Wind Power prediction with Machine Learning

Project Final Report

Geoffroy Dunoyer
Jean-Baptiste Gourlet
Guillaume Griehser
Magali Mellon

ABSTRACT

Wind energy is one of the cleanest and the most renewable resources that we have on earth. Wind's instability makes its speed and its direction difficult to forecast. Thus, to improve the reliability of this resource, we need to be able to predict its parameters with accuracy.

In recent years, artificial intelligence techniques have been developed in order to apprehend and solve this problem: in particular, machine learning algorithms. We implemented two models: the linear regression's and the k nearest neighbour to predict wind power.

We used the data provided by the *National Renewable Energy Laboratory* to analyze a turbine in California.

One of the asset of linear regression is the fact that there is no parameter to optimize whereas with kNN we have to find the k which fits the best with our model. However, the results we obtained show that the kNN algorithm is better for predicting wind power than linear regression.

Our predictions were made with a set time horizon of n×10 minutes. We obtained a linear regression between n and the mean squared error that we wanted to minimize. Then, our work shows that the k nearest neighbour model is good for predicting wind power from 10 to 30 minutes.

1 INTRODUCTION - MOTIVATION

Wind energy is playing an increasingly important part for ecologically friendly power supply. The volatility of wind power is inhibiting the performance of wind power plant. That is why we expect to develop a method to get a forecast of wind power thanks to machine learning.

Nowadays developing new ways to produce energy is a critical issue. Indeed, the need to reduce the carbon footprint and the risks related to the use of nuclear power plants we need to develop cleaner and more efficient sources of energy. Wind power is a renewable energy which use dates back to the seventh century and the invention of windmills.
Nevertheless wind is known to be a very unstable and random resource. It is very difficult to forecast wind speed and wind direction, as well as other meteorological parameters.

In order to increase the sharing of wind power it is necessary to improve the wind electricity quality and reliability. Some solutions to solve wind variations as balance management and energy storage have already been studied. But they are not enough efficient yet. To improve these solutions, power generation forecast and wind speed forecast could be useful.

Researchers, system operators and utilities have already worked on improving wind scheduling and dispatching methods. Thus, the influence of the intermittent nature of wind power is reduced and wind energy harvesting is improved. The goal is to create a model with a sample of wind data and test it with other data. As a second step, we will compare our outcome with the one from other studies.

With all the work already done during the past few years, in particular with machine learning models or artificial intelligence techniques, some ways to forecast wind power have already emerged. Wind power predictions can be grouped by their time scales into four categories: very short-term (few seconds - 30 min), short-term (30 min - 6 h), medium-term (6 h - 1 day) and long-term (longer than 1 day)

[1]. A similar domain to wind power prediction is time series prediction for solar power outpour.

2 PROBLEM DEFINITION

2.1 Wind Power

Wind is a moving object (the air) which has a mass and a kinetic energy. Thus, we can believe that it is possible to produce more energy when the wind speed is higher. We'll show that with equations:

$$Kinetic\,Energy = \frac{1}{2}mv^2 = \frac{1}{2}\rho V v^2$$

With m the mass, V the volume, ρ the density and v the speed.

If S is the area that wind passes across during time t, we have:

$$Wind\,Energy = \frac{1}{2}\rho S t v^3$$

Then,

$$Wind\,Power = \frac{E}{t} = \frac{1}{2}\rho S v^3$$

The last equation gives the total power within the wind. However, man is able to transform only a part of this power into useful energy.

2.2 Wind Turbine Power

The potential power of a wind turbine is calculated as watt (W). The energy is calculated as a function of time in W-hour: 1 kW of power delivered during one hour gives 1 kW.h.

Wind turbines recover the wind kinetic energy by slowing the wind in the space determined by the surface of their rotor. As we saw in the previous paragraph, this power is proportional to:

$$Power = \frac{1}{2}\rho S v^3$$

We denote by r the radius of the blades,

$$Power = \frac{1}{2}\rho \pi^2 r^2 v^3$$

Other elements that determine the output power of a wind turbine are [2]:

- Number and shape of blades
- The mechanical efficiency of the rotor towards the axis of the generator
- The electrical efficiency of the generator

We observe that theoretically, the power is proportional to the cube of wind speed. However, in reality, this power is limited by a speed range: there is a minimum starting speed and a maximum safety speed.

