame of this project not available, virtual sensors have been introduced. For each variable, a global information is computed by averaging all the localized variables coming from the set of turbines installed on the wind farm. This kind of data is a first step to forecast and helps to quantify the

Table 2: Modeling performances using deteriorated wind measures (average).

| | Method | Mean of RMSE | Sd of RMSE | % of IP |
|---|---|---|---|---|
| | Persistence | 855.52 | 141.14 | 6.96 |
| using wind speed only | Linear Regression | 393.09 | 77.25 | 3.20 |
| | Logistic Regression | 541.37 | 103.15 | 4.40 |
| | **Polynomial Log. Reg.** | **288.28** | **75.23** | **2.34** |
| | CART | 349.17 | 53.20 | 2.84 |
| | CART+Bagging (=RF) | 293.26 | 48.96 | 2.38 |
| using all variables | Linear Regression | 387.71 | 89.73 | 3.15 |
| | Logistic Regression | 524.30 | 92.58 | 4.26 |
| | Polynomial Log. Reg. | 297.16 | 92.79 | 2.42 |
| | LASSO | 387.44 | 89.86 | 3.15 |
| | CART | 349.17 | 53.20 | 2.84 |
| | **CART + Bagging** | **228.75** | **43.35** | **1.86** |
| | RF | 447.77 | 161.84 | 3.64 |
| | SVM | 424.15 | 143.02 | 3.45 |
| | kNN | 428.05 | 125.84 | 3.48 |

loss of accuracy due to the replacement of all the localized data with a unique global information.

The same methods have been used and the results are available in Table 2. The deterioration of the prediction can easily be seen in Figure 4. Polynomial logistic regression is remarkably robust, performing similarly to the context with local measures, contrary to SVM and kNN. When only wind speed is considered, polynomial logistic regression competes with CART-Bagging, whereas the latter outperforms all the considered procedures when dealing with all the variables.

**Comparison of different farms** Just as in the previous framework, CART-Bagging and the turbine builder's power curve prediction have been tested on several farms. Figure 5 stresses the good results of CART-Bagging, which seems robust to the difference between the mean wind speed and the local wind speed on each turbine, contrary to the use of the power curve, suffering from the aggregation of sensor data.

# 6 Conclusion

As can be seen in this study, the behavior of a polynomial logistic regression with only the wind speed as covariate is at the same time simple and effective for power prediction.

It is interesting to notice that agnostic methods are in this case very appropriate for prediction and show promising results with the best stability, particularly for CART-Bagging.

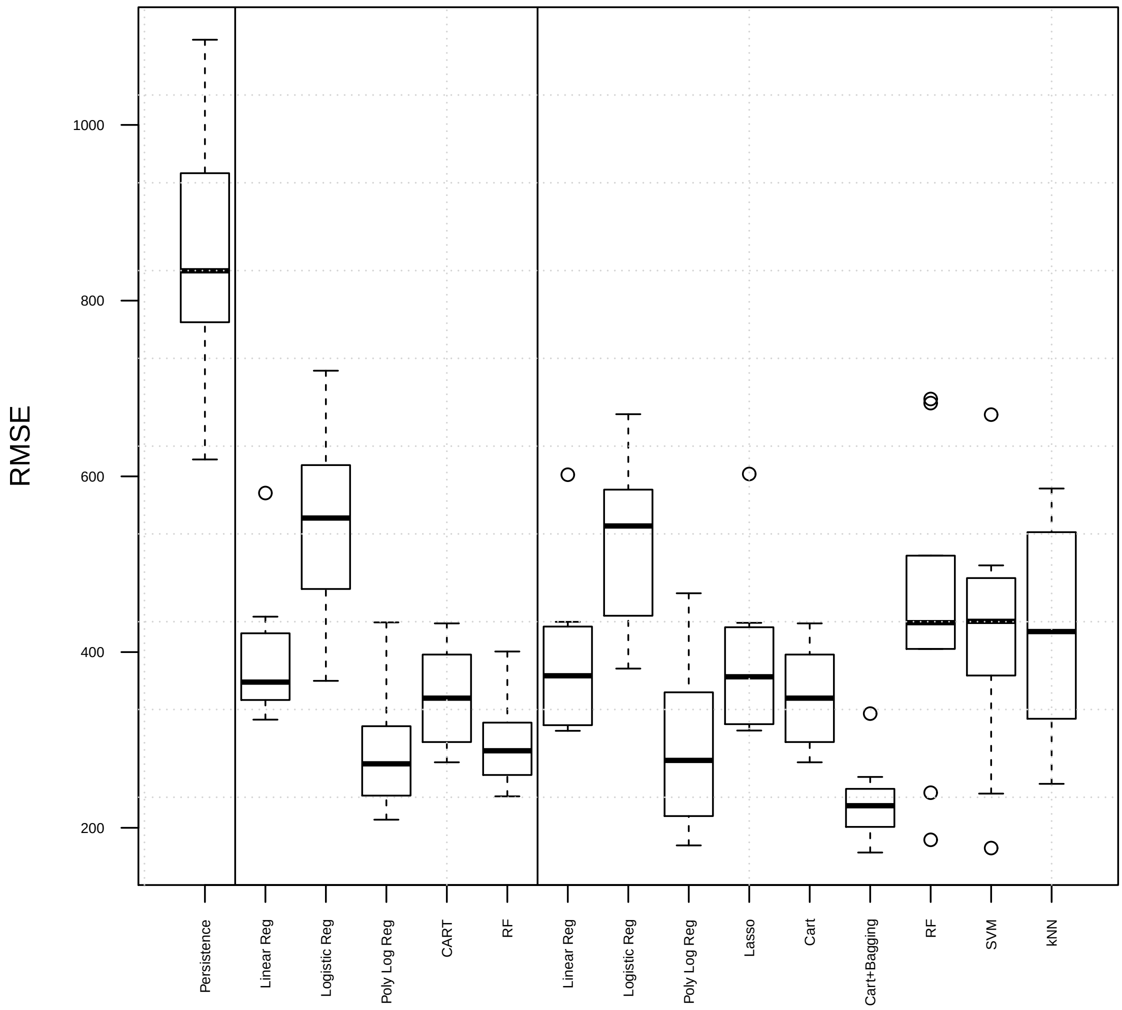


Figure 4: Boxplots of the RMSE for the different procedures using the mean of wind speed measures.