2.3 Power Prediction

In our model, we want to predict the wind power production for a single turbine at a date $t+\lambda\Delta t$, $\lambda \in \mathbb{N}$, knowing past power measurements at dates t,...,$t-\mu$, $\mu \in \mathbb{N}$.

λ is the horizon of the prediction.

2.3.1 *Notations.*

The features will be:

$$\mathbf{X} = (\mathbf{x}_0, \ldots, \mathbf{x}_\mu) \in \mathbb{R}_{(\mu+1)\times m}$$

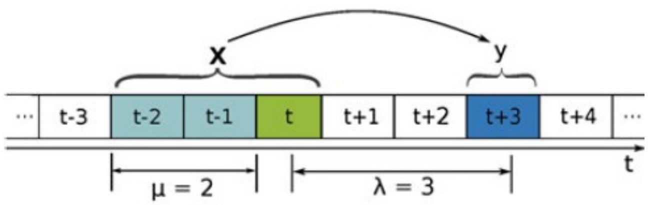
And the labels:

$$Y = [y_0, \ldots, y_\mu] \in \mathbb{R}_{(\mu+1)\times 2}$$

Where:

$$\forall t, \begin{cases} \mathbf{x}_t & = & [year(t), \ldots] \\ y_t & = & [power(t),\ wind\,speed(t)] \end{cases}$$

We want to predict with regression the label $y' = y_{t+\lambda\Delta t}$ for a $\mathbf{x}' = \mathbf{x}_{t+\lambda\Delta t}$ with regression.



2.3.2 *Accuracy Function.*

We will use the mean squared error technique to get the best accuracy. Thus, we want to minimise the error function:

$$E = \frac{1}{\mu + 1} \sum_{t=0}^{\mu} \|f(\mathbf{x}_t) - y_t\|_2^2$$

2.3.3 *Linear Regression.*

First we will use linear regression to predict wind power production.

We assume that

$$Y = \mathbf{X}w$$

The goal is to compute w by minimizing $\|Y - \mathbf{X}w\|_2^2$ If we compute the derivate and set equal to zero, we have an analytical solution:

$$w = (X^T X)^{-1} X^T Y$$

In practice, we won't use this formula, we prefer using the linear regression function of scikit-learn which is more effective.

2.3.4 *The k Nearest Neighbours.*

Then we will use an other machine learning algorithm to compare the results and find which one fits the best with our problem.

3 RELATED WORK

In their work [3], Treiber, Heinermann and Kramer focus on short term wind power prediction and use data from the National Renewable Energy Laboratory (NREL), which are designed for a wind integration study in the western part of the United States. First, they make prediction for an individual turbine and then they generalize and show that it is possible to predict turbine power for an entire wind park. Their pattern is based on statistical approaches. They consider the prediction problem as a regression one and implement different techniques such as linear regression, kNN and SVM.

We was mainly inspired by their work and we tried to implement by ourselves their model in order to compare our results. However, as we will see later, we only managed to work on a single turbine, not on an entire park.

In his thesis [4], Yiqian Liu studied seven wind farms in Ontario, Canada. He implements nine different machine learning and deep learning algorithms to find the best one. He proves that support vector machine has the best overall performances for both short-term and longer prediction. However, deep learning does not fit to basic wind power predictions.

One way to improve linear regression techniques is to use machine learning ensembles, which is done by Justin Heinermann and Oliver Kramer in [5]. They worked on large wind time series data which come from real measurements to compare homogeneous and heterogeneous machine learning ensembles. They show that heterogeneous ensembles made of many base algorithms are much more efficient and run faster than homogeneous ensemble regressors composed by decision trees, kNN and SVM.

In a conference paper [6], Kramer et al. describe WindML, a Python-based framework which linked wind energy to machine learning techniques. They work on developing tools to analyze the growing wind energy data. They introduce different use cases and some modules of WindML. These modules concern both basic machine learning algorithms and sophistical techniques for dealing with high-dimensional time series.

4 METHODOLOGY

4.1 Approaches used

The two algorithms we used were computed using the Python module scikit-learn in order to save time and to be sure that the algorithm is well optimised

4.1.1 Linear Regression. Here there is no parameters to optimise. That is why we will choose this algorithm to set the labels we will introduce.