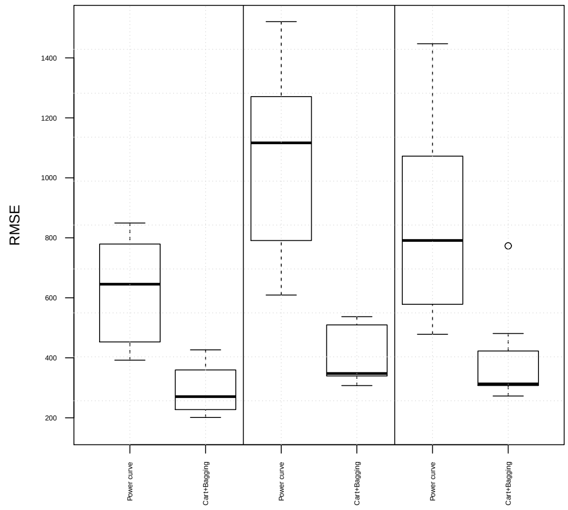


Figure 5: Comparison of the RMSE for the turbine builder's power curve and the CART-Bagging procedure on several farms using local measures.

# 7 Acknowlegement

We are extremely grateful to Nicolas Girard and Sophie Guignard from the Maïa Eolis Company, for providing the data and for helpful discussions.
This research has been supported by a public grant overseen by the French National Research Agency (ANR) as part of the project FOREWER (reference: ANR-14-CE05-0028).

# References

[1] L. Breiman. Bagging predictors. In *Machine Learning*, pages 123–140, 1996.

[2] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, October 2001.

[3] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and regression trees*. Wadsworth Statistics/Probability Series. Wadsworth Advanced Books and Software, Belmont, CA, 1984.

[4] C. Carrillo, A. F. Montaño Obando, J. Cidrás, and E. Díaz-Dorado. Review of power curve modelling for wind turbines. *Renewable and Sustainable Energy Reviews*, 21:572 – 581, 2013.

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

[6] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V Vapnik. Support vector regression machines. In M. C. Mozer, M. I. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems 9*, pages 155–161. MIT Press, 1997.

[7] E. Fix and J. L. Hodges. Discriminatory analysis, nonparametric discrimination: Consistency properties. *US Air Force School of Aviation Medicine*, Technical Report 4(3):477+, January 1951.

[8] A. M. Foley, P. G. Leahy, A. Marvuglia, and E. J. McKeogh. Current methods and advances in forecasting of wind power generation. *Renewable Energy*, 37(1):1–8, 2012.

[9] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl. The state-of-the-art in short-term prediction of wind power: A literature overview. Technical report, ANEMOS. plus, 2011.

[10] J. Jung and R. P. Broadwater. Current status and future advances for wind speed and power forecasting. *Renewable and Sustainable Energy Reviews*, 31:762–777, 2014.

[11] O. Kramer and F. Gieseke. Short-term wind energy forecasting using support vector regression. In *Soft Computing Models in Industrial and Environmental Applications, 6th International Conference SOCO 2011*, pages 271–280. Springer, 2011.

[12] O. Kramer, F. Gieseke, and B. Satzger. Wind energy prediction and monitoring with neural computation. *Neurocomputing*, 109:84–93, 2013.

[13] A. Kusiak, H. Zheng, and Z. Song. Wind farm power prediction: a data-mining approach. *Wind Energy*, 12(3):275–293, 2009.

[14] M. Lydia, S. S. Kumar, A. I. Selvakumar, and G. E. P. Kumar. A comprehensive review on wind turbine power curve modeling techniques. *Renewable and Sustainable Energy Reviews*, 30:452 – 460, 2014.

[15] E. Mangalova and O. Shesterneva. K-nearest neighbors for {GEFCom2014} probabilistic wind power forecasting. *International Journal of Forecasting*, pages –, 2016.

[16] M. Milligan, M. Schwartz, and Y. H. Wan. Statistical wind power forecasting for us wind farms. In *the 17th Conference on Probability and Statistics in the Atmospheric Sciences/2004 American Meteorological Society Annual Meeting Seattle, Washington, January11-15*, 2004.

[17] H. Quan, D. Srinivasan, and A. Khosravi. Short-term load and wind power forecasting using neural network-based prediction intervals. *IEEE transactions on neural networks and learning systems*, 25(2):303–315, 2014.

[18] G. Sideratos and N. D. Hatziargyriou. Probabilistic wind power forecasting using radial basis function neural networks. *IEEE Transactions on Power Systems*, 27(4):1788–1796, 2012.

[19] Syndicat des Énergies Renouvelables. Panorama de l'électricité renouvelable en 2015. Technical report, Syndicat des Énergies Renouvelables, 2015.

[20] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[21] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong. Probabilistic forecasting of wind power generation using extreme learning machine. *IEEE Transactions on Power Systems*, 29(3):1033–1044, 2014.

[22] B. Wu, M. Song, K. Chen, Z. He, and X. Zhang. Wind power prediction system for wind farm based on auto regressive statistical model and physical model. *Journal of Renewable and Sustainable Energy*, 6(1):013101, 2014.