4.1.2 k Nearest Neighbours. Here it is important to set the number of neighbours. We will set it in the optimisation section.

4.2 Description of the data set:

The data we used was provided by the *National Renewable Energy Laboratory* [7] [8] [9] [10]. It provides a toolkit for scientists we can use online and get data related to wind in the United States of America. What interested us in our project is the power produced by a turbine at an altitude of 100m and the wind speed. We chose a turbine whose GPS coordinates are (38.0625;-122.125), in Benicia, California and we got data from year 2004 to 2006. The measures were done each 10 minutes. In a second time, we selected 4 more places to be sure that the situation was not special. These cities are:

- Las Cruces, New Mexico
- Atlantic City, Nevada
- Brockway, Oregon
- Lincoln Park, Michigan

4.3 Data Selection:

Data selection is directly related with the time scale performances we want. So, in order to make this selection and given the fact that the minimum time prediction we can make here is 10 minutes, we arbitrary fixed the time horizon to 30 minutes.

4.3.1 Selection of past power measurements: The first approach could be to only consider the wind power at a time *t*. In fact the best compromise is to set μ to two or three that is to say to consider the measures 10 or 20 minutes before *t*.

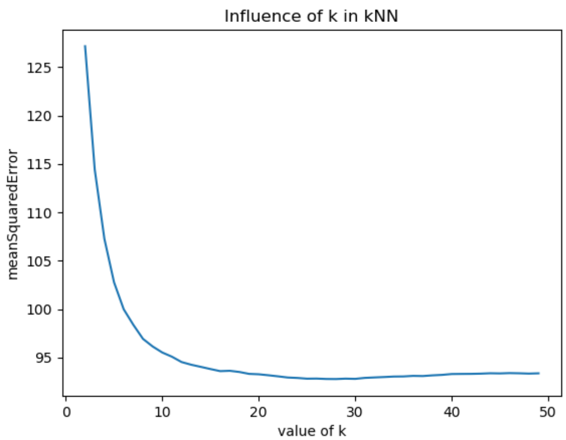
4.3.2 Wind influence: In the thesis we used, they did not implement the wind speed. But adding these features considerably decrease the mean squared error. A justification of this decrease is that there can be a power output equals to zero when the wind is not powerful enough. So the precision is higher when adding the wind. That makes sense to add this data because on each turbine it is easy to set a wind sensor.

4.4 Optimisation of parameters

4.4.1 Number of past measurements: In this part, as we say, we use the linear regression model.

city	μ=0	μ=1	μ=2	μ=3	μ=4	μ=5
Benicia	4.971	4.535	4.476	4.461	**4.456**	4.456
Las Cruces	9.925	9.514	9.516	**9.512**	9.516	9.519
Atlantic City	6.779	6.329	6.219	6.186	6.183	**6.181**
Brockway	7.054	6.684	6.647	**6.643**	6.654	6.658
Lincoln Park	17.49	16.79	16.75	16.76	16.74	**16.73**

4.4.2 Number of neighbours in kNN: . After we studied what is the best number k of closest neighbours. We only did that on sample from Las Cruces with μ=3.



As we can see on this graphic we should take between 25 and 30 neighbours.

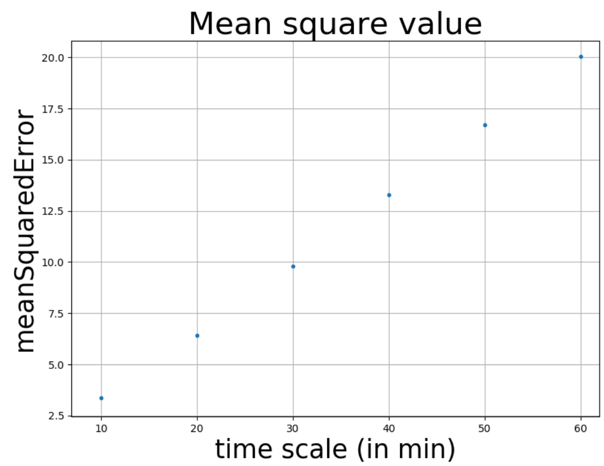
5 EVALUATION

5.1 Method of evaluation

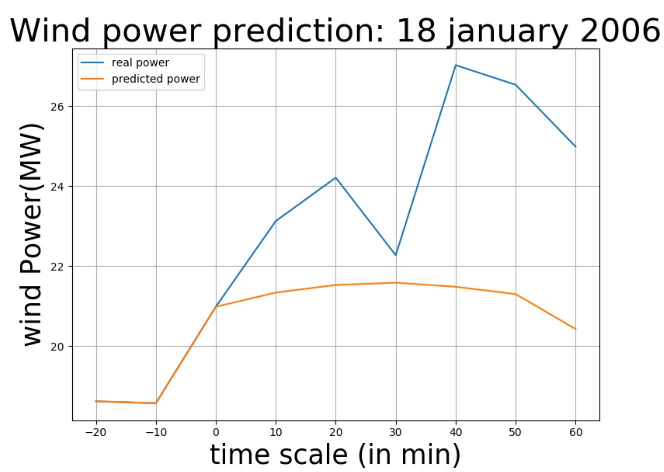
Once we optimised our models, we can run qualitative experiments. For that we use the dataset of year 2005 to train our model and we run our experiments on year 2006 for each turbines.That makes sense if we consider an industrial point of view. In order to limit the number of results, we only worked with the k Nearest Neighbours model.

What we clearly want to figure out here is the influence of time scale prediction on our performance.

5.2 Results



In this graph, we clearly see that the relations between time scale prediction and mean squared error is linear and as we expected, the error increase when we want to predict in a longer time.



What we plotted here is the prediction of wind (red curve) and the real wind (blue curve) at a date arbitrary chosen for the turbine located in Las Cruces, for 1 hour. The result is not very meaningful and it appears that the best prediction is made for a 30 minutes scale, which is normal, because we optimised our parameters considering it was the scale we wanted. What could have been done is to optimise the numbers of parameters selected for each time scale to have better results and prevent overfitting.

6 CONCLUSION

As a conclusion we would like to link our work with concrete industrial questions. Here we were able to predict the power for one turbine, we could extend it to a wind park for example. Moreover, what is interesting is more an idea of the power value rather than a precise figure.

We could as always, try to optimise more and more our models but the main issue of this not that, we think. For example, it might be interesting to link our models with the real electric consumption of American people in order to set other production means for example.

REFERENCES

[1] Saurabh S Soman, Hamidreza Zareipour, Om Malik, and Paras Mandal. A review of wind power and wind speed forecasting methods with different time horizons. In *North American power symposium (NAPS), 2010*, pages 1–8. IEEE, 2010.

[2] Specialized Ooreka authors. Puissance eolienne. https://eolienne.ooreka.fr/astuce/voir/352953/puissance-eolienne.

[3] Nils André Treiber, Justin Heinermann, and Oliver Kramer. Wind power prediction with machine learning. In *Computational Sustainability*, pages 13–29. Springer, 2016.

[4] Yiqian Liu. Machine learning for wind power prediction. Shandong University, 2013.

[5] Justin Heinermann and Oliver Kramer. On heterogeneous machine learning ensembles forwind power prediction. pages 54–59. Department of Computing Science, University of Oldenburg, 2015.

[6] Oliver Kramer, Fabian Gieseke, Justin Heinermann, Jendrik Poloczek, and Nils André Treiber. A framework for data mining in wind power time series. In Wei Lee Woon, Zeyar Aung, and Stuart Madnick, editors, *Data Analytics for Renewable Energy Integration*, pages 97–107, Cham, 2014. Springer International Publishing.

[7] C Draxl, BM Hodge, A Clifton, and J McCaa. Overview and meteorological validation of the wind integration national dataset toolkit. *NREL/TP-5000-61740. Golden (CO): National Renewable Energy Laboratory (forthcoming), Tech. Rep.*, 2015.

[8] Caroline Draxl, Andrew Clifton, Bri-Mathias Hodge, and Jim McCaa. The wind integration national dataset (wind) toolkit. *Applied Energy*, 151:355–366, 2015.

[9] W Lieberman-Cribbin, Caroline Draxl, and Andy Clifton. *A guide to using the WIND Toolkit validation code.* National Renewable Energy Laboratory, 2014.

[10] Jack King, Andy Clifton, and Bri-Mathias Hodge. *Validation of power output for the WIND Toolkit.* National Renewable Energy Laboratory, 2014.